Political Economy Lecture Notes

Daron Acemoglu

Contents

These are preliminary notes. They include typos and mistakes.

Part 1

Issues and Evidence

CHAPTER 1

General Issues

The first lecture is a broad overview of the main issues and the evidence. The rest of the course will be about formal models. However, it is useful to begin with a relatively broad discussion, without formalizing every claim. Some of the statements here will be formalized in the models discussed later, and for those that are not formalized, you will have the tools to formalize them depending on your own interests.

1.1. What Are Institutions?

1.1.1. Different views of institutions. One of the first things you learn in economics as an undergraduate is that the way that societies are organized, or briefly their "institutions", are an important determinant of their economic performance—though perhaps without being clear what "institutions" are. For instance, you learn that competitive markets, in conjunction with well-defined and enforced property rights, allocate resources efficiently. An economy with such a set of institutions is then contrasted to other ways of organizing the economy. For example, the most common argument is that the Soviet Union did less well economically because it tried to allocate resources via central planning (though see Allen, 2003). This debate about the pros and cons of socialist versus capitalist institutions dates back at least to the 1920s.

One argument (now obviously discredited) was that socialism would not face incentive problems because people's preferences would change when society was changed (just another example that one has to be careful about claims regarding "manipulation of preferences"). In response, von Mises raised the 'calculation problem.' He argued that it would be simply impossible for a socialist central planner to compute efficiently the allocation of resources.

In the 1930s Lange (1936, 1937) famously argued that since the central planner knew the Walrasian system of supply and demand equations, he would be able to mimic the market and solve the equations for the allocation of resources. The famous response to Lange came from Hayek (1945), who argued that the price system was able to communicate and aggregate information much more efficiently than an individual or a computer could do. Of course, in hindsight, the incentive problems of socialism seem even more first-order than those calculation problems.

Another example of a traditional discussion of the importance of institutions comes from economic history. Modern capitalistic economies are often contrasted with the feudal economies of the middle ages where labor and land were not allocated via the market. Many economic histories indeed emphasize how economic development is linked to the emergence of markets, trade, exchange etc. (see Hicks, 1969, or Wrightson, 2000). This is therefore an example of the evolution of a specific set of institutions playing an important role in economic development.

A final obvious example of institutions is the *tragedy of the commons* and problems with common ownership of property, which feature prominently in both discussions of pre-capitalist land systems and in todays environmental debates. These problems refer to excessive use of a common resource, such as land or a stock of game or an environmental resource. The tragedy of commons is often viewed as a failure of "institutions". A set of institutions establishing clear property rights would (or might) avoid the tragedy of the commons.

There is a counter-tradition in economics emanating from Coase (1960). The so-called "Coase Theorem" claims that in a world without 'transactions costs' rational individuals can always negotiate to an allocation of resources that is efficient. To be exact, this theorem actually requires an initial allocation of property rights that are secure, so it does presume some type of institutional background (though it argues that the specifics of who has property rights is not that important). Moreover, the Coase Theorem is in fact not exactly a theorem. It holds only under a very restrictive set of assumptions, even though most of the discussion in applied work that appeals to it makes it sound as if it holds under a very general set of assumptions. In the political economy context, we will highlight one specific reason why Coase Theorem type reasoning will not hold, which is related to holdup/commitment type problems (Farrell, 1987, discusses how informational asymmetries may lead the Coase Theorem to fail). However, even without introducing such problems or asymmetric information or other "imperfections," the Coase Theorem may not hold. A very interesting paper by Jackson and Wilkie (2005) considers a static game in which individuals can make action-dependent offers to each other and show that generally the Coase Theorem type outcomes will not arise, because individuals will not make their offers simply to achieve efficiency, but also to extract resources from each other.

For our purposes, the Coase Theorem is a useful benchmark, because it leads to a specific view of institutions, which we will refer to as *efficient institutions view* below, where even if institutions matter, they themselves are the outcome of some type of negotiation among individuals and social groups, so that dysfunctional and highly inefficient institutions can be avoided.

Leaving these issues aside for a second, it is also useful to note that the institutional structure, while important in economic models, is often left implicit and almost always taken as exogenous. For example, in the Arrow-Debreu model people have endowments and there are competitive markets in which they can buy and sell. It is clear who has what and nobody gets to try to take someone else's endowment and nobody refuses to honor a contract.

However, even when they explicitly enter into the analysis (for example, the vast literature on missing or imperfect markets), most often these institutions are treated as exogenous, and there is little effort to understand why these institutions vary across countries. Usually market imperfection of markets are tied to some fundamental "technological" problem, such as an informational asymmetry, rather than institutional differences across societies.

Perhaps more surprisingly, much of political economy also treats institutions as exogenous. There are many theoretical and empirical works on the effect of different electoral systems or governance structures on economic performance, but they typically start from an exogenously-given set of institutional rules, such as a given set of electoral rules, a dictatorship, a particular form of democracy etc..

This course is explicitly about understanding both the effects of institutions on economic outcomes and the *emergence and persistence* of institutions in equilibrium.

1.1.2. What are institutions? There is no correct answer to this question, and the appropriate answer most probably depends primarily on the use that we want to put the notion of institutions to. Douglass North, for example, emphasizes the role of institutions as "to reduce uncertainty by establishing a stable (but not necessarily efficient) structure to human interaction." This is useful but somewhat too loose a definition.

An alternative, but related, definition is that institutions refer to "the rules of the game" or more specifically to the extensive form of the exact game that the agents are playing (e.g., as suggested by David Kreps). This definition is also useful, but may be too encompassing. At least in the context of the models we look at in this course, what the institutions are will be clear. Social scientists are still struggling for the most useful general definition.

Probably the first question we should address is whether, and how well, institutions adapt to the economic requirements of the society. For example, the classical Marxist approach views the "superstructure," which roughly corresponds to the notion of institutions here, as simply determined by the underlying economic forces and not having large independent effects.

More pertinent to the discussion here, following Acemoglu (2003a) and Acemoglu, Johnson and Robinson (2005), we may want to distinguish between several different views of institutions. In coming up with this list we abstract from other interesting approaches which are beyond the scope of this course, for instance the evolutionary models of Nelson and Winter (1982) and Young (1998).

(1) *Efficient institutions view:* As already hinted above, according to this view, institutions may matter for economic outcomes, but societies will choose the institutions that maximize their total surplus. How this surplus will be distributed among different groups or agents does not affect the choice of institutions. An example of this view would be Demsetz's theory of property rights, which is also very similar to the theory of property rights that is advanced in North and Thomas. According to Demsetz (1967), enforcing property rights has some costs, but also is beneficial economically. When the benefits exceed the costs, property rights will be enforced. As an example, he offers variations in property rights in land among American Indians. In many Indian societies, there were no property rights in land because land was abundant, and the inefficiencies from overhunting were relatively small. However, property rights developed following the commercialization of the fur trade, because with the possibility of selling fur, the overhunting problem ('tragedy of the commons') became more serious, so the benefits from establishing property in over land increased. This caused the emergence of property rights in land among the tribes engaged in fur trading. The underlying reasoning of this view again comes from the Coase Theorem. Therefore, even in a world where institutions matter, when different economic parties can negotiate costlessly, they will be able to bargain to internalize potential externalities. The farmer, who suffers from the pollution created by the nearby factory, can pay the factory owner to reduce pollution.

The same reasoning can be applied to political situations. If the current laws or institutions benefit a certain group while creating a disproportionate cost for another, these two groups can negotiate to change the institutions. By doing so they will increase the size of the total surplus ("the pie" that they have to divide between themselves), and they can then bargain over the distribution of this additional surplus.

The efficient institutions view suggests that institutions differ because countries have different underlying characteristics which make different economic institutions efficient. For example, property rights may be insecure in Mali but this is because, since a large part of Mali is the Sahara desert, even secure property rights would generate few economic benefits relative to the cost of creating them. This view, taken to its logical extreme (e.g., Djankov et al, 2003) may even suggest that Soviet socialism was an efficient way of organizing the economy given the circumstances that faced Stalin. Returning to Demsetz, this view would also imply that even though during the Apartheid era in South Africa Africans were not allowed by law to own property in white areas and that all land had to be held communally in the African homelands/Bantustans (13% of the country), this was efficient.

For our purposes, the most important implication of the efficient institutions view is that we should not look at differences in institutions as a key determinant of economic development or economic performance, since societies will have adopted the "right" set of institutions for their circumstances. Therefore, the evidence discussed in the next lecture that institutional differences have a significant causal effect on economic outcomes will go against the efficient institutions view.

(2) *The Social conflict view:* An alternative is that institutions emerge as a result of economic agents' choices, but are not necessarily efficient. But why isn't a set of institutions that maximize output chosen?

Because according to this view, institutions are not always chosen by the whole society (and not for the benefit of the whole society), but by the groups that control political power at the time (perhaps as a result of conflict with other groups demanding more rights). These groups will choose the institutions that maximize their own rents, and the institutions that result may not coincide with those that maximize total surplus. North (1981, Chapter 3), in the same vein, argues that in all societies there is a: "persistent tension between the ownership structure which maximizes the rents to the ruler (and his group) and an efficient system that reduces transaction costs and encourages economic growth".

For example, institutions that enforce property rights by restricting state predation will not be in the interest of a ruler who wants to appropriate assets in the future. By establishing property rights, this ruler would be reducing his own future rents, so may well prefer institutions that do not enforce property rights, and therefore do not constrain him from appropriating assets in the future to those that do.

Therefore, equilibrium institutions will not be those that maximize the size of the overall pie, but the slice of the pie taken by the powerful groups.

Why doesn't a Coase theorem type reasoning apply? This is what we will discuss in detail in many different parts of this course. But for now, it is useful to note that one possible and obvious reason for why the Coase theorem may not apply in politics is *commitment problems*. If a ruler has political power concentrated in his hands, he cannot commit not to expropriate assets or revenues in the future. The enforcement of property rights, which would encourage investment by the agents, requires that he credibly relinquishes political power to some extent. But according to the Coasian bargain, he has to be compensated for what he could have received using this power. Herein lies the problem. When he relinquishes his power, then he has no guarantees that he will receive the promised payments in the future. Therefore, by their very nature, institutions that regulate political and social power create commitment problems, and prevent Coasian bargains that are necessary to reach efficient outcomes.

(3) *The ideology/beliefs view:* According to this view societies may have different institutions because people have different beliefs about what is best for society. Some societies get it right and some get it wrong and those that get it right ex post are prosperous. It is clear that belief differences and ideology do matter in practice, the real question is whether this can systematically explain the massive variation in outcomes we observe in the world. It is also possible of course that people care about the organization of society for non-economic reasons and are prepared to sacrifice output in order to have a set of institutions that they feel are better.

(4) *The incidental institutions view:* While the efficient institutions view is explicitly based on economic reasoning, a different approach, which downplays choices and emphasizes the development of institutions as a by-product of other social interactions, is more popular among many political scientists and sociologists. According to this view, the set of political and economic institutions emerge not as a choice of economic actors, but is an incidental consequence of other actions. An interesting example of this is the work by Tilly (1990). Building on the Weberian tradition, Tilly proposed a theory of the formation of modern states, which argues that modern state institutions such as fiscal systems, bureaucracy and parliaments are closely related to the need to raise resources to fight wars and thus arose in places with incessant inter-state competition. Building on this work Herbst (2000) argued that

the absence of such processes led to a very different pattern of state formation historically in Africa, something which left African countries with states which were incapable of providing order or public goods. Another example would be the literature on legal origins where the fact that Latin American countries have French legal origin stems from the coincidence that they were colonized by the Spanish who happened to have a legal system which was more compatible with the Civil Code than the Common law.

What distinguishes the efficient institutions view from the other three is that according to the efficient institutions view, there should not be meaningful institutional differences translating into different economic outcomes— institutional differences should simply reflect differences in economic environments, rather than cause such differences. Therefore, as noted above, empirical evidence that shows that "exogenous" institutional differences matter for differences in economic outcomes will support one of the other three views. We will discuss this type of empirical evidence below.

What distinguishes the first two approaches from the incidental institutions view is that according to the incidental institutions view, we cannot try to understand institutional differences as emerging from different economic calculations. As a result, we cannot ask questions of the following form: "why aren't the existing set of institutions being replaced by a new set of institutions that are more beneficial for the whole society or for certain groups?" These types of counterfactual questions are a very attractive feature of the economic approach. The first two approaches are therefore more in line with economic research in general, and will be in the starting point of the approach in this class.

But there are important differences between these two views as well. In the social conflict view, conflict between social groups is an essential element of institutions and differences in the nature of this conflict will lead to different sets of institutions. In contrast, in the efficient institutions view, conflict between different groups or agents is not important, and institutional differences will mostly emerge from differences in the economic environment or the costs of designing institutions.

The approach in this course will be very much based on the social conflict view of institutions. While the other views may also contribute to differences in institutions we observe in practice, the perspective in this course will be that the bulk of the differences in institutions emerges as an equilibrium outcome from well-specified games, in which there is conflict of interest between individuals or social groups. This is a natural perspective for an economic approach to institutions, since it emphasizes both economic interests and equilibrium. It also tends to suggest that we have to think of dynamic interactions as an important element of the overall picture, and in fact, we will see that institutions become much more meaningful in dynamic situations.

1.1.3. Types of institutions. Let us start with an approach similar to the social conflict institutions view. But now, we need to understand what institutions are.

For this purpose, we can distinguish between the following types of institutions (or different roles of institutions):

- Economic institutions: We can think of economic institutions as determining the "economic rules of the game"—in particular, the degree of property rights enforcement, the set of contracts that can be written and enforced, and some of the rules and regulations that determine the economic opportunities open to agents. Common examples of economic institutions would therefore include individual property rights, commercial law, contract law, patent law, the type of credit arrangements, etc.
- Political institutions: In contrast, political institutions determine the "rules of the political game." Political institutions help to regulate the limits of political power and determine how political power changes hands. Common examples of political institutions would include the constitution, electoral rules, constraints imposed on the power of the executive by other branches of the government, the number of veto players, the extent of checks and balances etc.

(In addition, we can think of "social institutions" or "legal institutions" though we can also think of, for example, legal institutions, as part of economic institutions—to the extent that they affect contracting—and part of political institutions—to the extent that they regulate the allocation of political power in society.)

These distinctions between different types of institutions are useful, but they are not completely tight and we do not often know how to map them to the data. These distinctions therefore should not give the impression that we know exactly what these different "institutions" do in practice. There is considerable uncertainty about the role of these different objects for economic outcomes (more on this below).

At this point, it is also useful to note that in some situations we might want to distinguish economic policies from economic institutions. For example, a tax on capital by the government is a policy not an institution. For many of the applications, economic policies will be very similar to economic institutions, but one might want to bear in mind that they may be easier to change than economic institutions. Again this distinction is not very tight. We refer to labor market policies, such as the extent of firing costs or whether or not trade unions have a closed shop agreement as 'labor market institutions.' Is it important to distinguish between policy and institutions? Consider the following example: currently in Zimbabwe the government of President Robert Mugabe is expropriating the property of the white farmers. In the 1960s and 1970s a succession of governments in Ghana used the monopoly purchasing power of the Cocoa Marketing Board to tax farmers at punitively high rates (Bates, 1981). Though governments in Ghana did not try to take the land of the cocoa farmers, they made the value of their assets almost zero. What is the difference?

It is conventional to subdivide institutions into:

- Formal institutions, for example, whether the country in question has a Supreme Court, separation of power, parliamentary system etc.
- Informal institutions, which determine how a given set of formal rules and informal institutions function in practice. For example, many Latin American countries have a presidential system similar to the U.S., but in practice, they have very different "political institutions".

[...Currently, we have very little understanding of how informal institutions work, and this might be an interesting area for future research...] One terminology that we have used (Acemoglu, Robinson and Verdier, 2003) is to say that societies where behavior is strongly conditioned by the formal institutions are 'strongly institutionalized' while those where it is not are 'weakly institutionalized.' An interesting example of the difference comes from the history of the Supreme Court in the US and Argentina. Starting in 1935 the Supreme Court began to rule against President Rooselvelt's New Deal Policies including ruling that the National Industrial Recovery Act was unconstitutional. In response after his landslide 1936 re-election, Roosevelt proposed judicial 'reform' which in effect would give him the right to nominate 6 new judges (there were 9 at the time) and ensure a majority. Despite huge majorities in Congress and Senate and a clear mandate, this created a massive outcry in the press and opposition in the Senate. In response the Supreme Court made some compromises, ruling that the Social Security Act and the National Labor Relations Act, two key pieces of New Deal legislation, were constitutional. Moreover, one of the most conservative judges resigned, allowing FDR to appoint a democratic. This made it impossible for FDR to get the change in the rules through the Senate and he had to drop it. This is an interesting example of how the formal rules which constitute the separation of powers and checks and balances have to be supported by informal norms.

Contrast this to the experience of Argentina under Perón. When Perón (who had been secretary of labor in the previous military Junta which had ruled since 1943) was first democratically elected president in 1946 the Supreme Court had recently ruled unconstitutional an attempt to create a new national labor relations board. The labor movement was a key pillar of the Peronist movement and he sought the impeachment of 4 or the 5 members of the Court. In the end 3 were removed and the Chamber of Deputies and the Senate supported this. Following this the Court did not provide any checks on Perón's policies. There then followed a sequence of transitions between civilian and military governments in Argentina (see Acemoglu and Robinson, 2006, Chapter 1). The 1946 impeachment established a new norm so that whenever a political transition too place, the incoming regime either replaced the entire existing Supreme Court or impeached most of its members (Helmke, 2005, Chapter 4). The attached figure from Iaryczower, Spiller and Tommasi (2002) dramatically shows the declining average tenure of Supreme Court judges in Argentina over this period. Fascinatingly in 1990 when the first transition between democratically elected governments occurred, the incoming Peronist President, Menem, complained that the existing Supreme Court, which had be appointed after the transition to democracy in 1983 by the Radical President Alfonsín, would not support him. He then proposed an expansion of the Court from 5 to 9 members which was duly passed and which allowed him to name 4 new judges.

We will sometimes refer to the cluster of institutions, consisting of economic institutions, political power and formal and informal political institutions, simply as "institutions".

It may also be tempting to follow political scientists and sociologists and classify institutions into two groups (with the implicit understanding that most real world institutions will fall somewhere in between):

- Predatory ("bad") institutions: as institutions that do not encourage investment and economic development.
- Developmental ("good") institutions: institutions that permit or encourage investment and growth.

Even though, this may be a useful device for discussion or even more formal thinking, there is a potential problem. Certain arrangements that are good for economic development may later become bad for in investment and development. We will see formal models illustrating this issue later in the course. For now, when convenient, we will sometimes talk of bad and good institutions.

1.1.4. Conceptions of the State. Most of the institutions we have been talking about are collective choices. When a society moves from communal to private property, this is going to be a collective not an individual decision. Though often in the economics literature the creation of private property comes about when someone decides to build a fence to keep people off their land (Tornell, 1997, for such a model) this is actually not what often happens. Before the existence of private property different people may have overlapping and conflicting claims over assets. This was particularly true historically of land. Allen's (1982) seminal study showed that the first-order effect of the enclosure movement in England, was distributional and it had large redistributive effects because the switch to private ownership effectively expropriated many people's claims (see Firmin-Sellars, 1995, for a fascinating study of the evolution of property rights in Africa).

Thus to talk about how institutions arise we need to consider how preferences are aggregated. At the center of this whole picture therefore is political power: whose preferences count? For example, in 17th and 18th century Europe, it is commonly accepted that the landed aristocracy had most political power. But what does this mean? There were clearly barriers against and checks on the exercise of this power, as exemplified by peasant revolts, beheadings of kings, and bourgeois revolutions.

Preferences can be aggregated in different ways. For example, the person with the most guns might be able to force the rest of society to accept his preferences. In the modern world we more usually think of political institutions as determining how preferences are aggregated. Central to political institutions is the institution of the state. Perhaps it is first useful to think of "the state" as the locus of political power. Probably the most common and useful definition of the state is as a monopoly (or near-monopoly) of violence, or of coercion power. That is, the state has the means to coerce other agents to perform certain tasks and abstain from certain behavior.

For example, Max Weber puts this as follows:

"the modern state is a compulsory association which organizes domination. It has been successful in seeking to monopolize the legitimate use of physical force as a means of domination within a territory.... The right to use physical force is ascribed to other institutions or individuals only to the extent to which the state permits it".

But then, who controls the state? And who and what constrains the state?

There are a number of different ways of thinking of the state:

- *The state as a non-actor:* the simplest view of the state, common in many economics and public finance textbooks, treats the state without agency—that is, the state does not have its own objective function, nor does it represent the interests of some groups in the society. It is there to enforce property rights and contracts, and provide public goods. There is also little discussion of incentives, or what is sometimes called "opportunistic" behavior, and little sense in which the state needs to have the monopoly of violence in society. This view of the state very naturally leads to calls for the state/the government to intervene when there are market failures. The driving force of the new political economy of institutions is the recognition that this is not a satisfactory view, and we will not dwell on it in the rest of this class.

- *The state as a nexus of cooperation.* This view recognizes the presence of "opportunistic" behavior on the part of the agents, but does not emphasize conflict between groups of agents (such as workers versus capitalists, or Hutu versus Tutsi). The state, by virtue of its coercive power, encourages cooperation among agents. This view is related to Hobbes' conception of the Leviathan with the monopoly on coercion serving the interest of all the citizens. According to Hobbes, without the Leviathan individuals live in "fear and danger of violence death" and their lives are "solitary, poor, nasty, brutish and short," because every man is fighting for himself and not cooperating with others. The state encourages cooperation and orderly behavior. This view is also very close to the so-called "populist" political philosophy originating from Rousseau. Rousseau argued that the state should be a reflection of the "general will" of the people— "obedience to a law we have prescribed for ourselves." When all citizens obey this general will or law, welfare in this society will be higher.

  This view is not identical to the efficient institutions view, since the potential for institutional failure is present. Nevertheless, it is closely related to viewing institutions as evolving in order to solve some potential market failures in society.

- *The state as the agent of a social group.* In these theories, the state represents the interests of a social group, such as the landowners, the capitalists, an ethnic group or some sort of elite, and uses its monopoly over violence in order to further the interests of this group. Marxist theories of the state generally fall in this category, since they view the state as controlled by the capitalists or more generally by the ruling class. However, many non-Marxist theories, perhaps going as far back as Aristotle, are also in this category. For example, it seems plausible that before the 18th-century the state in Europe looked after the interests of the landed aristocracy and the King. More recently, much political economy of Africa sees the state as the instrument of a particular ethnic group.

A related view, common in the Chicago political economy circles (e.g., Becker, 1983) and US political science in the 1950s and 1960s, sees the state as a potential aggregator of the demands of different (interest) groups in society. Here the state has no preferences per se, but reflects the net effect of the different pressures placed upon it. Nevertheless, this view is closely related to that which envisages the state as the agent of a social group, because at the end the state plays a crucial role in the allocation of resources and in creating winners and losers in the society, but it has no identity separate from that of the groups that act through it.

Consequently, both of these views are closely related to each other, and they can be thought of as the applications of the social conflict view of institutions to the role of the state.

- *The state as the grabbing hand:* in this view, the state is controlled by the bureaucrats or the politicians, who use their power to look after their own interests. This view goes back to Buchanan and Tullock, and recently has been popularized by Shleifer. In this case, the crucial question is how to control bureaucrats while ensuring that they perform the functions they are supposed to.

  The major difference between this view and the previous one is that here the conflict is not between groups that have well-defined economic interests, specific assets or association to social groups, but between whoever are the "politicians" and "bureaucrats" and the society that is supposed to monitor them.

- *The state as the autonomous bureaucracy. Weberian theories of the state.* In this view, like the previous one, the state is controlled by the bureaucrats or the politicians in the sense that they can take actions that agents themselves may not have taken. However, in this view, somehow the state could represent interests other than the narrow interests of its members. For example, in some modern Marxist theories of the state, such as Poulantzas' theory, the state looks after the interest of the capitalist class better than individual capitalists themselves would be able to. In the Weberian theories, such as Evans's embedded autonomy approach, or in Tilly's theory, the bureaucracy could be developmental and defend the interests of the whole society. For example, Evans attributes failures of many societies to states that are "weak because diffused fragments of society have stayed strong, retaining at the local level the ability to frustrate state actions". Therefore, in these theories, states need to be able to "dominate" the society in order to enact useful change. In this class of theories scholars often talk about the state being 'strong' or 'weak' which typically refers to their capacity to get what they want.

All of these views are obviously simple, and they can be combined with each other, or extended. For example, we can imagine a situation in which the state represents the interests of a social group, but which social group this is varies over time. Alternatively, one can be much more specific on the reasons why the state emerges as the monopoly supplier of coercion, and link this to the grabbing hand view of the state. We will see examples of models of this sort below.

It is also useful at this point to mention an important ingredient of models that view social groups (such as classes) as the key actors. At the end of the day, decisions are taken by individuals, so if we are going to treat social groups as the key actors, all individuals within the social groups must find a profitable to take the same actions, and often, take actions that are in the interest of the group as a whole. This leads to what Olson has termed the "free rider" problem: individuals may free ride and not undertake actions that are costly for themselves but beneficial for the group. Therefore, any model that uses social groups as the actor must implicitly use a way of solving the free-rider problem. The usual solutions are

- Ideology: groups may develop an ideology that makes individuals derive utility from following the group's interests.
- Repeated interactions: if individuals within groups interact more often with each other, certain punishment mechanisms may be available to groups to coerce members to follow the group's interests.
- Exclusion: certain groups might arrange the benefits from group action such that those who free ride do not receive the benefits of group action.

[...Currently, there is little systematic work in economics on how social groups solve the free-rider problem, and this may be an important area for future work...]

We will return to discuss some of these issues later in the class.

1.2. Developmental Vs. Predatory Institutions

It is now useful to return to a discussion of how we should think of developmental vs. predatory institutions according to the different theories.

What follows is a very schematic representation:

	Hobbesian Theories	Social Conflict Theories	Grabbing Hand View	Weberian Theories
Strong states?	Good	Ambiguous	Bad	Good
Institutional differences?	Accidental/ costs of inst. design	Due to economic incentives of various groups	Accidental/ strength of bureaucracy	Accidental
Developmental institutions?	Yes	Ambiguous	No if state is strong and can grab a lot	Yes if state is strong
Agreement on institutions?	Yes	Generally no	Generally yes	Generally yes
Which view of institutions?	Efficient institutions view	Social Conflict institutions view	Incidental institutions view	Incidental institutions view

Whether strong states are useful for economic development is a first-order question both for economic policy and economic research. It is clear that strong states are good in the Hobbesian and the Weberian theories, and bad in the grabbing hand view. As the above models illustrate, they can be good or bad in the social-conflict theories, depending on whether the state is being controlled by the group that has the investments that are more important at the margin.

Another major question that we will discuss further in the class is why there are institutional differences across countries. The Weberian theories fall in the incidental-institutions view of the world, and institutional differences reflect historical accidents. As we mentioned above, Tilly argues that in Europe there was a lot of inter-state competition, leading to strong states which played an important developmental role, while Herbst argues that there wasn't enough inter-state competition in Africa and this underlies the problems of economic development in that continent.

Also in the Hobbesian theories, institutional differences must have accidental causes, or simply reflect differences in the costs of designing the appropriate institutions. In these theories, everybody agrees what the right set of institutions should be (at least behind the veil of ignorance). But because of some "accidents", they may be unable to develop these institutions. Greif's (1994) theory of why Maghribi merchants did not develop modern institutions is that they had better informal ways of detecting cheating, so they did not need to develop these institutions. Ultimately, this lack of institutional development was costly for the Maghribi traders because they made it difficult to take advantage of expanded trade opportunities. In contrast the Genoese were not able to rely on their social networks to enforce contracts at low cost and instead had to develop formal contract enforcement institutions which adapted much better to changed environments and new opportunities.

In the grabbing hand view, everybody, other than the politicians, agree what the "correct" set of institutions are. Therefore, institutional differences either reflect accidents, or the strength of these politicians/bureaucrats.

In contrast, in the social-conflict theories, institutions serve the interests of different classes/groups, so there will be a conflict over what type of institutions should emerge, and economic incentives will determine equilibrium institutions.

1.3. Institutional Origins in the Social Conflict Theories

At this point, it is useful to push our organizing framework, the social conflict view of institutions, somewhat further, and ask more systematically what set of factors would determine differences in institutions. At a basic level, the social conflict view, as opposed to the efficient institutions view, emphasizes that different sets of institutions create different groups of winners and losers. This immediately leads to an obvious question: how does a society decide which set of institutions emerge in equilibrium (or persist or change etc.)? If winners from a given set of institutions could compensate losers, then we would be in the realm of the efficient institutions view. For example, the society could undertake an institutional reform that improves efficiency, and those benefiting from the reform compensate those who have lost out. The building block of the social conflict theories is that such ex post compensation is not possible or will not happen in equilibrium. Then how do equilibrium institutions get determined?

The answer is related to *political power*. When there is a conflictual situation, the party with greater "power" is likely to have its way. This applies to the political realm as well, and the relevant power is naturally political power. In this context, we may want to distinguish between democratic and non-democratic politics. In democratic politics, political power is the aggregation of the wishes of certain different segments of society via voting, lobbying and other political activities. In non-democratic politics, we will think of an existing *political elite*, which has sufficient political power to play a crucial role in the determination of policies and the reform of institutions (this does not necessarily mean they will always implement their ideal policy, but they have disproportionate effect on outcomes).

Anticipating what will come next, it is also useful to distinguish between two different types of political power: *de jure* and *de facto* political power. De jure political power is allocated by political institutions (such as constitutions or electoral systems), while de facto political power emerges from the ability to engage in collective action, use brute force, paramilitaries, armies, or other channels such as lobbying or bribery. Equilibrium policies will be an outcome of total political power, which consists of the composition of these two sources of power. We will see below how the interplay of these two types of political power will lead to a framework for thinking of both the effect of institutions on economics outcomes and the equilibrium determination of institutions.

The crux of many approaches to institutions (especially "inefficient institutions" as will be defined later) is the unwillingness of individuals or groups with political power to allow policies or institutions that increase the size of the "social pie" (i.e., for example, increase investment or GDP, or economic growth etc.). In this context, it is useful to mention three specific but related mechanisms for why those with political power (for example in many societies the "elites") may be unwilling to embrace efficient institutions. These are:

(1) *Hold-up:* Imagine a situation in which an individual or a group holds unconstrained political power. Also suppose that productive investments can be undertaken by a group of citizens or producers that are distinct from the "political elites", i.e., the current power holders. The producers will only undertake the productive investments if they expect to receive the benefits from their investments. Therefore, a set of economic institutions protecting their property rights are necessary for investment. Can the society opt for a set of economic institutions ensuring such secure property rights? The answer is often no (even assuming that "society" wants to do so).

The problem is that the political elites—those in control of political power—cannot commit to respect the property rights of the producers once the investment are undertaken. Naturally, ex ante, before investments are undertaken, they would like to promise secure property rights. But the fact that the monopoly of political power in their hands implies that they cannot commit to not *hold-up* producers once the investments are sunk.

(2) *Political Losers:* Another related source of inefficient economic institutions arises from the desire of political elites to protect their political power. Political power is the source of the incomes, rents, and privileges of the elite. If their political power were eroded, their rents would decline. Consequently, the political elite should

evaluate every potential economic change not only according to its economic consequences, such as its effects on economic growth and income distribution, but also according to its political consequences. Any economic change that will erode the elites' political power is likely to reduce their economic rents in the long run.

(3) *Economic Losers:* A distinct but related source of inefficiency stems from the basic supposition of the social conflict view that different economic institutions imply different distributions of incomes. This implies that a move from a bad to a better set of economic institutions will make some people or groups worse off (and will not be Pareto improving). This in turn implies that such groups will have an incentive to block or impede such institutional changes even if they benefit the whole of society in some aggregate sense. [This naturally raises the question as to how and why groups that have the political power to block change cannot use the same political power in order to redistribute some of the gains from beneficial reform to themselves after reform takes place; this may be related to the fact that there are only limited fiscal and redistributive instruments, or that the economic losers question is intimately linked to that of political losers, in that after the reform, previous elites may no longer have as much political power].

Hold-up, political loser and economic loser considerations lead to some interesting comparative static results which can be derived by considering the political institutions that lie behind these phenomena.

First, the perspective of hold-ups immediately suggests that situations in which there are constraints on the use of political power, for example, because there is a balance of political power in society or a form of separation of powers between different power-holders, are more likely to engender an environment protecting the property rights of a broad cross-section of society. When political elites cannot use their political power to expropriate the incomes and assets of others, even groups outside the elite may have relatively secure property rights. Therefore, constraints and checks on the use of political power by the elite are typically conducive to the emergence of better economic institutions

Second, a similar reasoning implies that economic institutions protecting the rights of a significant cross-section are more likely to arise when political power is in the hands of a relatively broad group containing those with access to the most important investment opportunities. When groups holding political power are narrower, they may protect their own property rights, and this might encourage their own investments, but the groups outside the political elites are less likely to receive adequate protection for their investments (see Acemoglu, 2003b).

Third, good economic institutions are more likely to arise and persist when there are only limited rents that power holders can extract from the rest of society, since such rents would encourage them to opt for a set of economic institutions that make the expropriation of others possible.

Finally, considerations related to issues of political losers suggest that institutional reforms that do not threaten the power of incumbents are more likely to succeed. Therefore, institutional changes that do not strengthen strong opposition groups or destabilize the political situation are more likely to be adopted.

1.4. Towards a Framework

The discussion above illustrates some basic issues that arise once we start looking at the world, especially at the institutional differences across societies, through the lenses of social conflict view. But as emphasized before, a major part of our objective is to understand the equilibrium determination of institutions. This requires a dynamic framework, especially since the primary importance of institutions arises in the context of holdup/commitment (which is an inherently dynamic issue). Moreover, as emphasized above, a useful perspective is to think of political institutions as the source of de jure political power in society, i.e., the set of political institutions determined today regulates the distribution of de jure political power in the future (again leading to dynamic interactions).

The most ambitious objective of this course would be to provide a complete dynamic model with endogenous economic and political institutions that can be applied to a variety of situations. Unfortunately, the newly-emerging field of political economy is not there yet (though we have clues about important ingredients). This is both good and bad: it is good, since there is a lot to do in future research; and it is bad, since we are still far from a satisfactory understanding of many important political economy questions (for example, those related to how institutional reform can be undertaken etc.).

Let us start with a schematic representation of a particular dynamic model which could be useful in thinking about the equilibrium evolution of economic and political institutions (this is by no means the only possible model one could have, but one that we find useful and acts as an organizing framework for much of what will come next):

$$\begin{array}{ccc}
\text{political} & & \\
\text{institutions}_t & \Longrightarrow & \left.\begin{array}{c}\text{de jure}\\ \text{political}\\ \text{power}_t\\ \&\\ \text{de facto}\\ \text{political}\\ \text{power}_t\end{array}\right\} \begin{array}{c}\Longrightarrow \begin{array}{c}\text{economic}\\ \text{institutions}_t\end{array} \Longrightarrow \left\{\begin{array}{c}\text{economic}\\ \text{performance}_t\\ \&\\ \text{distribution}\\ \text{of resources}_{t+1}\end{array}\right.\\ \\ \Longrightarrow \begin{array}{c}\text{political}\\ \text{institutions}_{t+1}\end{array}\end{array} \\
\text{distribution} & & \\
\text{of resources}_t & \Longrightarrow &
\end{array}$$

This is a dynamical system, even though it is expressed schematically. The two state variables of this dynamical system are political institutions and the distribution of resources, and the knowledge of these two variables at time t is sufficient to determine all the other variables in the system. While political institutions determine the distribution of de jure political power in society, the distribution of resources influences the distribution of de facto political power at time t. These two sources of political power, in turn, affect the choice of economic institutions and influence the future evolution of political institutions. Economic institutions determine economic outcomes, including the aggregate growth rate of the economy and the distribution of resources at time $t + 1$. Although economic institutions are the essential factor shaping economic outcomes, they are themselves endogenous and determined by political institutions and distribution of resources in society.

There are two sources of persistence in the behavior of the system: first, political institutions are durable, and typically, a sufficiently large change in the distribution of political power is necessary to cause a change in political institutions, such as a transition from dictatorship to democracy. Second, when a particular group is rich relative to others, this will increase its de facto political power and enable it to push for economic and political institutions favorable to its interests. This will tend to reproduce the initial relative wealth disparity in the future. Despite these tendencies for persistence, the framework also emphasizes the potential for change. In particular, "shocks", including changes in technologies and the international environment, that modify the balance of (de facto) political power in society and can lead to major changes in political institutions and therefore in economic institutions and economic growth. [We will later argue that perhaps this economic channel for persistence might be overemphasized in some of the literature...]

Now let's investigate in more detail where this schematic representation comes from

1. Economic institutions matter for economic growth because they shape the incentives of key economic actors in society, in particular, they influence investments in physical and human capital and technology, and the organization of production. Although cultural and

geographical factors may also matter for economic performance, differences in economic institutions are the major source of cross-country differences in economic growth and prosperity. Economic institutions not only determine the aggregate economic growth potential of the economy, but also an array of economic outcomes, including the distribution of resources in the future (i.e., the distribution of wealth, of physical capital or human capital). In other words, they influence not only the size of the aggregate pie, but how this pie is divided among different groups and individuals in society. We summarize these ideas schematically as (where the subscript t refers to current period and $t+1$ to the future):

$$\text{economic institutions}_t \Longrightarrow \left\{ \begin{array}{c} \text{economic performance}_t \\ \text{distribution of resources}_{t+1} \end{array} \right. .$$

2. Economic institutions are endogenous. They are determined as collective choices of the society, in large part for their economic consequences. However, there is no guarantee that all individuals and groups will prefer the same set of economic institutions because, as noted above, different economic institutions lead to different distributions of resources. Consequently, there will typically be a *conflict of interest* among various groups and individuals over the choice of economic institutions. So how are equilibrium economic institutions determined? If there are, for example, two groups with opposing preferences over the set of economic institutions, which group's preferences will prevail? The answer depends on the *political power* of the two groups. Although the efficiency of one set of economic institutions compared with another may play a role in this choice, political power will be the ultimate arbiter. Whichever group has more political power is likely to secure the set of economic institutions that it prefers. This leads to the second building block of our framework:

$$\text{political power}_t \Longrightarrow \text{economic institutions}_t$$

3. Implicit in the notion that political power determines economic institutions is the idea that there are conflicting interests over the distribution of resources and therefore indirectly over the set of economic institutions. But why do the groups with conflicting interests not agree on the set of economic institutions that maximize aggregate growth (the size of the aggregate pie) and then use their political power simply to determine the distribution of the gains? Why does the exercise of political power lead to economic inefficiencies and even poverty? We will explain that this is because there are commitment problems inherent in the use of political power. Individuals who have political power cannot commit not to use it in their best interests, and this commitment problem creates an inseparability between efficiency and distribution because credible compensating transfers and side-payments cannot be made to offset the distributional consequences of any particular set of economic institutions.

4. The distribution of political power in society is also endogenous. Political institutions, similarly to economic institutions, determine the constraints on and the incentives of the key actors, but this time in the political sphere, and therefore regulate the distribution of de jure power in society. Examples of political institutions include the form of government, for example, democracy vs. dictatorship or autocracy, and the extent of constraints on politicians and political elites. For example, in a monarchy, political institutions allocate all de jure political power to the monarch, and place few constraints on its exercise. A constitutional monarchy, in contrast, corresponds to a set of political institutions that reallocates some of the political power of the monarch to a parliament, thus effectively constraining the political power of the monarch. This discussion therefore implies that:

$$\text{political institutions}_t \Longrightarrow \text{ de jure political power}_t$$

5. As noted above, there is more to political power than de jure power allocated by political institutions, however. A group of individuals, even if they are not allocated power by political institutions, for example as specified in the constitution, may nonetheless possess political power. Namely, they can revolt, use arms, hire mercenaries, co-opt the military, or use economically costly but largely peaceful protests in order to impose their wishes on society. We refer to this type of political power as de facto political power, which itself has two sources. First, it depends on the ability of the group in question to solve its collective action problem, i.e., to ensure that people act together, even when any individual may have an incentive to free ride. For example, peasants in the Middle Ages, who were given no political power by the constitution, could sometimes solve the collective action problem and undertake a revolt against the authorities. Second, the de facto power of a group depends on its economic resources, which determine both their ability to use (or misuse) existing political institutions and also their option to hire and use force against different groups. Since we do not yet have a satisfactory theory of when groups are able to solve their collective action problems, our focus will be on the second source of de facto political power, hence:

$$\text{distribution of resources}_t \Longrightarrow \text{de facto political power}_t$$

6. This brings us to the evolution of one of the two main *state variables* in our framework, political institutions (the other state variable is the distribution of resources, including distribution of physical and human capital stocks etc.). Political institutions and the distribution of resources are the state variables in this dynamic system because they typically change relatively slowly, and more importantly, they determine economic institutions and economic performance both directly and indirectly. Their direct effect is straightforward to

understand. If political institutions place all political power in the hands of a single individual or a small group, economic institutions that provide protection of property rights and equal opportunity for the rest of the population are difficult to sustain. The indirect effect works through the channels discussed above: political institutions determine the distribution of de jure political power, which in turn affects the choice of economic institutions. This framework therefore introduces a natural concept of *a hierarchy of institutions*, with political institutions influencing equilibrium economic institutions, which then determine economic outcomes.

Political institutions, though slow changing, are also endogenous. Societies transition from dictatorship to democracy, and change their constitutions to modify the constraints on power holders. Since, like economic institutions, political institutions are collective choices, the distribution of political power in society is the key determinant of their evolution. This creates a tendency for persistence: political institutions allocate de jure political power, and those who hold political power influence the evolution of political institutions, and they will generally opt to maintain the political institutions that give them political power. However, de facto political power occasionally creates changes in political institutions. While these changes are sometimes discontinuous, for example when an imbalance of power leads to a revolution or the threat of revolution leads to major reforms in political institutions, often they simply influence the way existing political institutions function, for example, whether the rules laid down in a particular constitution are respected as in most functioning democracies, or ignored as in current-day Zimbabwe. Summarizing this discussion, we have:

$$\text{political power}_t \Longrightarrow \text{political institutions}_{t+1}$$

Putting all these pieces together, gives the above representation.

To see whether this framework is useful, let us consider a brief example.

Consider the development of property rights in Europe during the Middle Ages. There is no doubt that lack of property rights for landowners, merchants and proto- industrialists was detrimental to economic growth during this epoch. Since political institutions at the time placed political power in the hands of kings and various types of hereditary monarchies, such rights were largely decided by these monarchs. Unfortunately for economic growth, while monarchs had every incentive to protect their own property rights, they did not generally enforce the property rights of others. On the contrary, monarchs often used their powers to expropriate producers, impose arbitrary taxation, renege on their debts, and allocate the productive resources of society to their allies in return for economic benefits or political support. Consequently, economic institutions during the Middle Ages provided little incentive to invest in land, physical or human capital, or technology, and failed to foster economic growth. These economic institutions also ensured that the monarchs controlled a large fraction of the economic resources in society, solidifying their political power and ensuring the continuation of the political regime.

The seventeenth century, however, witnessed major changes in the economic and political institutions that paved the way for the development of property rights and limits on monarchs' power, especially in England after the Civil War of 1642 and the Glorious Revolution of 1688, and in the Netherlands after the Dutch Revolt against the Hapsburgs. How did these major institutional changes take place? In England, for example, until the sixteenth century the king also possessed a substantial amount of de facto political power, and leaving aside civil wars related to royal succession, no other social group could amass sufficient de facto political power to challenge the king. But changes in the English land market and the expansion of Atlantic trade in the sixteenth and seventeenth centuries gradually increased the economic fortunes, and consequently the de facto power of landowners and merchants. These groups were diverse, but contained important elements that perceived themselves as having interests in conflict with those of the king: while the English kings were interested in predating against society to increase their tax incomes, the gentry and merchants were interested in strengthening their property rights.

By the seventeenth century, the growing prosperity of the merchants and the gentry, based both on internal and overseas, especially Atlantic, trade, enabled them to field military forces capable of defeating the king. This de facto power overcame the Stuart monarchs in the Civil War and Glorious Revolution, and led to a change in political institutions that stripped the king of much of his previous power over policy. These changes in the distribution of political power led to major changes in economic institutions, strengthening the property rights of both land and capital owners and spurred a process of financial and commercial expansion. The consequence was rapid economic growth, culminating in the Industrial Revolution, and a very different distribution of economic resources from that in the Middle Ages.

It is worth returning at this point to two critical assumptions in our framework. First, why do the groups with conflicting interests not agree on the set of economic institutions that maximize aggregate growth? So in the case of the conflict between the monarchy and the merchants, why does the monarchy not set up secure property rights to encourage economic growth and tax some of the benefits? Second, why do groups with political power want to change political institutions in their favor? For instance, in the context of the example above, why did the gentry and merchants use their de facto political power to change political institutions rather than simply implement the policies they wanted? The answers to both questions revolve around issues of *commitment* and go to the heart of our framework.

The distribution of resources in society is an inherently conflictual, and therefore political, decision. As mentioned above, this leads to major commitment problems, since groups with political power cannot commit to not using their power to change the distribution of resources in their favor. For example, economic institutions that increased the security of property rights for land and capital owners during the Middle Ages would not have been credible as long as the monarch monopolized political power. He could promise to respect property rights, but then at some point, renege on his promise, as exemplified by the numerous financial defaults by medieval kings. Credible secure property rights necessitated a reduction in the political power of the monarch. Although these more secure property rights would foster economic growth, they were not appealing to the monarchs who would lose their rents from predation and expropriation as well as various other privileges associated with their monopoly of political power. This is why the institutional changes in England as a result of the Glorious Revolution were not simply conceded by the Stuart kings. James II had to be deposed for the changes to take place.

The reason why political power is often used to change political institutions is related. In a dynamic world, individuals care not only about economic outcomes today but also in the future. In the example above, the gentry and merchants were interested in their profits and therefore in the security of their property rights, not only in the present but also in the future. Therefore, they would have liked to use their (de facto) political power to secure benefits in the future as well as the present. However, commitment to future allocations (or economic institutions) was not possible because decisions in the future would be decided by those who had political power in the future with little reference to past promises. If the gentry and merchants would have been sure to maintain their de facto political power, this would not have been a problem. However, de facto political power is often transient, for example because the collective action problems that are solved to amass this power are likely to resurface in the future, or other groups, especially those controlling de jure power, can become stronger in the future. Therefore, any change in policies and economic institutions that relies purely on de facto political power is likely to be reversed in the future. In addition, many revolutions are followed by conflict within the revolutionaries. Recognizing this, the English gentry and merchants strove not just to change economic institutions in their favor following their victories against the Stuart monarchy, but also to alter political institutions and the future allocation of de jure power. Using political power to change political institutions then emerges as a useful strategy to make gains more durable. The framework that we propose, therefore, emphasizes the importance of political institutions, and changes in political institutions, as a way of manipulating future political power, and thus indirectly shaping future, as well as present, economic institutions and outcomes.

This framework, though abstract and highly simple, enables us to provide some preliminary answers to our main question: why do some societies choose "good economic institutions"? At this point, we need to be more specific about what good economic institutions are. A danger we would like to avoid is that we define good economic institutions as those that generate economic growth, potentially leading to a tautology. This danger arises because a given set of economic institutions may be relatively good during some periods and bad during others. For example, a set of economic institutions that protects the property rights of a small elite might not be inimical to economic growth when all major investment opportunities are in the hands of this elite, but could be very harmful when investments and participation by other groups are important for economic growth (see Acemoglu, 2003b). To avoid such a tautology and to simplify and focus the discussion, throughout we think of good economic institutions as those that provide security of property rights and relatively equal access to economic resources to a broad cross-section of society. Although this definition is far from requiring equality of opportunity in society, it implies that societies where only a very small fraction of the population have well-enforced property rights do not have good economic institutions. Consequently, as we will see in some of the historical cases discussed below, a given set of economic institutions may have very different implications for economic growth depending on the technological possibilities and opportunities.

1.5. References

(1) Acemoglu, Daron (2003a) "Why Not a Political Coase Theorem?" *Journal of Comparative Economics*, 31, 620-652.

(2) Acemoglu, Daron (2003b) "The Form of Property Rights: Oligarchic versus Democratic Societies," NBER Working Paper #10037.

(3) Acemoglu, Daron, Simon Johnson and James Robinson (2005) "Institutions As the Fundamental Cause of Long-Run Growth" in Philippe Aghion and Steven Durlauf eds. *Handbook of Economic Growth*, Amsterdam; North-Holland.

(4) Acemoglu, Daron, James A. Robinson and Thierry Verdier (2004) "Kleptocracy and Divide-and-Rule: A Model of Personal Rule," The Alfred Marshall Lecture", *Journal of the European Economic Association Papers and Proceedings,* 2004, 162-192.

(5) Allen, Robert C. (1982) "The Efficiency and Distributional Consequences of Eighteenth Century Enclosures," *Economic Journal*, 92, 937-953.

(6) Allen, Robert C. (2003) *Farm to Factory: A Reinterpretation of the Soviet Industrial Revolution*, Princeton; Princeton University Press.

(7) Bates, Robert H. (1981) *Markets and States in Tropical Africa*, University of California Press, Berkeley CA.

(8) Becker, Gary S. (1983) "A Theory of Competition Among Pressure Groups for Political Influence," *Quarterly Journal of Economics*, 98, 371-400.

(9) Coase, Ronald H. (1960) "The Problem of Social Cost,"*Journal of Law and Economic*s, 3, 1-44.

(10) Demsetz, Harold (1967) "Toward a Theory of Property Rights," *American Economic Review*, 57, 61-70.

(11) Djankov, Simeon, Edward L. Glaeser, Florencio Lopez-de-Silanes, Rafael La Porta and Andrei Shleifer (2003) "The New Comparative Economics, *Journal of Comparative Economics*, 31, 595-619.

(12) Farrell, Joseph (1987) "Information and the Coase Theorem," *Journal of Economic Perspectives*, 1, 113-129.

(13) Firmin-Sellers, Kathryn (1995) "The Politics of Property Rights," *American Political Science Review*, 89, 867-881.

(14) Greif, Avner (1994) "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies," *Journal of Political Economy*, 102, 912-950.

(15) Hayek, Freidrich (1945) "The Use of Knowledge in Society," *American Economic Review*, 35, 519-530.

(16) Helmke, Gretchen (2005) *Courts under Constraints: Judges, Generals and Presidents in Argentina*, New York; Cambridge University Press.

(17) Herbst, Jeffery I. (2000) *States and Power in Africa: Comparative Lessons in Authority and Control,* Princeton University Press, Princeton NJ.

(18) Hicks, John R. (1969) *A Theory of Economic History*, New York; Oxford University Press.

(19) Iaryczower, Matias , Pablo T. Spiller and Mariano Tommasi (2002) "Judicial Decision Making in Unstable Environments: Argentina 1938-1998," *American Journal of Political Science*, 46, 699-716.

(20) Jackson, Matthew O. and Simon Wilkie (2005) "Endogenous Games and Mechanisms: Side Payments Among Players," *Review of Economic Studies*, 72, 543-566.

(21) Lange, Oskar (1936) "On the Economic Theory of Socialism: Part One," *Review of Economic Studies*, 4, 53-71.

(22) Lange, Oskar (1937) "On the Economic Theory of Socialism: Part Two," *Review of Economic Studies*, 5, 123-142.

(23) Nelson, Richard R. and Sidney G. Winter (1982) *An Evolutionary Theory of Economic Change*, Cambridge; Belknap Press.

(24) North, Douglass C. (1981) *Structure and Change in Economic History*, New York; W.W. Norton & Co.

(25) Tilly, Charles (1990) *Coercion, Capital and European States, AD 990-1990*, Blackwell, Cambridge MA.

(26) Tornell, Aaron (1997) "Economic Growth and Decline with Endogenous Property Rights," *Journal of Economic Growth*, 2, 219-50.

(27) Wrightson, Keith (2000) *Earthly Necessities: Economic Lives in early Modern Britain*, New Haven; Yale University Press.

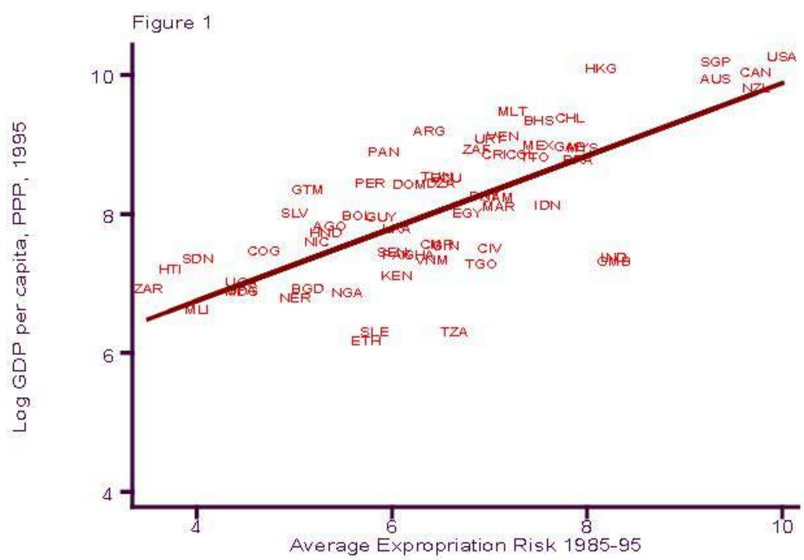
(28) Young, H. Peyton (1998) *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*, Princeton; Princeton University Press.

CHAPTER 2

Evidence

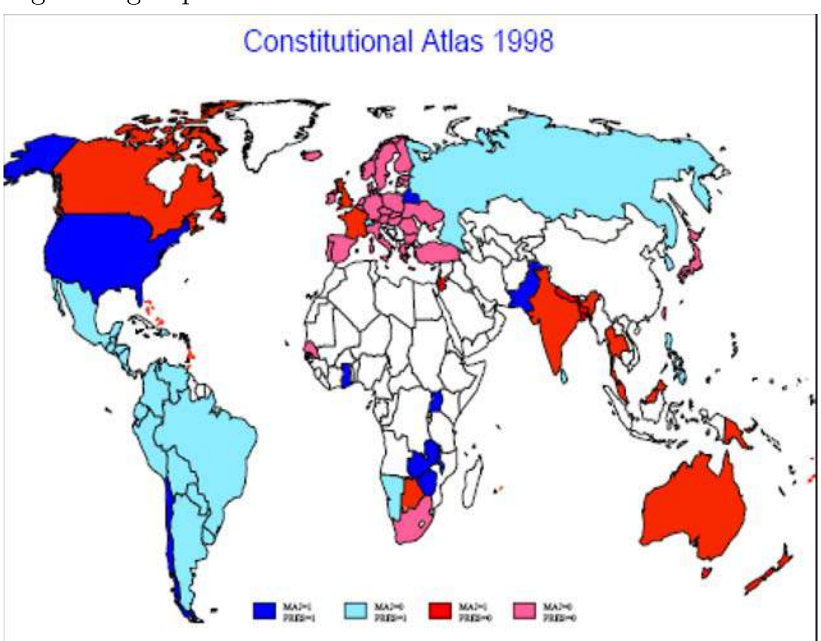
2.1. Aggregate Correlations

There is now a large literature documenting a positive correlation between measures of institutions and good governance on the one hand, and economic performance on the other. One of the earliest is the paper by Knack and Keefer (1995). They use measures of property rights, and find them to be strongly correlated with investment and growth (even after controlling for other potential determinants of growth). The next figure, for example, shows the relationship between income per capita and one measure of property rights enforcement.



Other authors have found similar relationships using political instability, corruption, and measures of rule of law. These variables are all institutional outcomes. For example, how secure property rights are in equilibrium. One may also be interested in the effects of more specific formal institutions. For instance, a natural idea could be that whether or not private property existed was important. Many African countries still have most land held communally with the potential for a tragedy of the commons. Nevertheless, it is hard to get at this question using macro data (as we'll see micro data suggests that establishing individual property rights can be important). Prominent in the literature examining the effects of formal institutions has been the work of Persson and Tabellini (2003, 2004) who have shown how various aspects of formal democratic constitutions, for example whether or not the system is a presidential one or a parliamentary one, or whether or not legislators are elected using proportional representation, seems to matter for the level and composition of government spending. Their recent research suggests that these types of institutional distinctions might matter for economic growth as well (Persson and Tabellini, 2005). There is less consensus about whether the presence or absence of democracy itself matters for economic outcomes (Barro, 1997) or policy outcomes (Mulligan, Gill and Sala-i-Martin, 2004). Another prominent set of papers based on differences in formal institutions is the work by Andrei Shleifer and his co-authors on how many differences in institutions (such as shareholder protection or barriers to entry) can be traced to the legal origin of the country.

Whether these types of institutional distinctions matter for policy and economic outcomes is important, partly because there is a large amount of variation in these institutions. The next figure gives a glimpse of this.



It is useful to look at some of the estimates of Persson and Tabellini. They investigate the effect of electoral rules and types of political institutions on policy outcomes in a panel of

61 democracies. They find, for instance, that in presidential regimes, the size of government is smaller than in parliamentary regimes. They also find that majoritarian elections lead to smaller transfers than proportional representation systems. The enclosed three tables (from their American Economic Review paper) give some of their results. The first just shows the distribution of different political institutions and the countries that are covered (MAJ signifies majoritarian electoral system as opposed to a Proportional Representation system, PREs is whether or not the system has a president).

Table 4 Composition of government and constitutions: OLS estimates

	(1)	(2)	(3)	(4)	(5)
Dep. var.	*ssw*	*ssw*	*ssw*	*ssw*	*ssw*
pres	-2.24		-0.25	-5.47	-4.28
	(1.11)**		(2.06)	(1.19)***	(1.30)***
maj	-2.25		-1.02	-2.66	-3.03
	(1.25)*		(1.36)	(1.52)*	(1.50)**
propres		-3.22			
		(1.74)*			
majpar		-3.14			
		(2.18)			
majpres		-3.91			
		(2.41)			
pres_newdem				4.97	
				(1.65)***	
maj_newdem				1.74	
				(1.77)	
newdem				-5.36	
				(1.69)***	
pres_baddem					5.61
					(2.00)***
maj_baddem					3.67
					(1.62)**
baddem					-4.24
					(1.75)**
F-test (*pres*)		0.83		0.17	0.83
F-test (*maj*)				0.65	0.19
Sample	90s	90s	72-77	90s	90s
Obs.	69	69	42	69	69
R2	0.81	0.81	0.77	0.84	0.82

These results are very suggestive, but leave the question of whether it is the political institution, or underlying conditions that lead to the establishment and maintenance of these political institutions, that are causing these results.

The problem is that these correlations do not establish that institutions have a causal effect on economic performance. First, there could be the standard type of reverse causality.

Moreover, we could simply be observing the fact that countries with different economic environments are choosing different institutions. Here is an interesting fact: all British colonies in Africa had parliamentary institutions set up by the British at the time of independence. With the exception of Lesotho, they all subsequently switched to presidential systems.

There is an omitted-variables problem. Economies that are different for a variety of reasons will differ both in their institutions and in their income per capita, and since it is impossible to control for these differences in practice, we may be assigning the effect of these omitted variables to institutional differences, greatly exaggerating the effect of institutions of economic performance.

Finally, there another version of the reverse causality problem: which arises because most measures of institutions are "subjective", and perhaps scholars see better institutions in places that perform better.

2.2. "Exogenous" Differences in Institutions

To solve this identification problem, we need to find exogenous differences in institutions. In practice, of course, truly exogenous variation does not exist, so we have to find the source of variation that is plausibly orthogonal to other determinants of current economic performance.

There are now several attempts to find such exogenous differences in institutions (or to use an IV strategy) in the literature.

(1) Mauro's (1995) work on corruption, where he uses ethnolinguistic fragmentation as an instrument for corruption. The reasoning here is that ethnolinguistic fragmentation will make it harder for principals to control agents, hence facilitate corruption.
(2) Hall and Jones' (1999) work which uses distance from the equator and the fraction of the population speaking English as instruments for a measure of institutions (which they call social infrastructure). The reasoning is that these variables proxy for the strength of the "good" European/British influence on a country's culture and institutions.
(3) Acemoglu, Johnson and Robinson (2001, extended in 2002) who use mortality rates faced by potential settlers at the time of colonization as an instrument for institutional development. The argument is that in places where the Europeans did not settle because of high mortality, they introduced worse institutions than in places where they settled.

(4) Persson and Tabellini (2003, 2004) have recently used various instruments for the form of democratic institutions. These include fraction of the population speaking one of the major European languages and latitude.

The econometric argument underlying Mauro's and Hall and Jones' instruments is not entirely convincing. In both cases, the instrument can have a direct affect. For example, ethnolinguistic fragmentation can arguably affect economic performance by creating political instability, while many authors think that there is a direct effect of climate and geography on performance. Moreover, the theoretical reasoning for the instrument of Hall and Jones is not strong. It is not easy to argue that the Belgian influence in the Congo, or Western influence in the Gold Coast during the era of slavery promoted good institutions or governance.

The problem with the instruments underlying the work and Persson and Tabellini (2003) may be different (see the extensive discussion in Acemoglu, 2005). They motivate the instruments on the basis of the arguments proposed in Hall and Jones, yet these ideas apply only to the former colonial world, not the entire world. Probably even more problematically, even to the extent that one can tell a story linking latitude to the form of the constitution, it seems highly unlikely that the form of the constitutions will be the *only* way that latitude will influence the size of government, for instance. If there are other channels of influence which we cannot control for then the estimate of the impact of the form of the constitution on the size of government will be biased.

Not surprisingly, we find the reasoning for the Acemoglu, Johnson Robinson paper more compelling. Here, the theory goes as follows:

(1) There were different types of colonization policies which created different sets of institutions. At one extreme, as in the Belgian colonization of the Congo, European powers set up "extractive states". These institutions did not introduce much protection for private property, nor did they provide checks and balances against government expropriation. In fact, the main purpose of the extractive state was to transfer as much of the resources of the colony to the colonizer. At the other extreme, many Europeans went and settled in a number of colonies, and tried to replicate European institutions, with great emphasis on private property, and checks against government power. Primary examples of this include Australia, New Zealand, Canada, and the United States.

(2) The colonization strategy was influenced by the feasibility of settlements. In places where the disease environment was not favorable to European settlement, the formation of the extractive state was more likely.
(3) The colonial state and institutions persisted even after independence.

These premises suggest that exogenous variation in whether Europeans could settle or not would be a good instrument for institutional development in the colonies, and hence a good instrument for current institutions. Acemoglu, Johnson and Robinson use mortality rates faced by potential settlers at the time of colonizations as an instrument for settlements and institutional development. Schematically:

(potential) settler mortality ⇒ settlements ⇒ early institutions ⇒ current institutions ⇒ current performance

The enclosed tables and figures give the details of the estimates from Acemoglu, Johnson and Robinson (2001). We start with the OLS estimates for comparisons, using the same measure of property rights enforcement, average protection against expropriation, shown in the above figure.

Table 2
OLS Regressions

	Whole World (1)	Base Sample (2)	Whole World (3)	Whole World (4)	Base Sample (5)	Base Sample (6)	Whole World (7)	Base Sample (8)
	Dependent Variable is log GDP per capita in 1995						Dep. Var. is log output per worker in 1988	
Average Protection Against Expropriation Risk, 1985-1995	0.54 (0.04)	0.52 (0.06)	0.47 (0.06)	0.43 (0.05)	0.47 (0.06)	0.41 (0.06)	0.45 (0.04)	0.46 (0.06)
Latitude			0.89 (0.49)	0.37 (0.51)	1.60 (0.70)	0.92 (0.63)		
Asia Dummy				-0.62 (0.19)		-0.60 (0.23)		
Africa Dummy				-1.00 (0.15)		-0.90 (0.17)		
"Other" Continent Dummy				-0.25 (0.20)		-0.04 (0.32)		
R-Squared	0.62	0.54	0.63	0.73	0.56	0.69	0.55	0.49
N	110	64	110	110	64	64	108	61

We also show the relationship between settler mortality and European settlements and early institutions and the persistence of early institutions in the next table.

Table 3
Determinants of Institutions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A	*Dependent Variable is Average Protection against Expropriation Risk in 1985-95*									
Constraint on Executive in 1900	0.32	0.26								
	(0.08)	(0.09)								
Democracy in 1900			0.24	0.21						
			(0.06)	(0.07)						
Constraint on Executive in First Year of Independence					0.25	0.22				
					(0.08)	(0.08)				
European Settlements in 1900							3.20	3.00		
							(0.61)	(0.78)		
Log European Settler Mortality									-0.61	-0.51
									(0.13)	(0.14)
Latitude		2.20		1.60		2.70		0.58		2.00
		(1.40)		(1.50)		(1.40)		(1.51)		(1.34)
R-Squared	0.2	0.23	0.24	0.25	0.19	0.24	0.3	0.3	0.27	0.3
Number of Observations	63	63	62	62	63	63	66	66	64	64
Panel B	*Dependent variable is Constraint on Executive in 1900*				*Dependent variable is Democracy in 1900*				*Dependent variable is European Settlements in 1900*	
European Settlements in 1900	5.50	5.40			8.60	8.10				
	(0.73)	(0.93)			(0.90)	(1.20)				
Log European Settler Mortality			-0.82	-0.65			-1.22	-0.88	-0.11	-0.07
			(0.17)	(0.18)			(0.24)	(0.25)	(0.02)	(0.02)
Latitude		0.33		3.60		1.60		7.60		0.87
		(1.80)		(1.70)		(2.30)		(2.40)		(0.19)
R-Squared	0.46	0.46	0.25	0.29	0.57	0.57	0.28	0.37	0.31	0.47
Number of Observations	70	70	75	75	67	67	68	68	73	73

The following table shows the "first stages" and the "two-stage least squares" estimates based on this identification strategy.

Table 4
IV Regressions of log GDP per capita

	Base Sample (1)	Base Sample (2)	Base Sample without neo-Europes (3)	Base Sample without neo-Europes (4)	Base Sample without Africa (5)	Base Sample without Africa (6)	Base Sample with Continent Dummies (7)	Base Sample with Continent Dummies (8)	Base Sample, dep. var. is log output per worker (9)
Panel A: Two Stage Least Squares									
Average Protection Against Expropriation Risk 1985-1995	0.94	1.00	1.28	1.21	0.58	0.58	0.98	1.10	0.98
	(0.16)	(0.22)	(0.36)	(0.35)	(0.10)	(0.12)	(0.30)	(0.46)	(0.17)
Latitude		-0.65		0.94		0.04		-1.20	
		(1.34)		(1.46)		(0.84)		(1.8)	
Asia Dummy							-0.92	-1.10	
							(0.40)	(0.52)	
Africa Dummy							-0.46	-0.44	
							(0.36)	(0.42)	
"Other" Continent Dummy							-0.94	-0.99	
							(0.85)	(1.0)	
Panel B: First-Stage for Average Protection against Expropriation Risk in 1985-95									
Log European Settler Mortality	-0.61	-0.51	-0.39	-0.39	-1.20	-1.10	-0.43	-0.34	-0.63
	(0.13)	(0.14)	(0.13)	(0.14)	(0.22)	(0.24)	(0.17)	(0.18)	(0.13)
Latitude		2.00		-0.11		0.99		2.00	
		(1.34)		(1.50)		(1.43)		(1.40)	
Asia Dummy							0.33	0.47	
							(0.49)	(0.50)	
Africa Dummy							-0.27	-0.26	
							(0.41)	(0.41)	
"Other" Continent Dummy							1.24	1.1	
							(0.84)	(0.84)	
R-Squared	0.27	0.30	0.13	0.13	0.47	0.47	0.30	0.33	0.28

The next three tables show the robustness of these results to variables that are potentially threatening to the identification strategy, such as murder geographic controls and controls for current health conditions.

Table 5
IV Regressions of log GDP per capita with Additional Controls

	Base Sample (1)	Base Sample (2)	British colonies only (3)	British colonies only (4)	Base Sample (5)	Base Sample (6)	Base Sample (7)	Base Sample (8)	Base Sample (9)
				Panel A: Two Stage Least Squares					
Average Protection Against Expropriation Risk, 1985-1995	1.10	1.16	1.07	1.00	1.10	1.20	0.92	1.00	1.10
	(0.22)	(0.34)	(0.24)	(0.22)	(0.19)	(0.29)	(0.15)	(0.25)	(0.29)
Latitude		-0.75				-1.10		-0.94	-1.70
		(1.70)				(1.56)		(1.50)	(1.6)
British Colonial Dummy	-0.78	-0.80							
	(0.35)	(0.39)							
French Colonial Dummy	-0.12	-0.06							0.02
	(0.35)	(0.42)							(0.69)
French legal origin dummy					0.89	0.96			0.51
					(0.32)	(0.39)			(0.69)
p-value for Religion Variables							[0.001]	[0.004]	[0.42]
				Panel B: First-Stage for Average Protection against Expropriation Risk in 1985-95					
Log European Settler Mortality	-0.53	-0.43	-0.59	-0.51	-0.54	-0.44	-0.58	-0.44	-0.48
	(0.14)	(0.16)	(0.19)	(0.14)	(0.13)	(0.14)	(0.13)	(0.15)	(0.18)
Latitude		1.97				2.10		2.50	2.30
		(1.40)				(1.30)		(1.50)	(1.60)
British Colonial Dummy	0.63	0.55							
	(0.37)	(0.37)							
French Colonial Dummy	0.05	-0.12							-0.25
	(0.43)	(0.44)							(0.89)
French legal origin					-0.67	-0.7			-0.05
					(0.33)	(0.32)			(0.91)
R-Squared	0.31	0.33	0.30	0.30	0.32	0.35	0.32	0.35	0.45
				Panel C: Ordinary Least Squares					
Average Protection Against Expropriation Risk, 1985-1995	0.53	0.47	0.61	0.47	0.56	0.56	0.53	0.47	0.47
	(0.19)	(0.07)	(0.09)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Number of Observations	64	64	25	25	64	64	64	64	64

Table 6
Robustness Checks for IV Regressions of Log GDP per capita

	Base Sample (1)	Base Sample (2)	Base Sample (3)	Base Sample (4)	Base Sample (5)	Base Sample (6)	Base Sample (7)	Base Sample (8)	Base Sample (9)
Panel A: Two Stage Least Squares									
Average Protection Against Expropriation Risk, 1985-1995	0.84 (0.19)	0.83 (0.21)	0.96 (0.28)	0.99 (0.30)	1.10 (0.33)	1.30 (0.51)	0.74 (0.13)	0.79 (0.17)	0.71 (0.20)
Latitude		0.07 (1.60)		-0.67 (1.30)		-1.30 (2.30)		-0.89 (1.00)	-2.5 (1.60)
p-value for Temperature Variables	[0.96]	[0.97]							[0.77]
p-value for Humidity Variables	[0.54]	[0.54]							[0.62]
Percent of European descent in 1975			-0.08 (0.82)	0.03 (0.84)					0.3 (0.7)
P-Value for Soil Quality					[0.79]	[0.85]			[0.46]
P-Value for Natural Resources					[0.82]	[0.87]			[0.82]
Dummy for being landlocked					0.64 (0.63)	0.79 (0.83)			0.75 (0.47)
Ethnolinguistic fragmentation							-1.00 (0.32)	-1.10 (0.34)	-1.60 (0.47)
Panel B: First-Stage for Average Protection against Expropriation Risk in 1985-95									
Log European Settler Mortality	-0.64 (0.17)	-0.59 (0.17)	-0.41 (0.14)	-0.4 (0.15)	-0.44 (0.16)	-0.34 (0.17)	-0.64 (0.15)	-0.56 (0.15)	-0.59 (0.21)
Latitude		2.70 (2.00)		0.48 (1.50)		2.20 (1.50)		2.30 (1.40)	4.20 (2.60)
R-Squared	0.39	0.41	0.34	0.34	0.41	0.43	0.27	0.30	0.59
Panel C: Ordinary Least Squares									
Average Protection Against Expropriation Risk, 1985-1995	0.41 (0.06)	0.38 (0.06)	0.39 (0.06)	0.38 (0.06)	0.46 (0.07)	0.42 (0.07)	0.46 (0.05)	0.45 (0.06)	0.38 (0.06)

Table 7
Geography and Health Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Instrumenting only for Average Protection Against Expropriation Risk						Instrumenting for all Right-Hand Side Variables			Instrument only Average Protect. Against Exprop. Risk	
	Panel A: Two Stage Least Squares										
Average Protection Against Expropriation Risk, 1985-1995	0.69	0.72	0.63	0.68	0.55	0.56	0.69	0.74	0.68	0.91	0.90
	(0.25)	(0.30)	(0.28)	(0.34)	(0.24)	(0.31)	(0.26)	(0.24)	(0.23)	(0.24)	(0.32)
Latitude		-0.57		-0.53		-0.1					
		(1.04)		(0.97)		(0.95)					
Malaria in 1994	-0.57	-0.60					-0.62				
	(0.47)	(0.47)					(0.68)				
Life Expectancy			0.03	0.03				0.02			
			(0.02)	(0.02)				(0.02)			
Infant Mortality					-0.01	-0.01			-0.01		
					(0.005)	(0.006)			(0.01)		
	Panel B: First-Stage for Average Protection against Expropriation Risk in 1985-95										
Log European Settler Mortality	-0.42	-0.38	-0.34	-0.30	-0.36	-0.29	-0.41	-0.40	-0.40		
	(0.19)	(0.19)	(0.17)	(0.18)	(0.18)	(0.19)	(0.17)	(0.17)	(0.17)		
Latitude		1.70		1.10		1.60	-0.81	-0.84	-0.84		
		(1.40)		(1.40)		(1.40)	(1.80)	(1.80)	(1.80)		
Malaria in 1994	-0.79	-0.65									
	(0.54)	(0.55)									
Life Expectancy			0.05	0.04							
			(0.02)	(0.02)							
Infant Mortality					-0.01	-0.01					
					(0.01)	(0.01)					
Mean Temperature							-0.12	-0.12	-0.12		
							(0.05)	(0.05)	(0.05)		
Distance from Coast							0.57	0.55	0.55		
							(0.51)	(0.52)	(0.52)		
Yellow Fever Dummy										-1.10	-0.81
										(0.41)	(0.38)
R-Squared	0.3	0.31	0.34	0.35	0.32	0.34	0.37	0.36	0.36	0.10	0.32
	Panel C: Ordinary Least Squares										
Average Protection Against Expropriation Risk, 1985-1995	0.35	0.35	0.28	0.28	0.29	0.28	0.35	0.29	0.29	0.48	0.39
	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)
Number of Observations	62	62	60	60	60	60	60	59	59	64	64

We also look briefly at some of the IV estimates of Persson and Tabellini, which confirmed their OLS estimates, though much of the identification is based on an instrumental variable strategy that is less clearly motivated and must rely on functional form assumptions.

Table 3 Size of government and constitutions: Instrumental-variable, Heckman and Matching Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. var.	*cgexp*	*cgexp*	*cgexp*	*cgexp*	*cgexp*	*cgexp*	*cgexp*
pres	-5.29	-11.52	-6.51	-4.22	-5.89	-3.23	-7.45
	(2.18)**	(4.54)**	(3.71)*	(3..99)	(3.02)*	(2.74)	(2.34)***
maj	-6.21	-6.77	-4.83	-4.18	-4.81	-5.34	-5.59
	(2.82)**	(1.98)***	(3.19)	(3.17)	(3.41)	(2.73)*	(2.61)**
Conts & Cols	Yes	Yes	*col_uka*	*col_uka, laam*			
Sample	90s	90s	90s	90s	90s	90s	90s
Endogenous selection	*maj*	*pres*	*pres* *maj*	*pres* *maj*	*pres* *maj*	*pres* *maj*	*pres* *maj*
Estimation	Heckman ML	Heckman ML	2SLS	2SLS	Stratification	Nearest neighbor	Kernel
Rho	0.05	0.62					
	(0.29)	(0.33)					
Chi-2			3.29	2.23			
Adj. R2			0.59	0.59			
Obs.	75	75	75	75	66(*pres*) 70(*maj*)	66(*pres*) 70(*maj*)	66(*pres*) 70(*maj*)

Table 5 Composition of government and constitutions: Instrumental variables, Heckman and Matching Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. var.	*ssw*	*ssw*	*ssw*	*ssw*	*ssw*	*ssw*	*ssw*
pres	0.20	-2.38*	0.75	0.49	-3.06	-2.28	-3.79
	(3.27)	(1.33)	(2.00)	(2.14)	(2.67)	(1.79)	(2.36)
maj	-2.05*	-4.27	-3.21	-3.21	-1.85	-1.90	-3.46
	(1.12)	(1.79)**	(1.61)*	(1.62)*	(1.91)	(1.67)	(1.84)*
Conts & Cols	Yes	Yes	*col_uka*	*col_uka laam*			
Sample	90s	90s	90s	90s	90s	90s	90s
Endogenous Selection	*pres*	*maj*	*pres* *maj*	*pres* *maj*	*pres* *maj*	*pres* *maj*	*pres* *maj*
Estimation	Heckman 2-step	Heckman 2-step	2SLS	2SLS	Stratification	Nearest neighbor	Kernel
Rho	-0.46	0.59					
Chi-2			9.53*	9.98*			
Adj. R2			0.78	0.78			
Obs.	64	64	64	64	64(*pres*) 70(*maj*)	64(*pres*) 70(*maj*)	64(*pres*) 70(*maj*)

2.3. A Sharper Natural Experiment

Perhaps the most extreme natural experiment useful to think of the effect of a broad cluster of institutions is the separation of Korea between North and South Korea.

Until the end of World War II, Korea was under Japanese occupation. Korean independence came shortly after the Japanese Emperor Hirohito announced the Japanese surrender on August 15, 1945.

After this date, Soviet forces entered Manchuria and North Korea and took over the control of these provinces from the Japanese under the leadership of Kim Il Sung.

The U.S. supported Syngman Rhee, who was in favor of separation rather than a united communist Korea. Elections in the South were held in May 1948, amidst a widespread boycott by Koreans opposed to separation.

The newly elected representatives proceeded to draft a new constitution and established the Republic of Korea to the south of the 38th parallel. The North became the Democratic People's Republic of Korea, under the control of Kim Il Sung.

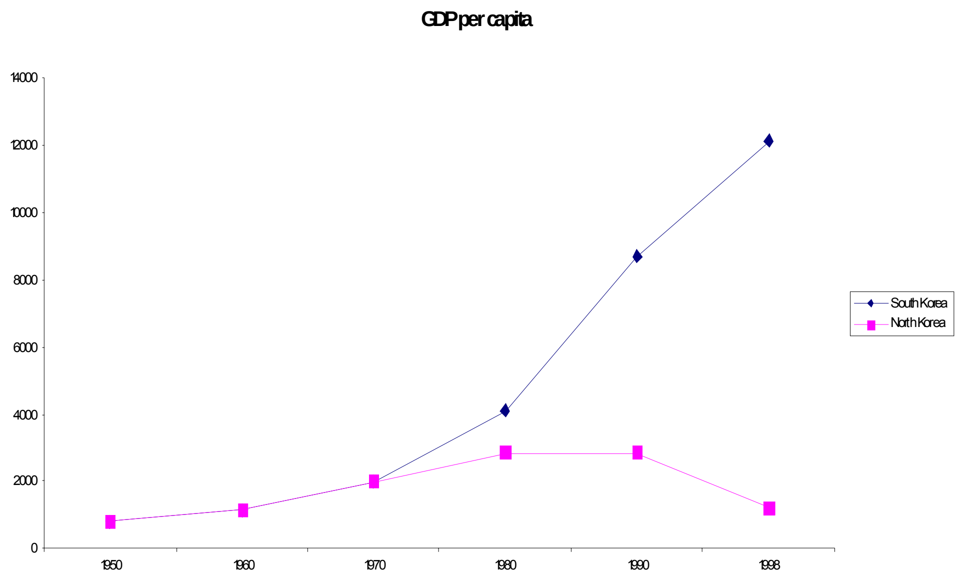
These two independent countries organized themselves in very different ways and adopted completely different sets of institutions. The North followed the model of Soviet socialism and the Chinese Revolution in abolishing private property of land and capital. Economic decisions were not mediated by the market, but by the communist state. The South instead maintained a system of private property and the government, especially after the rise to power of Park Chung Hee in 1961, attempted to use markets and private incentives in order to develop the economy.

Before this "natural experiment" in institutional change, North and South Korea shared the same history and cultural roots. In fact, Korea exhibited an unparalleled degree of ethnic, linguistic, cultural, geographic and economic homogeneity. There are few geographic distinctions between the North and South, and both share the same disease environment.

We can therefore think of the splitting on the Koreas 50 years ago as a natural experiment that we can use to identify the causal influence of a particular dimension of institutions on prosperity. Korea was split into two, with the two halves organized in radically different ways, and with geography, culture and many other potential determinants of economic prosperity held fixed. Thus any differences in economic performance can plausibly be attributed to differences in institutions.

Consistent with the hypothesis that it is institutional differences that drive comparative development, since separation, the two Koreas have experienced dramatically diverging paths of economic development:

By the late 1960's South Korea was transformed into one of the Asian "miracle" economies, experiencing one of the most rapid surges of economic prosperity in history while North Korea stagnated. By 2000 the level of income in South Korea was $16,100 while in North Korea it was only $1,000. By 2000 the South had become a member of the Organization of Economic Cooperation and Development, the rich nations club, while the North had a level of per-capita income about the same as a typical sub-Saharan African country. The next figure showsthis evolution using data from Madison.



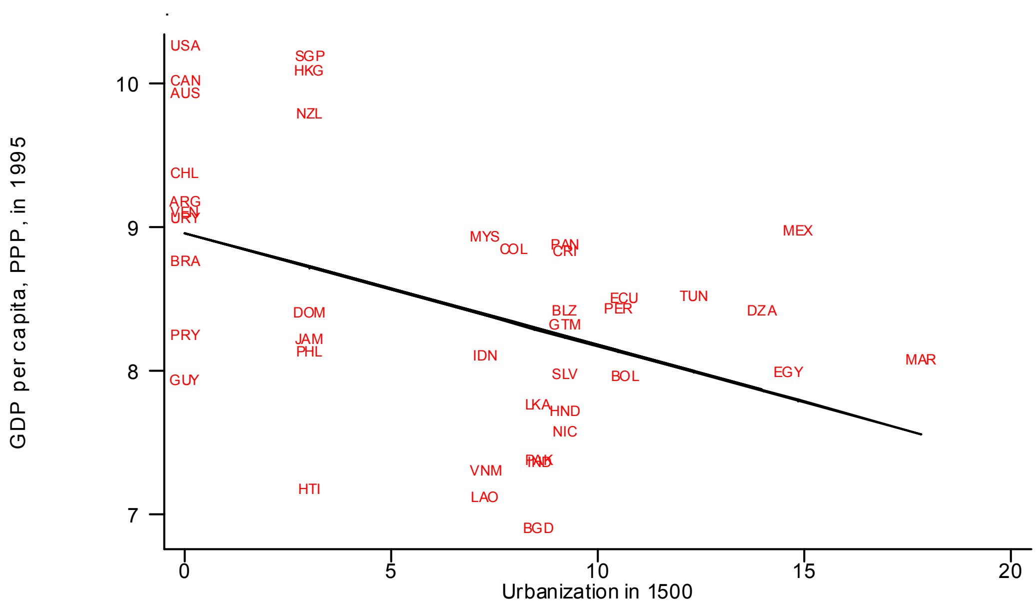
There is only one plausible explanation for the radically different economic experiences on the two Koreas after 1950: their very different institutions led to divergent economic outcomes. In this context, it is noteworthy that the two Koreas not only shared the same geography, but also the same culture.

We should note here however that while this is a good natural experiment for some questions, it won't help on others. For example, it won't tell us if particular aspects of the institutions of South Korea (greater financial depth) were the key (Acemoglu, 2005).

2.4. Reversal of Fortune

The pattern of evolution of prosperity among the former colonies also paints a picture similar to that of North vs. South Korea. Even leaving aside the issue of settler mortality, there is a remarkable pattern to how prosperity is changed in much of the globe after European colonization. The main pattern is dubbed "the reversal of fortune" by Acemoglu, Johnson and Robinson (2002) because places that were previously prosperous became relatively less prosperous after European colonization, and those that were non-urbanites, empty and less prosperous became relatively more prosperous. Throughout, since there are no national income accounts in 15th are 16th centuries anywhere in the world, we use proxies for prosperity, in particular urbanization rates (fraction of the population living in cities with 5000 inhabitants or more) and the population density. Acemoglu, Johnson and Robinson (2002) show that these are good proxies for pre-industrial prosperity using a variety of methods.

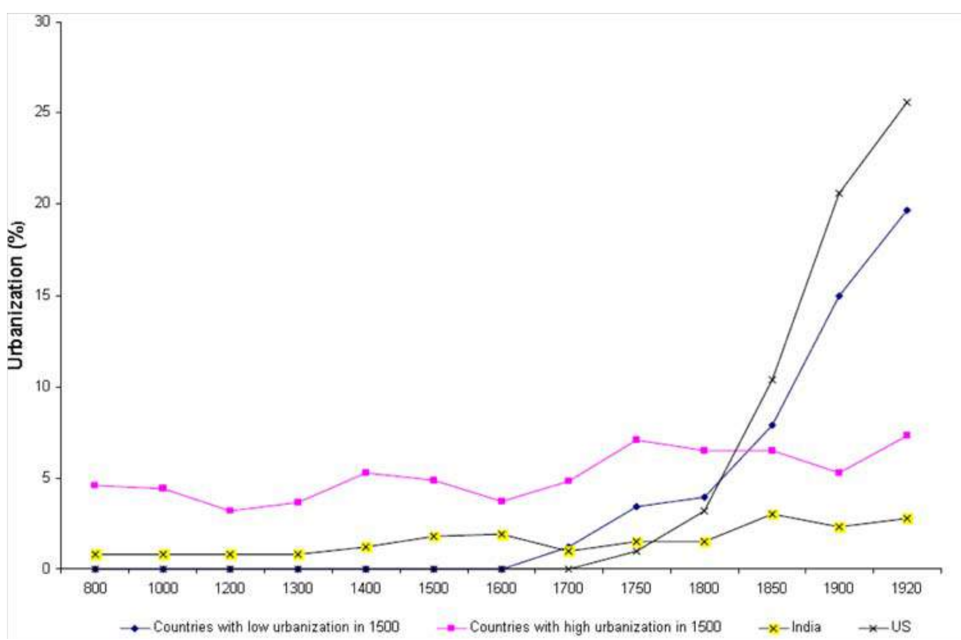
The next figure uses urbanization rates in 1500.

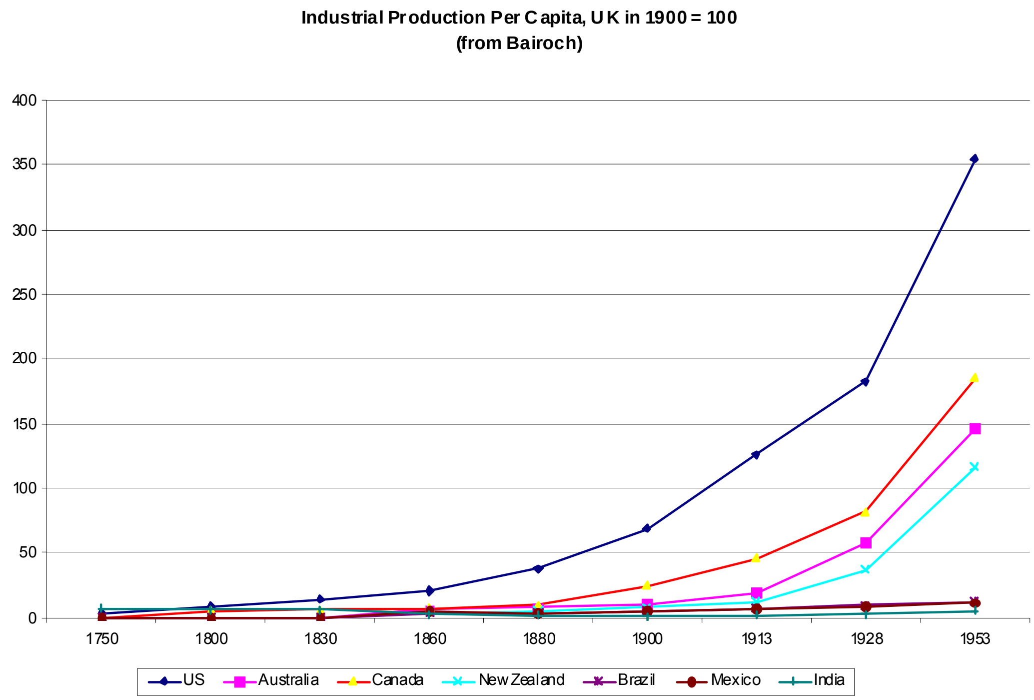


The pattern of reversal of fortune is clearly visible.

The next figure uses population density in 1500 and thus includes African countries, for which there are no reliable urbanization rates around this time. This makes the reversal of fortune even more striking.

This reversal did not take place immediately after the Europeans arrived. Instead, it was related to the fact that some nations took advantage of the wave of industrial technology at the end of the 18th and throughout the 19th centuries. Others did not. The next two figures illustrate this.





Finally, and perhaps most importantly, this reversal, in fact divergence, was related to institutions. It was the interaction between industrialization opportunities and institutions these countries developed, or saw imposed upon them, during colonization that determined whether they took advantage of the industrialization opportunities. Those with institutions that protected property rights and enabled new entrepreneurs and businessmen to enter did so, those with extractive institutions failed. The next table provides evidence in line with this hypothesis, looking at the effect of the interaction between institutions and frontier industrialization on development in a panel of former colonies.

Table IX
The Interaction of UK Industrialization and Institutions

	Former Colonies, using only pre-1950 data	Former Colonies, using data through 1980 (all data)	Former Colonies, using only pre-1950 data	Former Colonies, using only data pre-1950 and for independent countries	Former Colonies, with average institutions for each country, using only pre-1950 data	Former Colonies, with average institutions for each country, using only pre-1950 data	Former Colonies, with average institutions for each country, instrumenting using settler mortality, only pre-1950 data	Former Colonies, with average institutions for each country, instrumenting using settler mortality, only pre-1950 data	Former Colonies, with average institutions for each country, instrumenting using settler mortality, only pre-1950 data	Former Colonies, with average institutions for each country, instrumenting usin settler mortality, on pre-1950 data
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Dependent Variable is Industrial Production Per Capita										
UK Industrialization*Institutions	0.132	0.132	0.145	0.160	0.202	0.206	0.168	0.169	0.156	0.158
	(0.026)	(0.027)	(0.035)	(0.048)	(0.019)	(0.022)	(0.030)	(0.032)	(0.065)	(0.065)
Institutions	8.97	-3.36	10.51	7.48						
	(2.30)	(4.46)	(3.50)	(9.51)						
Independence			-14.3			-6.4		1.1		2.0
			(22.9)			(11.4)		(12.6)		(14.2)
UK Industrialization*Independence			-0.12			-0.042		0.046		0.06
			(0.21)			(0.12)		(0.13)		(0.17)
UK Industrialization*Latitude									0.13	0.12
									(0.50)	(0.48)
R-Squared	0.75	0.74	0.75	0.84	0.89	0.89	0.88	0.88	0.87	0.87
Number of Observations	59	75	59	32	59	59	59	59	59	59
Panel B: Dependent Variable is Log GDP Per Capita										
Log UK Industrialization*Institutions	0.078	0.060	0.073	0.079	0.135	0.130	0.159	0.150	0.116	0.111
	(0.022)	(0.017)	(0.027)	(0.025)	(0.021)	(0.026)	(0.032)	(0.038)	(0.067)	(0.073)
Institutions	-0.027	-0.084	-0.10	-0.11						
	(0.025)	(0.028)	(0.04)	(0.04)						
Independence			0.67			0.12		0.10		0.019
			(0.27)			(0.13)		(0.13)		(0.16)
Log UK Industrialization*Independence			0.035			-0.008		-0.042		0.016
			(0.12)			(0.093)		(0.11)		(0.14)
Log UK Industrialization*Latitude									0.42	0.42
									(0.49)	(0.54)
R-Squared	0.95	0.92	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.96
Number of Observations	79	131	79	46	79	79	79	79	79	79

Overall, the evidence is fairly conclusive that:

(1) Countries that are rich today, especially among the former colonies, are those that, for one reason or another, ended up with good institutions. Somewhat surprisingly, among the former colonies, these are countries that were relatively poor in 1500, before the colonization process started.

(2) These countries became rich, mostly by taking advantage of industrialization opportunities.

(3) There is a strong interaction between relatively good institutions and the capacity of an economy to take advantage of industrialization opportunities.

2.5. Weak and Strong Institutions?

There is also some evidence for the importance of weakly institutionalized environments. Jones and Olken (2005) examine the impact on growth rates of a leader dying in office. They look at whether this exogenous change in the identity of leaders has an impact on growth. The following table illustrates the kinds of events they are focusing on.

TABLE II
Deaths of National Leaders Due to Accidental or Natural Causes

Country	Leader	Year of death	Tenure (years)	Nature of death
Algeria	Houari Boumediene	1978	13.5	Waldenstrom's disease (blood disorder)
Angola	Agostinho Neto	1979	3.9	Cancer of the pancreas
Argentina	Juan Peron	1974	.7[a]	Heart and kidney failure
Australia	John Curtin	1945	3.7	Heart attack
Australia	Harold Holt	1967	1.9	Drowned while skin-diving in Port Philip Bay
Barbados	John (Tom) Adams	1985	8.5	Heart attack
Barbados	Errol Barrow	1987	1.0[a]	No cause of death announced
Bolivia	Rene Barrientos (Ortuna)	1969	2.7[a]	Helicopter crash
Botswana	Sir Seretse Khama	1980	13.8	Cancer of the stomach
Brazil	Arthur da Costa e Silva	1969	2.6	Paralytic stroke, then heart attack
China	Mao Tse-tung	1976	26.9	Parkinson's disease
China	Deng Xiaoping	1997	19.2	Parkinson's disease
Comoros	Prince Jaffar	1975	.4	While on pilgrimage to Mecca
Comoros	Mohamad Taki	1998	2.7	Heart attack
Cote d'Ivoire	Felix Houphouet-Boigny	1993	33.3	Following surgery for prostate cancer
Denmark	Hans Hedtoft	1955	1.3[a]	Heart attack in hotel in Stockholm
Denmark	Hans Hansen	1960	5.0	Cancer
Dominica	Roosevelt Douglas	2000	0.7	Heart attack
Ecuador	Jaime Roldos (Aguilera)	1981	1.8	Plane crash in Andes
Egypt	Gamal Abdel Nasser	1970	15.9	Heart attack
France	Georges Pompidou	1974	4.8	Cancer
Gabon	Leon Mba	1967	7.3	Cancer (in Paris)
Greece	Georgios II	1947	11.4	Heart attack
Grenada	Herbert Blaize	1989	5.0	Prostate cancer
Guinea	Sekou Toure	1984	25.5	Heart attack during surgery in Cleveland
Guyana	Linden Burnham	1985	19.2	During surgery
Guyana	Cheddi Jagan	1997	4.4	Heart attack a few weeks after heart surgery
Haiti	Francois Duvalier	1971	13.5	Heart disease
Hungary	Jozsef Antall	1993	3.6	Lymphatic cancer
Iceland	Bjarni Benediktsson	1970	6.7	House fire
India	Jawaharlal Nehru	1964	16.8	Stroke
India	Lal Bahadur Shastri	1966	1.6	Heart attack
Iran	Ayatollah Khomeini	1989	10.3	Following surgery to stem intestinal bleeding
Israel	Levi Eshkol	1969	5.7	Heart attack

They find large growth effects, though on average growth does not get better or worse (some leaders are good, some are bad). Some of the most fitting pictures are shown next.

DO LEADERS MATTER? 84

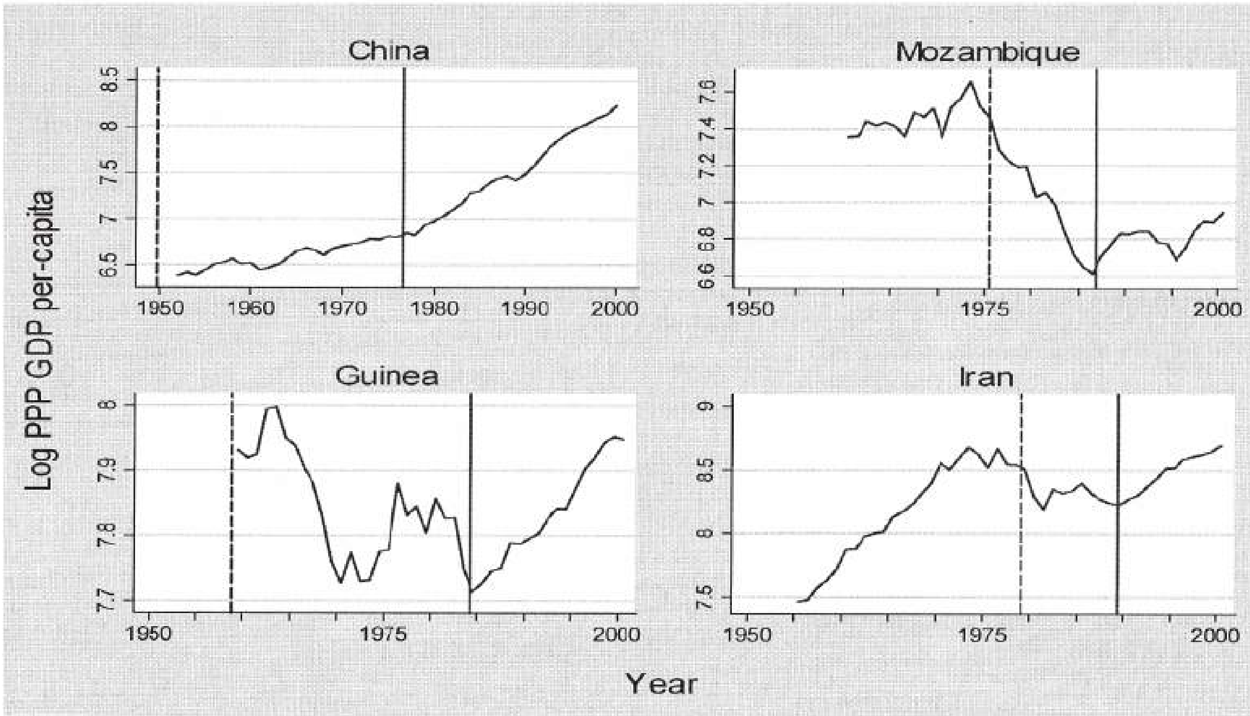


FIGURE I
Growth and Leader Deaths

More interesting than the result that "leaders matter" is the one that leaders matter only in autocracies and not in democracies. This is evidence in favor of the idea that when there are fewer formal constraints on individual politicians, such as the accountability inherent in democratic elections, they can have a large impact on growth.

They also find that among autocrats, there is a particularly strong leader effect in regimes without political parties, but no effects when there are political parties.

2.6. Which Institutions Matter?

The empirical discussion so far emphasized the importance of "institutions".

But these institutions are still something of an amalgam of different elements - a "black box".

For example, in Acemoglu, Johnson and Robinson the proxy for institutions is a measure of security of property rights. But there are two problems:

(1) many other dimensions of institutions are highly correlated with security of property rights, so it is difficult to know which of many institutional features matters more: the security of property rights, democracy, and independent judiciary etc.? For this reason, Acemoglu, Johnson and Robinson often refer to "a cluster of institutions". But this feeds into the second problem.

(2) suppose that we are convinced of the importance of institutions. Then what do we do? What features of institutions do we try to change? Do political institutions matter? If so which? Is it the formal or the informal institutions?

Empirical work on this topic would be very useful, but the problem is going to be one of identification: it is virtually impossible to find exogenous (simultaneous/independent) sources of variation in different components of institutions.

One recent paper attacking these issues is Acemoglu and Johnson (2005). They examine the independent role of secure property rights, what they call "property rights institutions" as compared to "contracting institutions." They think of the former as mediating the relationship between the private sector and the state, while the latter mediates relationships between private individuals. The first is measures by the types of variables used by Acemoglu, Johnson and Robinson, such as protection against expropriation risk. To measure the second they use data from Djankov et al. (2003) and the World Bank on the extent of legal formalism—for example, the number of procedures necessary to collect on a bounced check, and an index of procedural complexity, measuring the difficulties in resolving the case of an unpaid commercial debt. They instrument the property rights by settler mortality and contracting institutions by legal origins. The IV results suggest that while contracting institutions are important for such things as the form of financial intermediation, they do not have a significant effect on growth. Property rights institutions, on the other hand have statistically and quantitatively important have effects on both growth, investment and overall financial development.

We reproduce some key tables, which show the importance of "property rights institutions".

TABLE 3
First-Stage Regressions for Contracting and Property Rights Institutions
(OLS, Sample of Ex-Colonies)

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A. Measure of Contracting Institutions					
	Dependent Variable: Legal Formalism		Dependent Variable: Procedural Complexity		Dependent Variable: Number of Procedures	
English legal origin	−1.98	−1.79	−2.28	−2.24	−11.29	−12.39
	(.23)	(.20)	(.34)	(.29)	(3.31)	(2.88)
Log settler mortality	.09		−.08		1.59	
	(.09)		(1.32)		(1.29)	
Log population density in 1500		.04		−.13		−.38
		(.06)		(.86)		(.84)
R^2 in first stage	.64	.58	.47	.47	.23	.22
Observations	53	64	60	68	61	69
	Panel B. Measure of Property Rights Institutions					
	Dependent Variable: Constraint on Executive		Dependent Variable: Protection against Expropriation		Dependent Variable: Private Property	
English legal origin	−.002	.05	.60	.87	.72	.73
	(.48)	(.43)	(.31)	(.30)	(.22)	(.18)
Log settler mortality	−.66		−.71		−.30	
	(.19)		(.12)		(.09)	
Log population density in 1500		−.40		−.36		−.29
		(.13)		(.09)		(.05)
R^2 in first stage	.21	.15	.50	.35	.37	.47
Observations	51	60	51	57	52	60

Note.—Standard errors are in parentheses. All regressions are cross-sectional OLS with one observation per country. For detailed sources and definitions, see App. table A1.

TABLE 4
CONTRACTING VS. PROPERTY RIGHTS INSTITUTIONS: GDP PER CAPITA AND INVESTMENT-GDP RATIO (2SLS)

	Instrument for Property Rights Institutions					
	Log Settler Mortality (1)	Log Population Density (2)	Log Settler Mortality (3)	Log Settler Mortality (4)	Log Settler Mortality (5)	Log Settler Mortality (6)
	Panel A. Dependent Variable: Log GDP per Capita, Second Stage of 2SLS					
Legal formalism	.05	−.002			.35	.85
	(.24)	(.21)			(.15)	(.45)
Procedural complexity			.097			
			(.17)			
Number of procedures				.02		
				(.04)		
Constraint on executive	.99	.88	.84	.88		
	(.29)	(.27)	(.18)	(.23)		
Average protection against risk of expropriation					.99	
					(.16)	
Private property						2.45
						(.81)
	Results in Equivalent OLS Specification					
Measure of contracting institutions	−.16	−.13	−.050	−.013	.11	.01
	(.10)	(.10)	(.07)	(.009)	(.09)	(.10)
Measure of property rights institutions	.31	.29	.34	.32	.63	.74
	(.07)	(.07)	(.06)	(.06)	(.08)	(.14)
Observations	51	60	60	61	51	52
	Panel B. Dependent Variable: Investment-GDP Ratio, Second Stage of 2SLS					
Legal formalism	−.80	−1.34			.57	3.83
	(1.55)	(1.37)			(1.08)	(2.52)
Procedural complexity			−.60			
			(1.10)			
Number of procedures				−.08		
				(.23)		
Constraint on executive	4.70	4.24	4.21	4.06		
	(1.87)	(1.77)	(1.20)	(1.44)		
Average protection against risk of expropriation					4.68	
					(1.11)	
Private property						13.16
						(4.57)

TABLE 5
Contracting vs. Property Rights Institutions: Private Credit and Stock Market Capitalization (2SLS)

	Instrument for Property Rights Institutions					
	Log Settler Mortality (1)	Log Population Density (2)	Log Settler Mortality (3)	Log Settler Mortality (4)	Log Settler Mortality (5)	Log Settler Mortality (6)
	Panel A. Dependent Variable: Credit to Private Sector, Second Stage of 2SLS					
Legal formalism	−.08 (.08)	−.08 (.06)			−.01 (.07)	.16 (.14)
Procedural complexity			−.05 (.06)			
Number of procedures				−.010 (.012)		
Constraint on executive	.27 (.10)	.17 (.07)	.24 (.06)	.22 (.07)		
Average protection against risk of expropriation					.28 (.07)	
Private property						.70 (.25)
	Results in Equivalent OLS Specification					
Measure of contracting institutions	−.13 (.04)	−.11 (.04)	−.059 (.030)	−.006 (.003)	−.09 (.04)	−.08 (.04)
Measure of property rights institutions	.06 (.03)	.06 (.02)	.08 (.02)	.071 (.02)	.13 (.04)	.21 (.05)
Observations	51	60	60	61	51	52
	Panel B. Dependent Variable: Stock Market Capitalization, Second Stage of 2SLS					
Legal formalism	−.16 (.07)	−.14 (.05)			−.10 (.07)	.04 (.10)
Procedural complexity			−.11 (.06)			
Number of procedures				−.022 (.013)		
Constraint on executive	.20 (.09)	.13 (.07)	.19 (.06)	.14 (.08)		
Average protection against risk of expropriation					.21 (.07)	
Private property						.54 (.20)

Similarly, Acemoglu and Johnson (2005) show that these patterns are robust to a variety of controls and additional specification checks.

2.7. Within Country Variation

Problems of omitted variables are likely to be endemic to cross-country empirical work. One alternative strategy is to exploit variation within a country. There are several recent papers along these lines relevant to the discussion. Banerjee and Iyer (2005) and Iyer (2004) examined the impact of colonial institutions in India on variation in modern economic outcomes, such as agricultural productivity at the district level. They exploit variation in the timing of the creation of different system of taxation and different annexation policies to indentify the causal effect of taxation systems and British occupation. These papers find that districts where the British did tax farming in the 19th century (sold the right to levy taxes to Zamindars) and places which the British ruled directly (as opposed to being ruled indirectly through Indian Princes), have worse economic outcomes today.

Besley, Persson and Sturm (2005) have recently looked at the impact of political competition on economic growth using within-US evidence. They exploit political reforms of the 1960s, particularly the changes embodied in the 1965 voting rights act (abolition of poll taxes and literacy requirements) which were designed to enfranchise blacks in the US South. Using these changes as an instrument for political competition they find large positive effects of greater political competition on growth. They present evidence that the effect works through a more pro-business environment - greater political competition reduces total taxes and corporate taxes in relation to income.

For those of you who studied Meltzer and Richard's seminal (1981) paper these results are rather surprising. They were trying to argue that democratization historically in the US led to the median voter preferring greater amounts of redistribution, something which led to an expansion in the size of the government. What BPS find is that democratization in the US South in the 1960s, led to less redistribution!

We reproduce some key tables.

2.8. Micro Evidence

There is also careful micro evidence on the impact of institutions, such as that from Besley (1995), Field (2003, 2005), Galiani and Schargrodsky (2004) and Goldstein and Udry (2005).

For example, Besley investigates the effect of property rights over land in Ghana on investment. In particular, he looks at whether households with the right to sell, rent, mortgage and pledge the land invest more (he proxies investments by planting new trees, or drainage, land excavation, irrigation or manuring). He typically finds that households that report to have property rights over their land (the right to sell, mortgage etc.) undertake more investments.

A possible concern with these results, as with the aggregate correlations is that of omitted variable bias. Fields with good characteristics may have induced their owners to obtain property rights over them, and will also be naturally more productive, and therefore perhaps induce greater investments. Alternatively, differences in property rights and productivity may reflect heterogeneity among households. In these cases, the association between property rights and investment would not reflect the causal effect of better property rights enforcement on investment.

Besley tries to deal with this problem using an instrumental-variables approach, using the method of land acquisition as an instrument for whether there are property rights over the land. These IV estimates give similar results, but one might be concerned that these instruments do not really solve the endogeneity problem, since it may be different types of lands or households with different characteristics that engage in different types of land transactions.

Recent research by Field and Galiani and Schargrodsky has provided more convincing evidence of the impact of property rights. Both studies exploit very nice natural experiments where the allocation of property rights (titling) was exogenous in the sense of being independent of the underlying characteristics of the land or the people who benefitted. Galiani and Schargrodsky find a large impact of this on investment by people in their properties, but surprisingly find no effect on the credit market (the fact that people with a title have collateral for a loan does not seem to influence transactions in financial markets). Field (2005) finds very consistent results from her research in Peru.

Field (2003) also finds large effects of increasing the security of property rights on labor supply because people with insecure titles have to spend more time occupying their property to guarantee their rights.

Goldstein and Udry (2005) study the impact of property rights in land in rural Ghana. They show that women fail to fallow the land they farm, with large adverse effects of productivity. When women farm, they own the crops they plant, but the land remains communal property. The reason for the absence of fallowing is that if they fallow it, their user rights become insecure because land which is not being farmed can be re-allocated. In particular Goldstein and Udry show that it is because women lack political rights that they cannot influence how chiefs allocate user rights to plots and this is why they do not fallow. As with the Field paper this paper shows that "possession is 9/10ths of the law" but it shows that possession is socially costly.

The following table from Goldstein and Udry shows the role of "political power" in the village on. There evidence clearly shows how office holders, people with de jure political power, fallow more and thus have more productive land.

Table 1: Perceptions of Land Rights

	Percent of Cultivated Plots on which Respondent Claims Right to:				Percent of Plots Fallowed more than Six Years
	Determine Inheritance	Rent Out	Lend Out	Sell	
	(1)	(2)	(3)	(4)	(5)
Non-office holders	6	22	32	15	13
Office holders	26	53	60	32	22
t-test for equality	6.41	6.74	5.83	4.34	2.14
Number of observations	575	576	576	575	406

A point worth noting is that if the question of interest is whether institutional differences have aggregate consequences, this question is very difficult to answer with microdata. The fact that households that have property rights over their land in a given institutional structure behave differently does not imply that all households will start behaving differently once the aggregate extent of property rights enforcement changes. There will be composition, selection, and substitution effects. Therefore, micro evidence is not a perfect substitute for macro evidence, though issues of causality are often better addressed at the micro level.

Taken together, these results, nevertheless, weigh in favor of a view in which institutions are not simply adapting to differences in economic environments, but also cause an important part of these differences in economic environments and economic outcomes.

2.9. Interpreting the Evidence

How can we interpret the above evidence and a large body of other quantitative and qualitative work reaching similar conclusions?

We will argue that this evidence is inconsistent with the "efficient institutions" view and also largely inconsistent with the pure "ideology/beliefs" view.

In addition, many of the details make much more sense in the context of theories where individuals and social groups make decisions understanding their consequences, and this includes also the consequences of different sets of political and economic institutions.

This suggests that the umbrella of "social conflict theories" is the most appropriate one (of course, other approaches will be quite fruitful as well).

To emphasize the differences between various approaches and build a simple taxonomy, consider the following setup from Acemoglu (2003), with Y denoting aggregate output or consumption, which we take to represent social welfare (thus avoiding some of the complications that come from Pareto comparisons, and focusing on the main point here).

Moreover, suppose that we can write

$$Y = F(X, P),$$

where X is a vector of economic, geographic, social or other characteristics that are taken as given and directly influence economic outcomes, and P is a vector of policies and institutions that can potentially affect the outcomes of interest.

Define $\mathbb{P}(. \mid X)$ as the set of policies that maximize output, given a vector of characteristics X, i.e.,

$$P^*(X) \in \mathbb{P}(. \mid X) \iff P^*(X) \in \arg\max_P F(X, P).$$

The Political Coase Theorem (or the efficient institutions view) maintains that there are strong forces leading societies towards some $P^*(X)$ in $\mathbb{P}(. \mid X)$. The underlying idea is that if a society is pursuing a policy $P(X) \notin \mathbb{P}(. \mid X)$, then a switch to $P^*(X) \in \mathbb{P}(. \mid X)$ will create aggregate gains. If these gains correspond to a Pareto improvement, then all political systems will implement this change. If the change creates only a potential Pareto improvement, then part of the gains can be redistributed to those that are losing out via various mechanisms, or at the very least, the winners can lobby or vote for the beneficial change.

To the extent that $\mathbb{P}(. \mid X)$ is not a singleton, we can observe considerable policy differences across two identical societies, but the performance of these two societies should not be appreciably different. An example could be differences in policies regarding the role of the government in the economy between the Anglo-Saxon economies, in particular, the U.S. and the UK, and Continental European countries, which do not seem to lead to major differences in the economic performance between these two sets of countries.

The problem of interpreting evidence on the relationship between a measure of institutions (or policies, or regulation etc.) and economic outcomes is the following. For two societies with characteristics X and $X' \neq X$, we typically have $F(X, P^*(X)) \neq F(X', P^*(X'))$, and moreover, $F(X, P^*(X)) > F(X, P^*(X'))$ and $F(X', P^*(X')) > F(X', P^*(X))$.

This discussion implies that to refute the applicability of the Political Coase Theorem, we need to find systematic evidence that there are societies choosing P while $F(X, P) < F(X, P')$ for some feasible alternative P', or simply that $P \notin \mathbb{P}(. \mid X)$. That is, we need to show that there are societies that persistently pursue wrong policies, with significant output and welfare consequences.

At this point, it is also useful to introduce another variant of the efficient institutions view (or of PCT).

Theories of Belief Differences (Modified PCT): Assume that some subset of X, X_u, is uncertain. To simplify the notation while elaborating on this, suppose that $\mathbb{P}(. \mid X)$ is a singleton, in particular $\mathbb{P}(. \mid X) = P^*(X)$. Moreover, imagine that $X = (X_c, X_u)$, and suppose that $P^*(X_c, X_u) \neq P^*(X_c, X_u')$ whenever $X_u \neq X_u'$, that is, these uncertain characteristics affect which policies are right for the society. Suppose that politicians (or the society at large) have beliefs, denoted by $G(X_u)$, over the actual distribution of X_u. Also suppose that social welfare maximization corresponds to the maximization of expected aggregate output. Then define

$$P^*(X_c, G) \in \arg\max_P \int F(X_c, X_u, P)\, dG.$$

Now two societies with the same X_c, and the same ex post realization of X_u, may choose different policies because their ex ante beliefs over the payoff-relevant characteristics, the X_u's, are different. Given a particular realization of X_u, some societies among those with the same X_c and X_u will be richer than others, i.e., typically $F(X_c, X_u, P^*(X_c, G)) \neq F(X_c, X_u, P^*(X_c, G'))$ for $G \neq G'$.

For example, the North Koreans may be choosing socialist policies and government ownership because they believe those are the policies that will increase welfare, while South Korea, which presumably had the same characteristics, X_c and X_u, chose a capitalist development path. Ex post, the South Koreans turned out to be right, hence they were the ones who adopted the right policies, and the ones who prospered, while North Koreans today suffer poverty and famine.

To refute the class of models in this group, we need to show that there are societies that pursue policies that could not be the right policies under any plausible scenario. In other words, denoting the set of admissible beliefs by $\mathbb{G}$, if, for two feasible policies, P and P', $\int F(X_c, X_u, P')\, dG \geq \int F(X_c, X_u, P)\, dG$ for all $G \in \mathbb{G}$, then we should never observe P.

Finally, according to the *Theories of Social Conflict,* societies often, *knowingly,* choose some policy vector $P(X) \notin \mathbb{P}(. \mid X)$, because policies and institutions are chosen to maximize the payoffs of those who hold political power, *not* to maximize social welfare or aggregate income. To emphasize the difference between this approach and the Political Coase Theorem, imagine another vector of variables Z, which *do not* directly affect Y, thus $P^*(X)$ is independent of Z. These variables may nonetheless influence the "equilibrium" policy, so we can have $P(X, Z)$. Changes in Z will have no direct effect on output, but may have a powerful indirect impact by influencing the gap between $P(X, Z)$ and $P^*(X)$. In other words, we need to find a variable, Z, that is like an instrument in econometrics: it influences X, but has no direct effect on F.

At this level of generality, Theories of Social Conflict are more like a residual group; if we can show that certain societies systematically, and knowingly, pursue inefficient policies, we are in the realm of Theories of Social Conflict. But the usefulness of these theories depends, in turn, on whether they can pinpoint an interesting mechanism for why political and economic bargains are not struck to achieve better policies and institutions (i.e., what are the salient "transaction costs" preventing the PCT from applying?), and whether we can identify a range of institutional or other social variables, the Z's, that affect the degree of inefficiency of policies.

How do we interpret the evidence?

The sources of exogenous variation in the above discussions corresponds to Z in terms of this framework. They do not have a direct effect on what appropriate institutions should be, but they change the equilibrium institutions.

These Z-induced changes have had a big effect on equilibrium institutions and through that channel on economic outcomes. This suggests the importance of thinking through models in which interest of various actors and social groups, not some overarching efficiency objective, determine equilibrium institutions.

What about belief differences?

Is it possible that Kim Il Sung and Communist Party members in the North believed that communist policies would be better for the country and the economy in the late 1940s? Is it possible that European colonists in the Caribbean thought that slavery would be better for the slaves?

The answer to the second question is clearly no. That's a clear example of where it was not the least differences, but social conflict guiding the formation of equilibrium institutions.

What about the first? Perhaps early on that's how communist party members thought. However, by the 1980s it was clear that the communist economic policies in the North were not working. The continued efforts of the leadership to cling to these policies and to power can only be explained by those leaders wishing to look after their own interests at the expense of the population at large. Bad institutions are therefore kept in place, clearly not for the benefit of society as a whole, but for the benefit of the ruling elite, and this is a pattern we encounter in most cases of institutional failure that we discuss in detail below.

2.10. References

(1) Acemoglu, Daron (2005) "Constitutions, Politics and Economic Growth: Review Essay on Persson and Tabellini's *The Economic Effects of Constitutions,*" *Journal of Economic Literature*, XLIII, 1025-1048.

(2) Acemoglu, Daron and Simon Johnson (2005) "Unbundling Institutions," *Journal of Political Economy*, 113, 949-995.

(3) Acemoglu, Daron, Simon Johnson and James A. Robinson (2001) "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, December, 91, 5, 1369-1401.

(4) Acemoglu, Daron, Simon Johnson and James A. Robinson (2002) "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution," *Quarterly Journal of Economics*, 118, 1231-1294.

(5) Banerjee, Abhijit and Lakshmi Iyer (2005) "History, Institutions and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." *American Economic Review*, 95, 1190-1213.

(6) Barro, Robert J. (1997) *The Determinants of Economic Growth: A Cross-Country Empirical Study*, Cambridge; MIT Press.

(7) Besley, Timothy (1995) "Property Rights and Investment Incentives: Theory and Evidence from Ghana." *Journal of Political Economy*, 103, 903-937.

(8) Besley, Timothy, Torsten Persson and Daniel Sturm (2005) "Political Competition and Economic Performance: Theory and Evidence from the United States," Unpublished, London School of Economics.

(9) Djankov, Simeon, Florencio Lopez-de-Silanes, Rafael La Porta and Andrei Shleifer (2003) "Courts," *Quarterly Journal of Economics*, 118, 453-517.

(10) Field, Erica (2003) "Entitled to Work: Urban Property Rights and the Labor Supply in Peru," Unpublished, Harvard Department of Economics.

(11) Field, Erica (2005) "Property Rights and Investment in Urban Slums," Unpublished, Harvard Department of Economics.

(12) Galiani, Sebastián and Ernesto Schargrodsky (2004) "Property Rights for the Poor: The Effect of Land Titling," Unpublished, Universidad de San Andres.

(13) Goldstein, Markus and Christopher Udry (2005) "The Profits of Power: Land Rights and Agricultural Investment in Ghana," http://www.econ.yale.edu/~cru2/pdf/goldsteinudry.pdf.

(14) Hall, Robert E. and Charles I. Jones (1999) "Why Do Some Countries Produce so much more Output per Worker than Others?" *Quarterly Journal of Economics*, 114, 83-116.

(15) Jones, Benjamin F. and Benjamin A. Olken (2005) "Do Leaders Matter? National Leadership and Growth Since World War II," *Quarterly Journal of Economics*, 120, 835-864.

(16) Knack, Steven and Philip Keefer (1995) "Institutions and Economic Performance: Cross-Country Tests using Alternative Measures," *Economics and Politics*, 7, 207-227.

(17) Mauro, Paolo (1995) "Corruption and Growth, *Quarterly Journal of Economics*, 110, 681-712.

(18) Meltzer, Allan M. and Scott Richard (1981) "A Rational Theory of the Size of Government," *Journal of Political Economy*, 89, 914-927.

(19) Mulligan, Casey B., Gil, Richard and Sala-i-Martin, Xavier (2004) "Do Democracies Have Different Public Policies than Nondemocracies?" *Journal of Economic Perspectives*, 18, 51-74.

(20) Nunn, Nathan (2004) "Slavery, Institutional Development and Long-Run Growth in Africa, 1400-2000," Department of Economics, University of British Columbia.

(21) Persson, Torsten and Guido Tabellini (2003) *The Economic Effects of Constitutions: What Do the Data Say?* MIT Press, Cambridge.

(22) Persson, Torsten and Guido Tabellini (2004) "Constitutional rules and fiscal policy outcomes," *American Economic Review,* 94, 25-46.

(23) Persson, Torsten and Guido Tabellini (2005) "Democracy and development: The Devil in the details," forthcoming in *American Economic Review*, Papers and Proceedings, May 2006.

CHAPTER 3

A Review of Dynamic Games

In the rest of the lectures, we will make frequent use of dynamic models of politics, and this will necessitate analysis of dynamic games. Here I provide a brief review of a number of key concepts. Typically we will deal with infinity-repeated discounted dynamic games. The difference between dynamic games and infinitely repeated games is that in dynamic games, there is an underlying state, which evolves over time as a result of the actions by players and by nature. Dynamic games are also sometimes referred to as "stochastic games" following the early article by Lloyd Shapley on this topic.

3.1. Basic Definitions

Let us consider the following class of games. There is a set of players denoted by $\mathcal{N}$. This set will be either finite, or when it is infinite (especially uncountable), there will be more structure to make the game tractable and thus variants of the theorems here applicable. For now, let me focus on the case in which $\mathcal{N}$ is finite, consisting of N players. Each player $i \in \mathcal{N}$ has a strategy set $A_i(k) \subset \mathbb{R}^{n_i}$ at every date, where $k \in K \subset \mathbb{R}^n$ is the state vector, with value at time t denoted by k_t. A generic element of $A_i(k)$ at time t is denoted by a_{it}, and $a_t = (a_{1t}, ..., a_{Nt})$ is the vector of actions at time t, i.e.,

$$a_t \in A(k_t) \equiv \prod_{i=1}^{N} A_i(k_t).$$

I use the standard notation $a_{-it} = (a_{1t}, .a_{i-1,t}, a_{i+1,t}, .., a_{Nt})$ to denote the action vector without i's action, thus we can also write $a_t = (a_{it}, a_{-it})$.

Each player has an instantaneous utility function $u_i(a_t, k_t)$ where

$$u_i : A \times K \to \mathbb{R}$$

is assumed to be continuous and bounded.

Each player's objective at time t is to maximize their discounted payoff

$$U_{it} = \mathbb{E}_t \sum_{s=0}^{\infty} \beta^s u_i(a_{t+s}, k_{t+s}), \tag{3.1}$$

where $\beta \in (0,1)$ is the discount factor and $\mathbb{E}_t$ is the expectations operator conditional on information available at time t (here I am already not indexing $\mathbb{E}_t$ by i, since the focus will be on games with perfect monitoring or perfect observability; see below).

The law of motion of the state vector k_t is given by the following Markovian transition function

$$q\left(k_{t+1} \mid a_t, k_t\right), \tag{3.2}$$

which denotes the probability density that next period's state vector is equal to k_{t+1} when the time t action profile is $a_t \in A\left(k_t\right)$ and the state vector is $k_t \in K$. I refer to this transition function Markovian, since it only depends on the current profile of actions and the current state. Naturally,

$$\int_{-\infty}^{\infty} q\left(k \mid a_t, k_t\right) dk = 1 \text{ for all } a_t \in A\left(k_t\right) \text{ and } k_t \in K.$$

Next, we need to specify the information structure of the players. We focus on games with perfect observability or perfect monitoring, so that individuals observe realizations of all past actions (in case of mixed strategies, they observe realizations of actions not the strategies). Then, the public history at time t, observed by all agents up to time t, is therefore

$$h^t = \left(a_0, k_0, ..., a_t, k_t\right)$$

the history of the game up to and including time t. With mixed strategies, the history would naturally only include the realizations of mixed strategies not the actual strategy. Let the set of all potential histories at time t be denoted by $\mathcal{H}^t$. It should be clear that any element $h^t \in \mathcal{H}^t$ for any t corresponds to a subgame of this game.

Let a (pure) strategy for player i at time t be

$$\sigma_{it} : \mathcal{H}^{t-1} \times K \rightarrow A_i,$$

i.e., a mapping that determines what to play given the entire past history h^{t-1} and the current value of the state variable $k_t \in K$. This is the natural specification of a strategy for time t given that h^{t-1} and k_t entirely determine which subgame we are in.

A mixed strategy for player i at time t is

$$\sigma_{it} : \mathcal{H}^{t-1} \times K \rightarrow \Delta\left(A_i\right),$$

where $\Delta\left(A_i\right)$ is the set of all probability distributions over A_i. We are using the same notation for pure and mixed strategies to economize on notation. Let $\sigma_i = \left(\sigma_{i1}, \sigma_{i2} ..., \sigma_{i\infty}\right)$ the strategy profile of player i in the infinite game, and let $\sigma_i\left[t\right] = \left(\sigma_{it}, ..., \sigma_{i\infty}\right)$ the continuation strategy profile after time t induced by σ_i. Finally let S_i be the set of all feasible σ_i's, and $S_i[t]$ the set of all feasible $\sigma_i[t]$'s. As usual, $S = \prod_{i=1}^{N} S_i$, etc..

As is standard, define the best response correspondence as

$$BR\left(\sigma_{-i}[t] \mid h^{t-1}, k_t\right) = \{\sigma_i[t] \in S_i[t] : \sigma_i[t] \text{ maximizes (10.50) given } \sigma_{-i}[t] \in S_{-i}[t] \ \}$$

Definition 3.1. *A **Subgame Perfect Equilibrium (SPE)** is a strategy profile* $\sigma^* = (\sigma_1^*, ..., \sigma_N^*) \in S$ *such that* $\sigma_i^*[t] \in BR\left(\sigma_{-i}^*[t] \mid h^{t-1}, k_t\right)$ *for all* $\left(h^{t-1}, k_t\right) \in \mathcal{H}^{t-1} \times K$, *for all* $i \in \mathcal{N}$ *and for all* $t = 0, 1, ...$

Therefore, an SPE requires strategies to be best responses to each other given all possible histories, which is a minimal requirement. What is "strong" (or "weak" depending on the perspective) about the SPE is that strategies are mappings from the entire history. As a result, in infinitely repeated games, there are many subgame perfect equilibria. This has prompted game theorists and economists to focus on a subset of equilibria. One possibility would be to look for "stationary" SPEs, motivated by the fact that the underlying game itself is stationary, i.e., payoffs do not depend on calendar time. Another possibility would be to look at the "best SPEs," i.e., those that are on the Pareto frontier, and maximize the utility of one player subject to the utility of the remaining players not being below a certain level.

Perhaps the most popular alternative concept often used in dynamic games is that of Markov Perfect Equilibrium (MPE). The MPE differs from the SPE in only conditioning on the *payoff-relevant* "state". The motivation comes from standard dynamic programming (also known as Markov Decision Problems), where an optimal plan is a mapping from the state vector to the control vector. MPE can be thought of as an extension of this reasoning to game-theoretic situations. The advantage of the MPE relative to the SPE is that most infinite games will have many fewer MPEs than SPEs in general.

We could define payoff relevant history at time t in general as the smallest partition of $\mathcal{P}^t$ of $\mathcal{H}^t$ such that any two distinct elements of $\mathcal{P}^t$ necessarily lead to different payoffs or strategy sets for at least one of the players holding the action profile of all other players constant.

In this case, it is clear that given the Markovian transition function above, the payoff relevant state is simply $k_t \in K$. Then we define a pure Markovian strategy as

$$\hat{\sigma}_i : K \to A_i,$$

and a mixed Markovian strategy as

$$\hat{\sigma}_i : K \to \Delta(A_i).$$

Define the set of Markovian strategies for player i by $\hat{S}_i$ and naturally, $\hat{S} = \prod_{i=1}^{N} \hat{S}_i$.

Notice that I have dropped the t subscript here. Given the way we have specified the game, time is not part of the payoff relevant state. This is a feature of the infinite-horizon nature of the game. With finite horizons, time would necessarily be part of the payoff-relevant state. Naturally, it is possible to imagine more general infinite-horizon games where the payoff function is $u_i(a_t, k_t, t)$, with calendar time being part of the payoff-relevant state.

Note also that $\hat{\sigma}_i$ has a different dimension than σ_i above. In particular, while $\hat{\sigma}_i$ assigns an action (or a probability distribution over actions) to each state $k \in K$, while σ_i does so for each subgame, i.e., for all $(h^{t-1}, k_t) \in \mathcal{H}^{t-1} \times K$ and all t. To compare Markovian and non-Markovian strategies (and to make sure below that we can compare Markovian strategies to deviations that are non-Markovian), it is useful to consider an extension of Markovian strategies to the same dimension as σ_i. In particular, let $\hat{\sigma}'_i$ be an extension of $\hat{\sigma}_i$ such that

$$\hat{\sigma}'_i : K \times \mathcal{H}^{t-1} \to \Delta(A_i)$$

with $\hat{\sigma}'_i(k, h^{t-1}) = \hat{\sigma}_i(k)$ for all $h^{t-1} \in \mathcal{H}^{t-1}$ and $k_t \in K$. Define the set of extended Markovian strategies for player i by $\hat{S}'_i$ and naturally, $\hat{S}' = \prod_{i=1}^{N} \hat{S}'_i$. Moreover, as before, let $\hat{\sigma}'_{it}$ be the continuation strategy profile induced by $\hat{\sigma}'_i$ after time t, and $\hat{\sigma}'^{*}_{-it}$ be the continuation strategy profile of all players other than i induced by their Markovian strategies $\hat{\sigma}^*_{-i}$. I will refer both to $\hat{\sigma}_i$ and its extension $\hat{\sigma}'_i$ as "Markovian strategies".

Let us next define:

DEFINITION 3.2. *A* ***Markov Perfect Equilibrium (MPE)*** *is a profile of Markovian strategies* $\hat{\sigma}^* = (\hat{\sigma}^*_1, ..., \hat{\sigma}^*_N) \in \hat{S}$ *such that the extension of these strategies satisfy* $\hat{\sigma}'^*_i[t] \in BR(\hat{\sigma}'^*_{-i}[t] \mid h^{t-1}, k_t)$ *for all* $(h^{t-1}, k_t) \in \mathcal{H}^{t-1} \times K$, *for all* $i \in \mathcal{N}$ *and for all* $t = 0, 1, ...$

Therefore, the only difference between MPE and SPE is that we restrict attention to Markovian strategies. It is important to note that, as emphasized by the extension of the Markovian strategies to $\hat{\sigma}'^*_i \in \hat{S}'_i$ and the requirement $\hat{\sigma}'^*_i[t] \in BR(\hat{\sigma}'^*_{-i}[t] \mid h^{t-1}, k_t)$, which conditions on history h^t, *we do not restrict deviations to be Markovian.* In particular, for an MPE, a Markovian strategy $\hat{\sigma}^*_i$ must be a best response to $\hat{\sigma}^*_{i-1}$ among all strategies $\sigma_{it} : \mathcal{H}^{t-1} \times K \to \Delta(A_i)$ available at time t.

It should also be clear that an MPE is an SPE, since the extended Markovian strategy satisfies $\hat{\sigma}'^*_i[t] \in BR(\hat{\sigma}'^*_{-i}[t] \mid h^{t-1}, k_t)$, ensuring that $\hat{\sigma}^*_i$ is a best response to $\hat{\sigma}^*_{-i}$ in all subgames, i.e., for all $(h^{t-1}, k_t) \in \mathcal{H}^{t-1} \times K$ and for all t.

3.2. Some Basic Results

The following are some standard results and theorems that are useful to bear in mind for the rest of the course. First, we start with the eminently useful *one-stage deviation principle.* Recall that $\sigma_i[t] = (\sigma_{it}, ..., \sigma_{i\infty})$ denotes the continuation play for player i after date t, and therefore $\sigma_i[t] = (a_{it}, \sigma_i^*[t+1])$ designates the strategy involving action a_{it} at date t and then the continuation play given by strategy $\sigma_i^*[t+1]$.

THEOREM 3.1. ***(One-Stage Deviation Principle)*** *Suppose that the instantaneous payoff function of each player is uniformly bounded, i.e., there exists $B_i < \infty$ for all $i \in \mathcal{N}$ such that $\sup_{k\in K, a\in A(k)} u_i(a,k) < B_i$. Then a strategy profile $\sigma^* = (\sigma_1^*, ..., \sigma_N^*) \in S$ is an SPE [respectively $\hat{\sigma}^* = (\hat{\sigma}_1^*, ..., \hat{\sigma}_N^*) \in \hat{S}$ is an MPE] if and only if for all $i \in \mathcal{N}$, $\left(h^{t-1}, k_t\right) \in \mathcal{H}^{t-1} \times K$ and time t and for all $a_{it} \in A(k_t)$, $\sigma_i[t] = (a_{it}, \sigma_i^*[t+1])$ [resp. $\hat{\sigma}'_{it} = \left(a_{it}, \hat{\sigma}_i'^*[t+1]\right)$] yields no higher payoff to player i than $\sigma_i^*[t]$ [resp. $\hat{\sigma}_i'^*[t]$].*

PROOF. **(Basic Idea)** Fix the strategy profile of other players. Then the problem of individual i is equivalent to a dynamic optimization problem. Then since $\lim_{T\to\infty} \sum_{s=0}^{T} \beta^s u_i(a_{t+s}, k_{t+s}) = 0$ for all $\{a_{t+s}, k_{t+s}\}_{s=0}^{T}$ and all t given the uniform boundedness of instantaneous payoffs and $\beta < 1$, we can apply the principle of optimality from dynamic programming, to obtain of the one-stage deviation principle. □

This theorem basically implies that in dynamic games, we can check whether a strategy is a best response to other players' are you profile by looking at one-stage deviations, keeping the rest of the strategy profile of the deviating player as given. The uniform boundedness assumption can be weakened to require "continuity at infinity", which essentially means that discounted payoffs converge to zero along any history (and this assumption can also be relaxed further).

LEMMA 3.1. *Suppose that $\hat{\sigma}'^*_{-i}$ is Markovian (i.e., it is an extension of a Markovian strategy $\hat{\sigma}^*_{-i}$) and that for $h^{t-1} \in \mathcal{H}^{t-1}$ and $k_t \in K$, $BR\left(\hat{\sigma}'^*_{-i} \mid k_t, h^{t-1}\right) \neq \varnothing$. Then there exists $\hat{\sigma}_i'^* \in BR\left(\hat{\sigma}'^*_{-i} \mid k_t, h^{t-1}\right)$ that is Markovian.*

PROOF. **(Sketch)** Suppose $\hat{\sigma}'^*_{-i}$ is Markovian. Suppose, to obtain a contradiction, that there exists a non-Markovian strategy σ_i^* that performs strictly better against $\hat{\sigma}'^*_{-i}$ than all Markovian strategies. Then, by Theorem 3.1, there exists t, $\tilde{t} > t$, $k \in K$, $h^{t-1} \in H^{t-1}$ and $\tilde{h}^{\tilde{t}-1} \in H^{\tilde{t}-1}$ such that the continuation play following these two histories given $k \in K$ are not the same, i.e., $\sigma_i^*[t]\left(k, h^{t-1}\right) \in BR\left(\hat{\sigma}'^*_{-i} \mid k, h^{t-1}\right)$, $\sigma_i^*[\tilde{t}]\left(k, \tilde{h}^{\tilde{t}-1}\right) \in BR\left(\hat{\sigma}'^*_{-i} \mid k, \tilde{h}^{\tilde{t}-1}\right)$

and $\sigma_i^*[t](k, h^{t-1}) \neq \sigma_i^*[\tilde{t}]\left(k, \tilde{h}^{\tilde{t}-1}\right)$, where $\sigma_i^*[t](k, h^{t-1})$ denotes a continuation strategy for player i starting from time t with state vector k and history h^{t-1}. Now, construct the continuation strategy $\hat{\sigma}_i'^*[\tilde{t}]$ such that $\hat{\sigma}_i'^*[\tilde{t}]\left(k, \tilde{h}^{\tilde{t}-1}\right) = \sigma_i^*[t](k, h^{t-1})$. Since $\hat{\sigma}_{-i}'^*$ is Markovian, $\hat{\sigma}_{-i}'^*[t]$ is independent of $h^{t-1}, \tilde{h}^{t-1}$, and therefore $\hat{\sigma}_i'^*[t](k, h^{t-1}) = \hat{\sigma}_i'^*[t]\left(k, \tilde{h}^{\tilde{t}-1}\right) \in BR\left(\hat{\sigma}_{-i}'^* \mid k, h^{t-1}\right) \cap BR\left(\hat{\sigma}_{-i}'^* \mid k, \tilde{h}^{\tilde{t}-1}\right)$. Repeating this argument for all instances in which σ_i^* is not Markovian establishes that a Markovian strategy $\hat{\sigma}_i'^*$ is also best response to $\hat{\sigma}_{-i}'^*$. □

This lemma states that when all other players are playing Markovian strategies, there exists a best response that is Markovian for each player. This does not mean that there are no other best responses, but since there is a Markovian best response, this gives us hope in constructing Markov Perfect Equilibria. Consequently, we have the following theorem:

Theorem 3.2. *Let K and $A_i(k)$ for all $k \in K$ be finite sets, then there exists an MPE $\hat{\sigma}^* = (\hat{\sigma}_1^*, ..., \hat{\sigma}_N^*)$.*

Proof. **(Sketch)** Consider an extended game in which the set of players is an element (i, k) of $\mathcal{N} \times K$, with payoff function given by the original payoff function for player i starting in state k as in (10.50) and strategy set $A_i(k)$. The set $\mathcal{N} \times K$ is finite, and since $A_i(k)$ is also finite, the set of mixed strategies $\Delta(A_i(k))$ for player (i, k) is the simplex over $A_i(k)$. Therefore, the standard proof of existence of Nash equilibrium based on Kakutani's fixed point theorem applies and leads to the existence of an equilibrium $\left(\hat{\sigma}_{(i,k)}^*\right)_{(i,k)\in\mathcal{N}\times K}$ in this extended game. Now going back to the original game, construct the strategy $\hat{\sigma}_i^*$ for each player $i \in \mathcal{N}$ such that $\hat{\sigma}_i^*(k) = \hat{\sigma}_{(i,k)}^*$, i.e., $\hat{\sigma}_i^* : K \to \Delta(A_i)$. This strategy profile $\hat{\sigma}^*$ is Markovian. Consider the extension of $\hat{\sigma}^*$ to $\hat{\sigma}'^*$ as above, i.e., $\hat{\sigma}_i'^*(k, h^{t-1}) = \hat{\sigma}_i^*(k)$ for all $h^{t-1} \in \mathcal{H}^{t-1}$, $k_t \in K$, $i \in \mathcal{N}$ and t. Then, by construction, given $\hat{\sigma}_{-i}'^*$, it is impossible to improve over $\hat{\sigma}_i'^*$ with a deviation at any $k \in K$, thus Theorem 3.1 implies that $\hat{\sigma}_i'^*$ is best response to $\hat{\sigma}_{-i}'^*$ for all $i \in \mathcal{N}$, so is an MPE strategy profile. □

Similar existence results can be proved for countably infinite sets K and $A_i(k)$, and also for uncountable sets, but in this latter instance, some additional requirements are necessary, and these are rather technical in nature. Since they will play no role in what follows, we do not need to elaborate on these.

For the next result, let $\hat{\Sigma} = \left\{\hat{\sigma}^* \in \hat{S}\text{: } \hat{\sigma}^* \text{ is a MPE}\right\}$ be the set of MPE strategies and $\Sigma^* = \{\sigma \in S\text{: } \sigma^* \text{ is a SPE}\}$ be the set of SPE strategies. Let $\hat{\Sigma}'$ be the extension of $\hat{\Sigma}$ to include conditioning on histories. In particular, as defined before, recall that $\hat{\sigma}_i' : K \times H^{t-1} \to$

$\Delta(A_i)$ is such that $\hat{\sigma}'_i(k, h^{t-1}) = \hat{\sigma}_i(k)$ for all $h^{t-1} \in H^{t-1}$ and $k(t) \in K$, and let

$$\hat{\Sigma}' = \left\{ \begin{array}{c} \hat{\sigma}' \in S\colon\ \hat{\sigma}'_i(k, h^{t-1}) = \hat{\sigma}_i(k) \text{ for all } h^{t-1} \in H^{t-1},\, k(t) \in K \\ \text{and } i \in \mathcal{N} \text{ and } \hat{\sigma} \text{ is a MPE} \end{array} \right\}.$$

Theorem 3.3. ***(Markov Versus Subgame Perfect Equilibria)*** $\hat{\Sigma}' \subset \Sigma^*$.

Proof. This theorem follows immediately by noting that since $\hat{\sigma}^*$ is a MPE strategy profile, the extended strategy profile, $\hat{\sigma}'^*$, is such that $\hat{\sigma}'^*_i$ is a best response to $\hat{\sigma}^*_{-i}$ for all $h^{t-1} \in H^{t-1}$, $k(t) \in K$ and for all $i \in \mathcal{N}$, thus is subgame perfect. □

This theorem implies that every MPE strategy profile corresponds to a SPE strategy profile and any equilibrium-path play supported by a MPE can be supported by a SPE.

Theorem 3.4. ***(Existence of Subgame Perfect Equilibria)*** *Let K and $A_i(k)$ for all $k \in K$ be finite sets, then there exists a SPE $\sigma^* = (\sigma^*_1, ..., \sigma^*_N)$.*

Proof. Theorem 3.2 shows that a MPE exists and since a MPE is a SPE (Theorem 3.3), the existence of a SPE follows. □

When K and $A_i(k)$ are uncountable sets, existence of pure strategy SPEs can be guaranteed by imposing compactness and convexity of K and $A_i(k)$ and quasi-concavity of $U_i[t]$ in $\sigma_i[t]$ for all $i \in \mathcal{N}$ (in addition to the continuity assumptions above). In the absence of convexity of K and $A_i(k)$ or quasi-concavity of $U_i[t]$, mixed strategy equilibria can still be guaranteed to exist under some very mild additional assumptions.

Finally, a well-known theorem for SPE from repeated games also generalizes to dynamic games. Let $p(a \mid \sigma)$ be the probability distribution over the equilibrium-path actions induced by the strategy profile σ, with the usual understanding that $\int_{a \in A} p(a \mid \sigma)\, da = 1$ for all $\sigma \in S$, where A is a set of admissible action profiles. With a slight abuse of terminology, I will refer to $p(a \mid \sigma)$ as the equilibrium-path action induced by strategy σ. Then, let

$$U_i^M(k) = \min_{\sigma_{-i} \in \Sigma_{-i}} \max_{\sigma_i \in \Sigma_i} \mathbb{E} \sum_{s=0}^{\infty} \beta^s u_i(k_{t+s}, a_{t+s}),$$

starting with $k_t = k$ and with k_{t+s} given by (3.2) be the minmax payoff of player i starting with state k. Moreover, let

$$U_i^N(k) = \min_{\sigma \in \Sigma} \mathbb{E} \sum_{s=0}^{\infty} \beta^s u_i(k_{t+s}, a_{t+s}), \tag{3.3}$$

be the minimum SPE payoff of player i starting in state $k \in K$. In other words, this is player i's payoff in the equilibrium chosen to minimize this payoff (starting in state k). Then:

THEOREM 3.5. ***(Punishment with the Worst Equilibrium)*** *Suppose $\sigma^* \in S$ is a pure strategy SPE with the distribution of equilibrium-path actions given by $p(a \mid \sigma^*)$. Then, there exists a SPE $\sigma^{**} \in S$ (possibly equal to σ^*) such that $p(a \mid \sigma^*) = p(a \mid \sigma^{**})$ and σ^{**} involves a continuation payoff of $U_i^N(k)$ to player i, if i is the first to deviate from σ^{**} at date t after some history $h^{t-1} \in H^{t-1}$ and when the resulting state in the next period is $k(t+1) = k$.*

PROOF. **(Sketch)** If σ^* is a SPE, then no player wishes to deviate from it. Suppose that i were to deviate from σ^* at date t after history $h^{t-1} \in H^{t-1}$ and when $k_t = k$. Denote his continuation payoff starting at time t, with $k(t)$, and h^{t-1} by $U_i^d[t]\left(k_t, h^{t-1} \mid \sigma^*\right)$ and denote his equilibrium payoff under σ^* by $U_i^c[t]\left(k_t, h^{t-1} \mid \sigma^*\right)$. σ^* can be a SPE only if

$$U_i^c[t]\left(k_t, h^{t-1} \mid \sigma^*\right) \geq \max_{a_{it} \in A_i(k)} \mathbb{E}\left\{u_i\left(a_{it}, a_{-i}\left(\sigma^*_{-i}\right) \mid k_t, h^{t-1}\right) + \beta U_i^d[t+1]\left(k_{t+1}, h^t \mid \sigma^*_{-i}\right)\right\},$$

where $u_i\left(a_{it}, a_{-i}\left(\sigma^*_{-i}\right) \mid k_t, h^{t-1}\right)$ is the instantaneous payoff of individual i when he chooses action a_{it} in state k_t following history h^{t-1} and other players are playing the (potentially mixed) action profiles induced by σ^*_{-i}, denoted by $a_{-i}\left(\sigma^*_{-i}\right)$. $U_i^d[t+1]\left(k_t, h^{t-1} \mid \sigma^*\right)$ is the continuation payoff following this deviation, with k_{t+1} following from the transition function $q\left(k_{t+1} \mid k_t, a_{it}, a_{-i}\left(\sigma^*_{-i}\right)\right)$ and h^t incorporating the actions $a_{it}, a_{-i}\left(\sigma^*_{-i}\right)$. Note that by construction, the continuation play, following the deviation, will correspond to a SPE, since σ^*_{-i} specifies a SPE action for all players other than i in all subgames, and in response, the best that player i can do is to play an equilibrium strategy.

By definition of a SPE and the minimum equilibrium payoff of player i defined in (3.3),

$$U_i^d[t+1]\left(k_{t+1}, h^t \mid \sigma^*\right) \geq U_i^N(k_{t+1}).$$

The preceding two inequalities imply

$$U_i^c[t]\left(k_t, h^{t-1} \mid \sigma^*\right) \geq \max_{a_{it} \in A_i(k)} \mathbb{E}\left\{u_i\left(a_{it}, a_{-i}\left(\sigma^*_{-i}\right) \mid k_t, h^{t-1}\right) + \beta U_i^N(k_{t+1})\right\}.$$

Therefore, we can construct σ^{**}, which is identical to σ^* except replacing $U_i^d[t+1]\left(k_{t+1}, h^t \mid \sigma^*\right)$ with $U_i^N(k_{t+1})$ following the deviation by player i from σ^* at date t after some history $h^{t-1} \in H^{t-1}$ and when in the next period, we have $k_{t+1} = k$. Since $U_i^N(k_{t+1})$ is a SPE payoff, σ^{**} will also be a SPE. □

This theorem therefore states that in characterizing the set ofsustainable payoffs in SPEs, we can limit attention to SPE strategy profiles involving the most severe equilibrium punishments. A stronger version of this theorem is the following:

THEOREM 3.6. ***(Punishment with Minmax Payoffs)*** *Suppose $\sigma^* \in S$ is a pure strategy SPE with the distribution of equilibrium-path actions given by $p(a \mid \sigma^*)$. Then, there exists $\bar{\beta} \in (0,1)$ such that for all $\beta \geq \bar{\beta}$, there exists a SPE $\sigma^{**} \in S$ (possibly equal to σ^*) with $p(a \mid \sigma^*) = p(a \mid \sigma^{**})$ and σ^{**} involves a continuation payoff of $U_i^M(k)$ to player i, if i is the first to deviate from σ^{**} at date t after some history $h^{t-1} \in H^{t-1}$ and when $k(t) = k$.*

PROOF. The proof is identical to that of Theorem 3.5, except that it uses $U_i^M(k)$ instead of $U_i^N(k)$. When β is high enough, the minmax payoff for player i, $U_i^M(k)$, can be supported as part of a SPE. The details of this proof can be found in Abreu (1988) and a further discussion is contained in Fudenberg and Tirole (1994). □

3.3. Application: Repeated Games With Perfect Observability

For repeated games with perfect observability, both SPE and MPE are easy to characterize. Suppose that the same stage game is played an infinite number of times, so that payoffs are given by

$$U_i[t] = \mathbb{E}_t \sum_{s=0}^{\infty} \beta^s u_i(a_{t+s}), \tag{3.4}$$

which is only different from (10.50) because there is no conditioning on the state variable $k(t)$. Let us refer to the game $\{u_i(a), a \in A\}$ as the stage game. Define

$$m_i = \min_{a_{-i}} \max_{a_i} u_i(a),$$

as the minmax payoff in this stage game. Let $V \in \mathbb{R}^N$ be the set of feasible per period payoffs for the N players, with v_i corresponding to the payoff to player i (so that discounted payoffs correspond to $v_i/(1-\beta)$). Then:

THEOREM 3.7. ***(The Folk Theorem for Repeated Games)*** *Suppose that $\{A_i\}_{i \in \mathcal{N}}$ are compact. Then, for any $v \in V$ such that $v_i > m_i$ for all $i \in \mathcal{N}$, there exists $\bar{\beta} \in [0,1)$ such that for all $\beta > \bar{\beta}$, v can be supported as the payoff profile of a SPE.*

PROOF. **(Sketch)** Construct the following punishment strategies for any deviation: the first player to deviate, i, is held down to its minmax payoff m_i (which can be supported as a SPE). Then, the payoff from any deviation $a \in A_i$ is $D_i(a \mid \beta) \leq d_i + \beta m_i/(1-\beta)$ where d_i is the highest payoff player i can obtain by deviating, which is finite by the fact that u_i is continuous and bounded and A_i is compact. v_i can be supported if

$$\frac{v_i}{1-\beta} \geq d_i + \beta \frac{m_i}{1-\beta}.$$

Since d_i is finite and $v_i > m_i$, there exists $\bar{\beta}_i \in [0, 1)$ such that for all $\beta \geq \beta_i$ this inequality is true. Letting $\bar{\beta} = \max_{i \in \mathcal{N}} \bar{\beta}_i$ establishes the desired result. □

THEOREM 3.8. ***(Unique Markov Perfect Equilibrium in Repeated Games)*** *Suppose that the stage game has a unique equilibrium a^*. Then, there exists a unique MPE in which a^* is played at every date.*

PROOF. The result follows immediately since K is a singleton and the stage payoff has a unique equilibrium. □

This last theorem is natural, but also very important. In repeated games, there is no state vector, so strategies cannot be conditioned on anything. Consequently, in MPE we can only look at the strategies that are best response in the stage game.

EXAMPLE 3.1. **(Prisoner's Dilemma)** Consider the following standard prisoner's dilemma, which, in fact, has many applications in political economy.

	D	C
D	$(0,0)$	$(4,-1)$
C	$(-1,4)$	$(2,2)$

The stage game has a unique equilibrium, which is (D,D). Now imagine this game being repeated an infinite number of times with both agents having discount factor β. The unique MPE is playing (D,D) at every date.

In contrast, when $\beta \geq 1/2$, then (C,C) at every date can be supported as a SPE. To see this, recall that we only need to consider the minmax punishment, which in this case is $(0, 0)$. Playing (C,C) leads to a payoff of $2/(1-\beta)$, whereas the best deviation leads to the payoff of 4 now and a continuation payoff of 0. Therefore, $\beta \geq 1/2$ is sufficient to make sure that the following *grim strategy profile* implements (C,C) at every date: for both players, the strategy is to play C if h^t includes only (C,C) and to play D otherwise.

Why the grim strategy profile is not a MPE is also straightforward to see. This profile ensures cooperation by conditioning on past history, that is, it conditions on whether somebody has defected at any point in the past. This history is not payoff relevant for the future of the game given the action profile of the other player—fixing the action profile of the other player, whether somebody has cheated in the past or not has no effect on future payoffs.

3.4. Application: Common Pool Games

A particularly easy example of dynamic games is common pool games, where individuals decide over time how much to exploit a common resource. This class of games also has multiple applications in political economy, since different individuals or groups in political situations often share the same resource. This class of games also illustrates a number of nice features of MPEs and how they can be computed in practice.

3.4.1. Basic Setup. Imagine the society consists of $N+1<\infty$ players. Denote the set of players again by the set $\mathcal{N}$. Player i has utility function

$$\mathbb{E}_t \sum_{s=0}^{\infty} \beta^s \log\left(c_{t+s}^i\right)$$

at time t where $\beta \in (0,1)$ and c_t^i denotes consumption of individual $i \in \mathcal{N}$ at time t. The society has a common resource, denoted by K_t, which can be thought of as the capital stock at time t. This capital stock follows the non-stochastic law of motion:

$$K_{t+1} = AK_t - \sum_{i \in \mathcal{N}} c_t^i,$$

where $A>0$, K_0 is given and $K_t \geq 0$ must be satisfied in every period. This equation is the transition function, and shows how capital accumulates. If $A=1$, this would correspond to a fixed resource game, where the resource is being run down. But $A>1$ is possible, thus the potential for growth in the capital stock is also allowed.

The stage game is as follows: at every date all players simultaneously announce $\left\{c_t^i\right\}_{i \in \mathcal{N}}$. If $\sum_{i \in \mathcal{N}} c_t^i \leq AK_t$, then each individual consumes c_t^i. If $\sum_{i \in \mathcal{N}} c_t^i > AK_t$, then AK_t is equally allocated among the $N+1$ players.

3.4.2. Single Person Decision Problem. Before discussing MPEs, let us first look at the single-person decision problem, which is useful both to develop some of the basic dynamic programming language and tools, and as a comparison for the MPE below. Imagine $\left\{c_t^i\right\}_{i \in \mathcal{N}}$ is being chosen by a benevolent planner, wishing to maximize the total discounted payoff of all the agents in the society:

$$\mathbb{E}_t \sum_{i \in \mathcal{N}} \sum_{s=0}^{\infty} \beta^s \log\left(c_{t+s}^i\right).$$

By concavity, it is clear that this planner will allocate consumption equally across all individuals at every date. Thus let total consumption at time t be C_t, then each individual will consume $C_t/(1+N)$. This implies that the program of the planner at time $t=0$ is:

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\log\left(C_t\right) - \log\left(1+N\right)\right] \tag{3.5}$$

subject to

$$K_{t+1} = AK_t - C_t. \tag{3.6}$$

Clearly, the second term in the objective function (3.5) does not depend on any choice, and can be dropped. One can also note that this problem is identical to the consumption choice of an agent facing a constant gross interest rate equal to A.

Let us write this as a dynamic programming recursion. In particular, dropping expectations (since there is no uncertainty in this case), defining S as the savings (capital stock) left for next period, so that $C = AK - S$, we can write

$$V(K) = \max_{S \leq AK} \{\log(AK - S) + \beta V(S)\} \tag{3.7}$$

as the dynamic programming recursion. A solution to this problem is a pair of functions $V(K)$ and $h(K)$, such that $S \in h(K)$ is optimal. Standard arguments of dynamic programming imply that:

(1) $V(K)$ is uniquely defined, is continuous, concave and also differentiable whenever $S \in (0, AK)$ [these results follow from the following observations: (3.7) can be written as $v(\cdot) = T[v(\cdot)]$, where T is a contraction over the space of continuous bounded functions, thus has a unique fixed-point, denoted by $V(\cdot)$. This function satisfies $V(K_0) = \max_{\{C_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \log C_t$ subject to (3.6) and starting with K_0. Moreover, the fact that the instantaneous payoff function is continuous and concave is sufficient to establish that $V(K)$ is continuous, concave and differentiable in the interior of the constraint set.]

(2) $h(K)$ is single valued. [This follows from the strict concavity of the instantaneous payoff function].

Given these, the optimal plan $S = h(K)$ must satisfy the following necessary and sufficient first-order condition, whenever S is interior:

$$\frac{1}{AK - S} = \beta V'(S). \tag{3.8}$$

Moreover, since $V(K)$ is differentiable, using the envelope condition, we have

$$V'(K) = \frac{A}{AK - S} \tag{3.9}$$

[This is just the same as the standard envelope condition; $V'(K) = A/(AK - S) + (\beta V'(S) - 1/(AK - S)))dS/dK$, and the term multiplying dS/dK is equal to zero from the first-order condition, (3.8)]. Thus we have the following first-order condition

$$\frac{1}{AK - h(K)} = \frac{\beta A}{Ah(K) - h(h(K))},$$

which solves for $h(\cdot)$ as the optimal policy function.

In fact, the easiest way of making progress is to use the "guess and verify" method. This method is powerful in the context of dynamic programming, since we know that $V(\cdot)$ and

$h(\cdot)$ are uniquely defined. So if we can find one pair of functions that satisfy the necessary conditions, we are done.

Let us (with a little bit of experience and foresight) guess:

$$V(K) = \delta_0 + \delta_1 \log K \tag{3.10}$$
$$h(K) = \gamma_0 + \gamma_1 K$$

This implies

$$V'(K) = \frac{\delta_1}{K},$$

which from (3.9) gives:

$$\frac{\delta_1}{K} = \frac{A}{AK - S} = \frac{A}{AK - (\gamma_0 + \gamma_1 K)},$$

where the second equality substitutes from (3.10). It is therefore clear that $\gamma_0 = 0$ and

$$\delta_1 = \frac{A}{A - \gamma_1}.$$

Next using (3.8) together with (3.10), we have

$$\frac{1}{AK - \gamma_1 K} = \beta \frac{\delta_1}{\gamma_1 K},$$

or

$$\frac{1}{A - \gamma_1} = \frac{\beta A}{\gamma_1 (A - \gamma_1)},$$

which implies

$$\gamma_1 = \beta A$$

as the optimal savings rate of the single-person decision problem.

This is naturally the same as the usual optimal savings decision, with A playing the role of the gross interest rate. If the gross interest rate is greater than the inverse of the discount factor, individuals accumulate, otherwise they run down their assets. Continuing with the substitutions, we also have

$$\delta_1 = \frac{A}{A - \gamma_1} = \frac{A}{A - \beta A} = \frac{1}{1 - \beta}.$$

Finally, to find δ_0 (which is, in fact, not such an interesting parameter), note that

$$\delta_0 + \delta_1 \log K = \log (AK - \gamma_1 K) + \beta \left[\delta_0 + \delta_1 \log (\gamma_1 K)\right],$$

so that

$$\begin{aligned}\delta_0 &= \frac{\log (1-\beta) A + \log K + \beta \log (K) / (1-\beta) + \beta \log (\beta A) / (1-\beta) - \log K / (1-\beta)}{1-\beta} \\ &= \frac{(1-\beta) \log (1-\beta) + \beta \log (\beta) + \log A}{(1-\beta)^2}.\end{aligned}$$

Therefore, the unique value function can be written as:

$$V(K) = \frac{(1-\beta)\log(1-\beta) + \beta\log(\beta) + \log A}{(1-\beta)^2} + \frac{\log K}{1-\beta}.$$

3.4.3. Markov Perfect Equilibria in the Common Pool Game. The common pool game has some uninteresting MPEs. For example, all individuals announcing $c_0^i = AK_0$ is an MPE, since there are no profitable deviations by any agents. However, it yields negative infinite utility to all agents (since after this deviation $K_t = 0$ for all $t > 0$ and $\log(0) = -\infty$). Instead, we will look for a more interesting, continuous and symmetric MPE. The symmetry requirement is for simplicity, and implies that all agents will use the same Markovian strategy. Let that strategy be denoted by

$$c^N(K).$$

Given this, each individual's optimization problem can again be written recursively. Define individual consumption as $c_i = AK - \sum_{j \neq i} c_j - S$, with S as the capital stock left for next period. Given our restriction to symmetric Markovian strategies (and dropping conditioning on i), this gives

$$c = AK - Nc^N(K) - S,$$

so that we can write

$$V^N(K) = \max_{S \leq AK - Nc^N(K)} \left\{\log\left(AK - Nc^N(K) - S\right) + \beta V^N(S)\right\}, \tag{3.11}$$

where the solution is now a triple of functions $V^N(K)$, $c^N(K)$ and $h^N(K)$, such that $S \in h^N(K)$ is best response, where conditioning on N denotes that this refers to the game with $N+1$ players.

The method of solution is the same as before, though since this is now a solution to a game, $V^N(\cdot)$ depends on the function $c^N(\cdot)$, and we cannot establish that $V^N(K_N)$ is uniquely defined, or that it is continuous, concave or differentiable without knowing more about $c^N(\cdot)$. Nevertheless, let us assume that it is so (again this will be verified by the solution).

Assuming differentiability, the first-order condition of the maximization problem in (3.11) is

$$\frac{1}{AK - Nc^N(K) - S} = \beta\left(V^N\right)'(S). \tag{3.12}$$

The envelope condition is

$$\left(V^N\right)'(K) = \frac{A - \left(c^N\right)'(K)}{AK - Nc^N(K) - S}$$

Notice the term $\left(c^N\right)'(K)$ in the numerator. This is there because individuals realize that by their own action they will affect the state variable, and by affecting the state variable, they will influence the consumption decision of others. This is where the subtlety of dynamic games come in. Nevertheless, in some political economy applications, this effect is often ignored or forgotten, hence my emphasis on it here.

Now let us conjecture that:

$$\begin{aligned} V^N(K) &= \delta_0^N + \delta_1^N \log K, \\ h^N(K) &= \gamma_1^N K, \\ c^N(K) &= c_0^N K, \end{aligned} \tag{3.13}$$

where we could have allowed a constant in the savings and consumption functions, but with the same argument as in the single-person decision problem, these will turn out to be equal to zero.

Since the resource constraint of the economy in a symmetric equilibrium is $S \leq AK - (1+N)\,c^N(K)$ and will hold as equality, we also have

$$\gamma_1^N K = AK - (1+N)\,c_0^N K,$$

which implies

$$c_0^N = \frac{A - \gamma_1^N}{1+N}. \tag{3.14}$$

Now using the envelope condition together with (3.13), we have

$$\begin{aligned} \left(V^N\right)'(K) &= \frac{A - Nc_0^N}{AK - Nc_0^N K - \gamma_1^N K} \\ &= \frac{A - N\left(\frac{A-\gamma_1^N}{1+N}\right)}{AK - N\left(\frac{A-\gamma_1^N}{1+N}\right)K - \gamma_1^N K}, \end{aligned}$$

where the second line uses (3.14). Moreover, again from (3.13), we have

$$\left(V^N\right)'(K) = \frac{\delta_1^N}{K},$$

so

$$\begin{aligned} \delta_1^N &= \frac{(1+N)\,A - \left(A - \gamma_1^N\right)}{(1+N)\,A - N\left(A - \gamma_1^N\right) - \gamma_1^N} \\ &= \frac{A + N\gamma_1^N}{A + (N-1)\,\gamma_1^N}. \end{aligned}$$

Finally, using the first-order condition (3.12), we have

$$\begin{aligned}\frac{1}{\left(A - N\left(\frac{A-\gamma_1^N}{1+N}\right) - \gamma_1^N\right)K} &= \beta \cdot \frac{A - N\left(\frac{A-\gamma_1^N}{1+N}\right)}{\left(A - N\left(\frac{A-\gamma_1^N}{1+N}\right) - \gamma_1^N\right)S} \\ & \quad \beta \cdot \frac{A - N\left(\frac{A-\gamma_1^N}{1+N}\right)}{\left(A - N\left(\frac{A-\gamma_1^N}{1+N}\right) - \gamma_1^N\right)\gamma_1^N K},\end{aligned}$$

which implies

$$\gamma_1^N = \frac{\beta A}{1 + N - \beta N}$$

as the equilibrium savings rate in the MPE of the common pool problem. Notice that when $N = 0$, this is exactly equal to the optimal value $\gamma_1 = \beta A$ obtained from the single-person decision problem (or the social plan's problem). Moreover, it is obvious that

$$\frac{\partial \gamma_1^N}{\partial N} < 0,$$

that is, the more players there are drawing resources from the common pool, the lower is the savings rate of the economy. This captures the well-known problem, which sometimes goes under the names of the *free-rider problem* or *tragedy of the commons.* The inability of the players in this game to coordinate their actions leads to too much consumption and too little savings. For example, it is quite possible that

$$\beta A > 1 > \frac{\beta A}{1 + N - \beta N},$$

so that, the social planner's solution would involve growth, while the MPE would involve the resources shrinking over time.

It is useful to note that there are many Markov Perfect Equilibria in this game. We have already seen one at the beginning, which involved the entire capital stock of the economy being consumed in the first period. Perhaps unsurprisingly, there also exist non-symmetric MPEs. More interestingly, there may also exist discontinuous MPEs that implement the social planner's outcome. Let us return to those once we look at the subgame perfect equilibria of the game.

3.4.4. Subgame Perfect Equilibria in the Common Pool Problem. The common pool problem can also be used to illustrate the difference between SPE and MPE. Recall that for SPE, we can restrict attention to the most severe punishment, i.e., the minmax. We already saw that there exists an equilibrium in which all individuals receive negative infinite utility (the one in which they all play $c_0^i = AK_0$ at time $t = 0$). This immediately implies that for any discount factor β, any allocation can be supported as an SPE. In particular, the social planner's solution of saving the society's resources at the rate $\gamma_1 = \beta A$ is an SPE.

Let $\hat{h}^t$ be the time t history such that in all past periods, each individual has consumed

$$c_i(K) = \frac{(1-\beta) AK}{1+N}.$$

The (grim) strategy profile that would support this SPE is as follows:

$$c_i\left(h^t, K\right) = \begin{cases} \frac{(1-\beta)AK}{1+N} & \text{if } h^t = \hat{h}^t \\ AK & \text{if } h^t \neq \hat{h}^t \end{cases},$$

for all t, $K > 0$ and $i \in \mathcal{N}$. Therefore, this strategy profile follows the social planner's allocation until one agent deviates from it, and as soon as there is such a deviation, all agents switch to demanding the whole capital stock of the economy, which leads to infinite utility to all, and as seen above, is subgame perfect (i.e., each is playing a best response to the others' strategies).

Now finally, let us return to the question of whether there exist MPEs that achieve the same allocation as the best SPE. In the special case where $\beta A = 1$ the answer is yes. Suppose that the game starts with capital stock K_0 and consider following discontinuous strategy profile:

$$c_i(K) = \begin{cases} \frac{\beta AK}{1+N} & \text{if } K \geq K_0 \\ K & \text{if } K < K_0 \end{cases}.$$

It can be verified that when all players other than i' pursue this strategy, it is a best response for player i' to play this strategy as well, and along the equilibrium path, the social planner's allocation is implemented. Essentially, when somebody deviates and consumes too much, the capital stock of the economy will fall below K_0, and all individuals will enter a subgame perfect punishment phase. The remarkable thing is that this can be done with Markovian strategies, which only depend on the current state. Nevertheless, this particular equilibrium relies heavily on the fact that there are a finite number of players (and no noise). Moreover, it can be verified that when $\beta A \neq 1$, such a Markovian equilibrium cannot be constructed because the level of the capital stock consistent with "cooperation" will depend on past levels of capital stock, which are not part of the payoff relevant state. Therefore, this special type of MPE can only be supported when $\beta A = 1$.

3.5. References

(1) Abreu, Dilip (1988) "On the Theory of Repeated Games with Discounting" *Econometrica*, 56, 383-396.

(2) Fudenberg, Drew and Jean Tirole (1991) *Game Theory*, Cambridge; MIT Press. Chapters 4, 5 and 13.

(3) Stokey, Nancy, Robert E. Lucas and Edward Prescott (1989) *Recursive Methods in Economic Dynamics*, Harvard University Press, Cambridge, Chapter 4.

CHAPTER 4

Static Voting Models

Our first purpose is to understand how conflicting preferences are aggregated into collective choices under different political institutions. A natural starting point is voting, which is viewed as the prototypical set of institutions for decision-making in democratic societies. Many of the questions we are interested in require moving beyond models of static voting. But before we can do this, it is useful to review the general set of issues that arise in aggregating conflicting preferences into social choices and how voting over different alternatives aggregates (or fails to aggregate) individuals' preferences. We will see later that voting is not the only way in which decisions are made in democratic societies and many other decision-making procedures are used in non-democratic societies. But voting is still the natural place to start.

4.1. Arrow's Impossibility Theorem

The vast area of social choice theory is concerned with the fundamental question of political economy already discussed at the beginning of this chapter: how to aggregate the preferences of heterogeneous agents over policies (collective choices). Differently from the most common political economy approaches, however, social choice theory takes an axiomatic approach to this problem. Nevertheless, a quick detour into social chose theory as an introduction to the Median Voter Theorem is useful.

Let us consider an abstract economy consisting of a finite set of individuals $\mathcal{H}$, with the number of individuals denoted by H. Individual $i \in \mathcal{H}$ has a utility function

$$u\left(x_i, Y\left(x, p\right), p \mid \alpha_i\right).$$

Here x_i is his action, with a set of feasible actions denoted by X_i; p denotes the vector of political choices (for example, institutions, policies or other collective choices), with the menu of policies denoted by $\mathcal{P}$; and $Y\left(x, p\right)$ is a vector of general equilibrium variables, such as prices or externalities that result from all agents' actions as well as policies, and x is the vector of the x_i's. Instead of writing a different utility function u_i for each agent, I have parameterized the differences in preferences by the variable α_i. This is without loss of any generality (simply define $u_i(\cdot) \equiv u_i(\cdot \mid \alpha_i)$) and is convenient for some of the analysis that will follow. Clearly, the general equilibrium variables, such as prices, represented by $Y(x,p)$ here, need not be uniquely defined for a given set of policies p and vector of individual choices x. Since multiple equilibria are not our focus here, I ignore this complication and assume that $Y(x,p)$ is uniquely defined.

I also assume that, given aggregates and policies, individual objective functions are strictly quasi-concave so that each agent has a unique optimal action $x_i(p, Y(x,p), \alpha_i) = \arg\max_{x \in X_i} u(x_i, Y(x,p), p \mid \alpha_i)$. Substituting this maximizing choice of individual i into his utility function, we obtain his *indirect utility function* defined over policy as $U(p;\alpha_i)$. Next let us define the *preferred policy*, or the (political) bliss point, of voter i, and to simplify notation, suppose that this is uniquely defined and denote it by

$$p(\alpha_i) = \arg\max_{p \in \mathcal{P}} U(p;\alpha_i).$$

In addition, we can think of a more primitive concept of individual preference orderings, which captures the same information as the utility function $U(p;\alpha_i)$. In particular, if individual individual i weakly prefers p to p', we write $p \succeq_i p'$ and if he has a strict preference, we write $p \succ_i p'$. Under the usual assumptions on individual preferences (*completeness*, which allows any two choices to be compared; *reflexivity*, so that $z \succeq_i z$; and *transitivity*, so that $z \succeq_i z'$ and $z' \succeq_i z''$ implies $z \succeq_i z''$), we can equivalently represent individual preferences by the ordering $\succeq_i$ or by the utility function $U(p;\alpha_i)$. Throughout let us assume that individual preferences are transitive.

In this context, we can also think of a "political system" as a way of aggregating the set of utility functions, $U(p;\alpha_i)$'s, to a social welfare function $U^S(p)$ that ranks policies for the society. Put differently, a political system is a *mapping* from individual preference orderings to a social preference ordering. Arrow's Theorem shows that if this mapping satisfies some relatively weak conditions, then social preferences have to be "dictatorial" in the sense that they will exactly reflect the preferences of one of the agents. I next present this theorem.

Let us simplify the discussion by assuming that the set of feasible policies, $\mathcal{P}$, is finite and is a subset of the Euclidean space, that is, $\mathcal{P} \subset \mathbb{R}^K$ where $K \in \mathbb{N}$. Let $\Re$ be the set of all weak orders on $\mathcal{P}$, that is, $\Re$ contains information of the form $p_1 \succeq_i p_2 \succeq_i p_3$ and so on, and imposes the requirement of transitivity on these individual preferences. An individual ordering R_i is an element of $\Re$, that is, $R_i \in \Re$. This statement reiterates that we are only considering individuals with well-defined transitive preferences.

Since our society consists of H individuals, we define $\rho = (R_1, ..., R_H) \in \Re^H$ as the society's *preference profile.* That is, ρ gives the preference ordering of each individual $i \in \mathcal{H}$. Also $\rho_{|\mathcal{P}'} = \left(R_{1|\mathcal{P}'}, ..., R_{H|\mathcal{P}'}\right)$ is the society's preference profile when alternatives are restricted to some subset $\mathcal{P}'$ of $\mathcal{P}$.

Let $\Im$ be the set of all reflexive and complete binary relations on $\mathcal{P}$ (but notice *not necessarily* transitive). A social ordering $R^S \in \Im$ is therefore a reflexive and complete binary relation over all the policy choices in $\mathcal{P}$. Thus, a social ordering can be represented as

$$\phi : \Re^H \to \Im.$$

This mathematical formalism implies that $\phi(\rho)$ gives the social ordering for the preference profiles in ρ. We can alternatively think of ϕ as a political system mapping individual preferences into a social choice. A trivial example of ϕ is the dictatorial ordering making agent 1 the dictator, so that for any preference profile $\rho \in \Re^H$, ϕ induces a social order that entirely coincides with R_1.

Note that our formulation already imposes the condition of "*unrestricted domain*," which says that in constructing a social ordering we should consider all possible (transitive) individual orderings. Therefore, we are not limiting ourselves to a special class of individual orderings, such as those with "single-peaked" preferences as we will do later in this section.

We say that a social ordering is *weakly Paretian* if $[p \succ_i p'$ for all $i \in \mathcal{H}] \implies p \succ^S p'$, that is, if all individuals in the society prefer p to p', then the social ordering must also rank p ahead of p'. This is weakly Paretian (rather than strongly), since we require all agents to strictly prefer p to p'.

Next we say that, for a preference profile $\rho \in \Re^H$, a subset $\mathcal{D}$ of $\mathcal{H}$ is *decisive between* $p, p' \in \mathcal{P}$, if $[p \succeq_i p'$ for all $i \in \mathcal{D}$ and $p \succ_{i'} p'$ for some $i' \in \mathcal{D}] \implies p \succ^S p'$ (given ρ). If $\mathcal{D}' \subset \mathcal{H}$ is decisive between $p, p' \in \mathcal{P}$ for *all* preference profiles $\rho \in \Re^H$, then it is *dictatorial between* $p, p' \in \mathcal{P}$. We say that $\mathcal{D} \subset \mathcal{H}$ is *decisive* if it is decisive between any $p, p' \in \mathcal{P}$, and $\mathcal{D}' \subset \mathcal{H}$ is *dictatorial* if it is dictatorial between any $p, p' \in \mathcal{P}$, If $\mathcal{D}' \subset \mathcal{H}$ is dictatorial and a singleton, then its unique element is a *dictator*, meaning that social choices will exactly reflect his preferences regardless of the preferences of the other members of the society. In this case, we say that a social ordering ϕ is *dictatorial.*

Next a social ordering satisfies *independence from irrelevant alternatives*, if for any ρ and $\rho' \in \Re^H$ and any p, $p' \in \mathcal{P}$,

$$\rho_{|\{p,p'\}} = \rho'_{|\{p,p'\}} \Longrightarrow \phi(\rho)_{|\{p,p'\}} = \phi(\rho')_{|\{p,p'\}}.$$

The axiom of independence from irrelevant alternatives is essential for Arrow's Theorem. It states that if two preference profiles have the same choice over two policy alternatives, the social orderings that derive from these two preference profiles must also have identical choices over these two policy alternatives, regardless of how these two preference profiles differ for "irrelevant" alternatives. While this condition (axiom) at first appears plausible, it is in fact a reasonably strong one. In particular, it rules out any kind of interpersonal "cardinal" comparisons—that is, it excludes information on how strongly an individual prefers one outcome versus another.

The main theorem of the field of social choice theory is the following:

Theorem 4.1. ***(Arrow's (Im)Possibility Theorem)*** *If a social ordering, ϕ, is transitive, weakly Paretian and satisfies independence from irrelevant alternatives, then it is dictatorial.*

Proof. The proof is in two steps.

Step 1: Let a set $\mathcal{J} \subset \mathcal{H}$ be *strongly decisive* between $p_1, p_2 \in \mathcal{P}$ if for any preference profile $\rho \in \Re^H$ with $p_1 \succ_i p_2$ for all $i \in \mathcal{J}$ *and* $p_2 \succ_j p_1$ for all $j \in \mathcal{H} \backslash \mathcal{J}$, $p_1 \succ^S p_2$ ($\mathcal{H}$ itself is strongly decisive since ϕ is weakly Paretian). We first prove that if $\mathcal{J}$ is strongly decisive between $p_1, p_2 \in \mathcal{P}$, then $\mathcal{J}$ is dictatorial (and hence decisive for all $p, p' \in \mathcal{P}$ and for all preference profiles $\rho \in \Re^H$). To prove this, consider the restriction of an arbitrary preference profile $\rho \in \Re^H$ to $\rho_{|\{p_1,p_2,p_3\}}$ and suppose that we also have $p_1 \succ_i p_3$ for all $i \in \mathcal{J}$. Next consider an alternative profile $\rho'_{|\{p_1,p_2,p_3\}}$, such that $p_1 \succ'_i p_2 \succ'_i p_3$ for all $i \in \mathcal{J}$ and $p_2 \succ'_i p_1$ and $p_2 \succ'_i p_3$ for all $i \in \mathcal{H} \backslash \mathcal{J}$. Since $\mathcal{J}$ is strongly decisive between p_1 and p_2, $p_1 \succ'^S p_2$. Moreover, since ϕ is weakly Paretian, we also have $p_2 \succ'^S p_3$, and thus $p_1 \succ'^S p_2 \succ'^S p_3$. Notice that $\rho'_{|\{p_1,p_2,p_3\}}$ did not specify the preferences of individuals $i \in \mathcal{H} \backslash \mathcal{J}$ between p_1 and p_3, but we have established $p_1 \succ'^S p_3$ for $\rho'_{|\{p_1,p_2,p_3\}}$. We can then invoke independence from irrelevant alternatives and conclude that the same holds for $\rho_{|\{p_1,p_2,p_3\}}$, i.e., $p_1 \succ^S p_3$. But then, since the preference profiles and p_3 are arbitrary, it must be the case that $\mathcal{J}$ is dictatorial between p_1 and p_3. Next repeat the same argument for $\rho_{|\{p_1,p_2,p_4\}}$ and $\rho'_{|\{p_1,p_2,p_4\}}$, except that now $p_4 \succ_i p_2$ and $p_4 \succ'_i p_1 \succ'_i p_2$ for $i \in \mathcal{J}$, while $p_2 \succ'_j p_1$ and $p_4 \succ'_j p_1$ for all $j \in \mathcal{H} \backslash \mathcal{J}$. Then, the same chain of reasoning, using the facts that $\mathcal{J}$ is strongly decisive, $p_4 \succ'^S p_2$, ϕ is weakly Paretian and satisfies independence from irrelevant alternatives, implies that $\mathcal{J}$ is dictatorial between p_4 and p_2 (that is, $p_4 \succ^S p_2$ for any preference profile $\rho \in \Re^H$). Now once again using independence from irrelevant alternatives and also transitivity, for any preference profile $\rho \in \Re^H$, $p_4 \succ_i p_3$ for all $i \in \mathcal{J}$. Since $p_3, p_4 \in \mathcal{P}$ were arbitrary, this completes the proof that $\mathcal{J}$ is dictatorial (i.e., dictatorial for all $p, p' \in \mathcal{P}$).

Step 2: Given the result in Step 1, if we prove that some individual $h \in \mathcal{H}$ is strongly decisive for some $p_1, p_2 \in \mathcal{P}$, we will have established that it is a dictator and thus ϕ is dictatorial. Let $\mathcal{D}_{ab}$ be the strongly decisive set between p_a and p_b. Such a set always exists for any p_a, $p_b \in \mathcal{P}$, since $\mathcal{H}$ itself is a strongly decisive set. Let $\mathcal{D}$ be the minimal strongly decisive set (meaning the strongly decisive set with the fewest members). This is also well-defined, since there is only a finite number of individuals in $\mathcal{H}$. Moreover, without loss of generality, suppose that $\mathcal{D} = \mathcal{D}_{12}$ (i.e., let the strongly decisive set between p_1 and p_2 be the minimal strongly decisive set). If $\mathcal{D}$ a singleton, then Step 1 applies and implies that ϕ is dictatorial, completing the proof. Thus, to obtain a contradiction, suppose that $\mathcal{D} \neq \{i\}$. Then, by unrestricted domain, the following preference profile (restricted to $\{p_1, p_2, p_3\}$) is feasible

$$\begin{array}{cc} \text{for } i \in \mathcal{D} & p_1 \succ_i p_2 \succ_i p_3 \\ \text{for } j \in \mathcal{D} \backslash \{i\} & p_3 \succeq_j p_1 \succ_j p_2 \\ \text{for } k \notin \mathcal{D} & p_2 \succ_k p_3 \succ_k p_1. \end{array}$$

By hypothesis, $\mathcal{D}$ is strongly decisive between p_1 and p_2 and therefore $p_1 \succ^S p_2$. Next if $p_3 \succ^S p_2$, then given the preference profile here, $\mathcal{D} \backslash \{i\}$ would be strongly decisive between p_2 and p_3, and this would contradict that $\mathcal{D}$ is the minimal strongly decisive set. Thus $p_2 \succsim^S p_3$. Combined with $p_1 \succ^S p_2$, this implies $p_1 \succ^S p_3$. But given the preference profile here, this implies that $\{i\}$ is strongly decisive, yielding another contradiction. Therefore, the minimal strongly decisive set must be a singleton $\{h\}$ for some $h \in \mathcal{H}$. Then, from Step 1, $\{h\}$ is a dictator and ϕ is dictatorial, completing the proof. □

An immediate implication of this theorem is that any set of minimal decisive individuals $\mathcal{D}$ within the society $\mathcal{H}$ must either be a singleton, that is, $\mathcal{D} = \{i\}$, so that we have a dictatorial social ordering, or we have to live with intransitivities.

While this theorem is often referred to as Arrow's Possibility Theorem, it is really an "Impossibility Theorem". An alternative way of stating the theorem is that there exists *no* social ordering that is transitive, weakly Paretian, consistent with independence from irrelevant alternatives and non-dictatorial. Viewed in this light, an important implication of this theorem is that there is no way of avoiding the issue of conflict in preferences of individuals by positing a social welfare function. A social welfare function, respecting transitivity, can only replace the actual political economic process of decision making when it is dictatorial. Naturally, who will become the dictator in the society fundamentally brings back the issue of *political power*, which is also essential for any positive political economy analysis of collective decision-making. In addition, from a modeling point of view, Arrow's Theorem means that, if we are interested in non-dictatorial (and transitive) outcomes, we have to look at political systems that either restrict choices or focus on more concrete situations, where we have to be more specific about the distribution of political power and the political institutions regulating the decision-making process. This will be the basis of our analysis for the rest of this chapter and for the next chapter.

Often, economic models restrict the policy space and/or preferences of citizens in order to ensure that Arrow's Impossibility Theorem does not apply. Unfortunately, such restrictions on the policy space have more than technical implications. For example, they often force the modeler to restrict agents to use inefficient methods of redistribution. As a result, some of the inefficiencies that are found in political economy models are not a consequence of the logic of these models, but a consequence of the technical assumptions that the modelers make in restricting the policy space to a single policy. In some circumstances, limits on fiscal instruments might be justified on economic grounds.

One reaction to Arrow's Theorem might be that the problem of aggregating individual preferences arises because we are not looking at more relevant mechanisms such as voting. The next section shows that the same problems arise when collective choices are made by voting. In fact Arrow's Theorem applies to any possible way of aggregating individual preferences, and if voting were able to solve the problems raised by the theorem, it would be a contradiction to the theorem! Nevertheless, voting can be useful in situations where we put more structure on preferences and on how individuals vote, which will essentially amount to either giving up the "unrestricted domain" assumption on choices or relaxing the independence from irrelevant alternatives.

4.2. Voting and the Condorcet Paradox

Let us illustrate how voting also runs into exactly same problems as those highlighted by Arrow's Theorem by using a well-known example, *the Condorcet paradox.* The underlying reason for this paradox is related to Arrow's Theorem and will also illustrate why, to obtain the Median Voter Theorem below, we will have to introduce reasonably strong restrictions.

Example 4.1. Imagine a society consisting of three individuals, 1, 2, and 3 and three choices. The individuals' preferences are as follows:

$$\begin{array}{ll} 1 & a \succ c \succ b \\ 2 & b \succ a \succ c \\ 3 & c \succ b \succ a \end{array}$$

Moreover, let us make the political mechanism somewhat more specific, and assume that it satisfies the following three requirements, which together make up the "open agenda direct democracy" system.

A1. *Direct democracy.* The citizens themselves make the policy choices via majoritarian voting.

A2. *Sincere voting.* In every vote, each citizen votes for the alternative that gives him the highest utility according to his policy preferences (indirect function) $U(p;\alpha_i)$. This requirement is adopted for simplicity. In many situations, individuals may vote for the outcome that they do not prefer, anticipating the later repercussions of this choice (we refer to this type of behavior as "*strategic voting*"). Whether they do so or not is important in certain situations, but not for the discussion at the moment.

A3. *Open agenda.* Citizens vote over pairs of policy alternatives, such that the winning policy in one round is posed against a new alternative in the next round and the set of alternatives includes all feasible policies. Later, we will replace the open agenda assumption with parties offering policy alternatives, thus moving away from direct democracy some way towards indirect/representative democracy. For now it is a good starting point.

Now, using the three assumptions, consider a contest between policies a and b. In this contest, agents 2 and 3 will vote for b over a, so b is the majority winner. Next, by the open agenda assumption, the other policy alternative c will run against b. Now agents 1 and 3 prefer c to b, which is the new majority winner. Next, c will run against a, but now agents 1 and 2 prefer a, so a is the majority winner. Therefore, in this case we have "cycling" over the various alternatives, or put differently there is no "equilibrium" of the voting process that selects a unique policy outcome.

For future reference, let us now define a Condorcet winner as a policy choice that does not lead to such cycling. In particular,

Definition 4.1. *A* ***Condorcet winner*** *is a policy p^* that beats any other feasible policy in a pairwise vote.*

In light of this definition, there is no Condorcet winner in the example of the Condorcet paradox.

4.3. Single-Peaked Preferences and the Median Voter Theorem

Suppose now that the policy space is unidimensional, so that p is a real number, that is, $\mathcal{P} \subset \mathbb{R}$. In this case, a simple way to rule out the Condorcet paradox is to assume that preferences are *single peaked* for all voters. We will see below that the restriction that $\mathcal{P}$ is unidimensional is very important and single-peaked preferences are not well defined when there are multiple policy dimensions.

We say that voter i has single-peaked preferences if his preference ordering for alternative policies is dictated by their relative distance from his bliss point, $p(\alpha_i)$: a policy closer to $p(\alpha_i)$ is preferred over more distant alternatives. Specifically:

Definition 4.2. *Consider a finite set of $\mathcal{P} \subset \mathbb{R}$ and let $p(\alpha_i) \in \mathcal{P}$ be individual i's unique bliss point over $\mathcal{P}$. Then, the policy preferences of citizen i are **single peaked** iff:*

$$\begin{aligned} \text{For all } p'', p' &\in \mathcal{P}, \text{ such that } p'' < p' \leq p(\alpha_i) \text{ or } p'' > p' \geq p(\alpha_i), \\ \text{we have } U(p''; \alpha_i) &< U(p'; \alpha_i). \end{aligned}$$

Note that strict concavity of $U(p'; \alpha_i)$ is sufficient for it to be single peaked, but is not necessary. In fact, single-peakedness is equivalent to strict quasi-concavity. This definition could be weakened so that the bliss point of the individual is not unique (that is, it can be weakened from strict quasi-concavity to quasi-concavity). But this added generality is not important for our purposes.

We can easily verify that in the Condorcet paradox, not all agents possessed single-peaked preferences. For example, taking the ordering to be a, b, c, agent 1 who has preferences $a \succ c \succ b$ does not have single-peaked preferences (if we took a different ordering of the alternatives, then the preferences of one of the other two agents would violate the single-peakedness assumption).

The next theorem shows that with single-peaked preferences, there always exists a Condorcet winner. Before stating this theorem, let us define the *median voter* of the society. Given the assumption that each individual has a unique bliss point over $\mathcal{P}$, we can rank all individuals according to their bliss points, the $p(\alpha_i)$'s. Also, to remove uninteresting ambiguities, let us imagine that H is an odd number (i.e., $\mathcal{H}$ consists of an odd number of individuals). Then, the median voter is the individual who has exactly $(H-1)/2$ bliss points to his left and $(H-1)/2$ bliss points to his right. Put differently, his bliss point is exactly in the middle of the distribution of bliss points. We denote this individual by α_m , and his bliss point (ideal policy) is denoted by p_m.

Theorem 4.2. ***(The Median Voter Theorem)*** *Suppose that H is an odd number, that A1 and A2 hold and that all voters have single-peaked policy preferences over a given ordering of policy alternatives, $\mathcal{P}$. Then, a Condorcet winner always exists and coincides* *with the median-ranked bliss point, p_m. Moreover, p_m is the unique equilibrium policy (stable point) under the open agenda majoritarian rule, that is, under A1-A3.*

Proof. The proof is by a "separation argument". Order the individuals according to their bliss points $p(\alpha_i)$, and label the median-ranked bliss point by p_m. By the assumption that H is an odd number, p_m is uniquely defined (though α_m may not be uniquely defined). Suppose that there is a vote between p_m and some other policy $p'' < p_m$. By definition of single-peaked preferences, for every individual with $p_m < p(\alpha_i)$, we have $U(p_m; \alpha_i) > U(p''; \alpha_i)$. By A2, these individuals will vote sincerely and thus, in favor of p_m. The coalition voting for supporting p_m thus constitutes a majority. The argument for the case where $p'' > p_m$ is identical. $\square$

The assumption that the society consists of an odd number of individuals was made only to shorten the statement of the theorem and the proof. It is straightforward to generalize the theorem and its proof to the case in which H is an even number.

More important than whether there is an odd or even number of individuals in the society is the assumption of sincere voting. Clearly, rational agents could deviate from truthful reporting of their preferences (and thus from truthful voting) when this is beneficial for them. So an obvious question is whether the MVT generalizes to the case in which individuals do not vote sincerely? The answer is yes. To see this, let us modify the sincere voting assumption to strategic voting:

A2$'$. *Strategic voting.* Define a *vote function* of individual i in a pairwise contest between p' and p'' by $v_i(p', p'') \in \{p', p''\}$. Let a voting (counting) rule in a society with H citizens be $V: \{p', p''\}^H \to \{p', p''\}$ for any $p', p'' \in \mathcal{P}$. (For example, the majoritarian voting rule V^M picks p' over p'' when this policy receives more votes than p''). Let $V(v_i(p', p''), v_{-i}(p', p''))$ be the policy outcome from voting rule V applied to the pairwise contest $\{p', p''\}$, when the remaining individuals cast their votes according to the vector $v_{-i}(p', p'')$, and when individual i votes $v_i(p', p'')$. Strategic voting means that

$$v_i\left(p', p''\right) \in \arg\max_{\tilde{v}_i(p', p'')} U\left(V\left(\tilde{v}_i\left(p', p''\right), v_{-i}\left(p', p''\right)\right); \alpha_i\right).$$

In other words, strategic voting implies that each individual chooses the voting strategy that maximizes utility given the voting strategies of other agents.

Finally, recall that a *weakly-dominant* strategy for individual i is a strategy that gives weakly higher payoff to individual i than any of his other strategies regardless of the strategy profile of other players

THEOREM 4.3. ***(The Median Voter Theorem With Strategic Voting)*** *Suppose that H is an odd number, that A1 and A2′ hold and that all voters have single-peaked policy preferences over a given ordering of policy alternatives, $\mathcal{P}$. Then, sincere voting is a weakly-dominant strategy for each player and there exists a unique weakly-dominant equilibrium, which features the median-ranked bliss point, p_m, as the Condorcet winner.*

PROOF. The vote counting rule (the political system) in this case is majoritarian, denoted by V^M. Consider two policies $p', p'' \in \mathcal{P}$ and fix an individual $i \in \mathcal{H}$. Assume without loss of any generality that $U(p'; \alpha_i) \geq U(p''; \alpha_i)$. Suppose first that for any $v_i \in \{p', p''\}$, $V^M(v_i, v_{-i}(p', p'')) = p'$ or $V^M(v_i, v_{-i}(p', p'')) = p''$, that is, individual i is *not* pivotal. This implies that $v_i(p', p'') = p'$ is a best response for individual i. Suppose next that individual i is pivotal, that is, $V^M(v_i(p', p''), v_{-i}(p', p'')) = p'$ if $v_i(p', p'') = p'$ and $V^M(v_i(p', p''), v_{-i}(p', p'')) = p''$ otherwise. In this case, the action $v_i(p', p'') = p'$ is clearly a best response for i. Since this argument applies for each $i \in \mathcal{H}$, it establishes that voting sincerely is a weakly-dominant strategy and the conclusion of the theorem follows from Theorem 4.2. □

Notice that the second part of the Theorem 4.2, which applied to open agenda elections, is absent in Theorem 4.3. This is because the open agenda assumption does not lead to a well defined game, so a game-theoretic analysis and thus an analysis of strategic voting is no longer possible.

In fact, there is no guarantee that sincere voting is optimal in dynamic situations even with single-peaked preferences. The following example illustrates this:

EXAMPLE 4.2. Consider three individuals with the following preference orderings.

$$\begin{array}{ll} 1 & a \succ b \succ c \\ 2 & b \succ c \succ a \\ 3 & c \succ b \succ a \end{array}$$

These preferences are clearly single peaked (order them alphabetically to see this). In a one round vote, b will beat any other policy. But now consider the following dynamic voting set up: first, there is a vote between a and b. Then, the winner goes against c, and the winner of this contest is the social choice. Sincere voting will imply that in the first round players 2 and 3 will vote for b, and in the second round, players 1 and 2 will vote for b, which will become the collective choice.

Is such sincere voting "equilibrium behavior"? Exactly the same argument as above shows that in the second round, sincere voting is a weakly dominant strategy. But not necessarily in round one. Suppose players 1 and 2 are playing sincerely. Now if player 3 deviates and votes for a (even though she prefers b), then a will advance to the second round and would lose to c. Consequently, the social choice will coincide with the bliss point of player 3. What happens if all players are voting strategically?

Dynamic voting issues become more interesting, and open the way for agenda setting, when there are no Condorcet winners. The following example illustrates this.

Example 4.3. Consider the preference profile in Example 4.1 and the following political mechanism. First, all individuals vote between a and b, and then they vote over the winner of this contest and c. With sincere voting, b will win the first round, and then c wins the second round against b. Now consider agent 2. If he changes his vote in the first round to a (thus does not vote sincerely), the first-round winner will be a, which will also win against c, and player 2 prefers this outcome to the outcome of sincere voting, which was c.

This example can also be used to illustrate the role of "agenda setting". Suppose that in the above game, agent 1 decides the sequence of alternatives presented for voting. In particular, he has to choose between three options (a vs. b first, a vs. c first, and b vs. c, first). Anticipating strategic voting by player 2, he will choose the first option and will ensure that his most preferred alternative becomes the political choice of the society. In contrast, if agent 3 chose the sequence, he would go for a vs. c first, which would induce agent 1 to vote strategically for c, and lead to c as the ultimate outcome.

4.4. Party Competition and the Downsian Policy Convergence Theorem

The focus so far has been on voting between two alternative policies or on open agenda voting, which can be viewed as an extreme form of "direct democracy". The MVT becomes potentially more relevant and more powerful when applied in the context of indirect democracy, that is, when combined with a simple model of party competition. I now give a brief overview of this situation and derive the Downsian Policy Convergence Theorem, which is the basis of much applied work in political economy.

Suppose that there is a Condorcet winner, and there are two parties, A and B, competing for political office. Assume that the parties do not have an ideological bias, and would like to come to power. In particular, they both maximize the probability of coming to power, for example, because they receive a rent or utility of $Q > 0$ when they are in power.

Assume also that parties simultaneously announce their policy, and are committed to this policy. This implies that the behavior of the two parties can be represented by the following pair of maximization problems:

$$\text{(4.1)} \qquad \begin{aligned} \text{Party } A &: \quad \max_{p_A} \; \mathbb{P}(p_A, p_B)Q \\ \text{Party } B &: \quad \max_{p_B} \; (1 - \mathbb{P}(p_A, p_B))\, Q \end{aligned}$$

where Q denotes the rents of being in power and $\mathbb{P}(p_A, p_B)$ is the probability that party A comes to power when the two parties' platforms are p_A and p_B respectively. Let the bliss point of the median voter be p_m. When the median voter theorem applies, we have

$$\text{(4.2)} \qquad \begin{aligned} \mathbb{P}(p_A, p_B &= p_m) = 0, \; \mathbb{P}(p_A = p_m, p_B) = 1, \text{ and} \\ \mathbb{P}(p_A &= p_m, p_B = p_m) \in [0, 1]\,. \end{aligned}$$

This last statement follows since when both parties offer exactly the same policy, it is a best response for all citizens to vote for either party. However, the literature typically makes the following assumption:

A4. *Randomization:*

$$\mathbb{P}(p_A = p_m, p_B = p_m) = 1/2.$$

This assumption can be rationalized by arguing that when indifferent individuals, randomize between the two parties, and since there are many many individuals, by the law of large numbers, each party obtains exactly half of the vote.

We then have the following result:

THEOREM 4.4. ***(Downsian Policy Convergence Theorem)*** *Suppose that there are two parties that first announce a policy platform and commit to it and a set of voters $\mathcal{H}$ that vote for one of the two parties. Assume that A4 holds and that all voters have single-peaked policy preferences over a given ordering of policy alternatives, and denote the median-ranked bliss point by p_m. Then, both parties will choose p_m as their policy platform.*

PROOF. The proof is by contradiction. Suppose not, then there is a profitable deviation for one of the parties. For example, if $p^A > p^B > p_m$, one of the parties can announce p_m and win the election for sure. When $p^A \neq p_m$ and $p^B = p_m$, party A can also announce p_m and increase its chance of winning to 1/2. □

Interestingly, it is easy to see that this theorem does not hold without Assumption A4. However, in that case, the policy implemented will still be the one corresponding to the bliss point of the median-ranked voter. But there no longer need be convergence between the two parties.

This theorem is important because it demonstrates that there will be policy convergence between the two parties and that party competition will implement the Condorcet winner among the voters. Therefore, in situations in which the MVT applies, the democratic process of decision making with competition between two parties will lead to a situation in which both parties will choose their policy platform to coincide with the bliss point of the median voter. Thus the MVT and the Downsian Policy Convergence Theorem together enable us to simplify the process of aggregating the heterogeneous preferences of individuals over policies and assert that, under the appropriate assumptions, democratic decision-making will lead to the most preferred policy of the median voter. The Downsian Policy Convergence Theorem is useful in this context, since it gives a better approximation to "democratic policymaking" in practice than open agenda elections.

There is a sense in which Theorem 4.4 is slightly misleading, however. While the theorem is correct for a society with two parties, it gives the impression of a general tendency towards policy convergence in all democratic societies. Many democratic societies have more than two parties. A natural generalization of this theorem would be to consider three or more parties. Unfortunately, policy convergence does not generalize to three parties. Thus some care is necessary in applying the Downsian Policy Convergence Theorem in the context of different political institutions.

Another obvious question is what would happen in the party competition game when there is no Condorcet winner. Theorem 4.4 does not generalize to this case either. In particular, if we take a situation in which there is "cycling," like the above Condorcet paradox example, it is straightforward to verify that there is no pure strategy equilibrium in the political competition game.

4.5. Beyond Single-Peaked Preferences

Single-peaked preferences played a very important role in the results of Theorem 4.2 by ensuring the existence of a Condorcet winner. However, single peakedness is a very strong assumption and does not have a natural analog in situations in which voting is over more than one policy choice. When there are multiple policy choices (or when voting is over "functions" such as nonlinear taxation), much more structure needs to be imposed over voting procedures and agenda setting to determine equilibrium policies. Those issues are beyond the scope of my treatment here. Nevertheless, it is possible to relax the assumption of single-peaked preferences and also introduce a set of preferences that are "close" to single-peaked in multidimensional spaces, for example, the so-called *intermediate preferences*. Next consider a situation in which preferences are defined over multidimensional policies. However, maintain the assumption that individuals are heterogeneous only in one dimension. Under certain special circumstances we can still use a median-voter theorem. This requires all citizens to have "intermediate preferences".

Definition 4.3. *Voters have* ***intermediate preferences****, if their indirect utility function* $U(p;\alpha^i)$ *can be written as*

$$U(p;\alpha^i) = J(p) + K(\alpha^i)H(p),$$

where $K(\alpha^i)$ *is monotonic in* α^i*, for any* $H(p)$ *and* $J(p)$ *common to all voters.*

Theorem 4.5. ***(Median Voter Theorem with Intermediate Preferences)*** *Suppose that A2 holds and voters have intermediate preferences. Then a Condorcet winner always exists and coincides with the bliss point of the voter with the median value* α^i*,* $p(\alpha^m)$*.*

Proof. The proof is analogous to the proof of the above median-voter theorem. Since $p(\alpha^m)$ is the maximum for agent α^m, we have that

$$U(p(\alpha^m);\alpha^m) = J(p(\alpha^m)) + K(\alpha^m)H(p(\alpha^m)) \geq U(p;\alpha^m) = J(p) + K(\alpha^m)H(p),$$

for all $p \neq p(\alpha^m)$. Therefore,

$$\begin{aligned} K(\alpha^m) &\lesseqqgtr \frac{J(p(\alpha^m)) - J(p)}{H(p) - H(p(\alpha^m))} \text{ as } H(p) \gtreqqless H(p(\alpha^m)), \\ \text{for any } p &\neq p(\alpha^m). \end{aligned}$$

Suppose that $H(p) > H(p(\alpha^m))$ and $K(\alpha^i)$ is monotonically increasing (the case of monotonically decreasing is analogous). Then $K(\alpha^i) \geq K(\alpha^m)$ for all $\alpha^i > \alpha^m$, and these will all satisfy

$$K(\alpha^i) > \frac{J(p(\alpha^m)) - J(p)}{H(p) - H(p(\alpha^m))}$$

and therefore

$$U(p(\alpha^m);\alpha^i) = J(p(\alpha^m)) + K(\alpha^i)H(p(\alpha^m)) \geq U(p;\alpha^i) = J(p) + K(\alpha^i)H(p),$$

so they will support $p(\alpha^m)$ against p. The other cases are proved similarly. This shows that the policy $p(\alpha^m)$ always collects at least half the votes against any alternative policy. □

Example 4.4. Suppose all individuals have the same exogenous income $y^i = y$ and are subject to the same income tax τ. They are thus subject to the same budget constraint

$$c = y(1-\tau).$$

Government revenue per capita, τy, is spent on two types of public consumption, in per capita amounts p_1 and p_2, to satisfy the government budget constraint

$$p_1 + p_2 \leq \tau y.$$

Agents have heterogenous preferences for these publicly provided goods, however, summarized in the following utility function for voter i:

$$U^i = u(c) + \alpha^i G(p_1) + (1-\alpha^i)F(p_2)),$$

where the weight α^i is distributed in the population on the interval $[0,1]$. In this setting it is easy to derive the policy preferences of agent i over the two-dimensional policy $p \equiv (p_1, p_2)$, treating τ as residual. These preferences are

$$U(p_1, p_2; \alpha^i) = u(y - p_1 - p_2) + F(p_2) + \alpha^i (G(p_1) - F(p_2)).$$

Clearly, these preferences satisfy the "intermediate preferences" property, despite the fact that policies are two-dimensional. Then party competition will lead to the median voter's choice, which is given by

$$\max_{p_1,p_2} U(p;\alpha^m) = u(y - p_1 - p_2) + F(p_2) + \alpha^m (G(p_1) - F(p_2)).$$

The first-order conditions are

$$\begin{aligned} -u'(y-p_1-p_2) + \alpha^m G'(p_1) &= 0 \\ -u'(y-p_1-p_2) + (1-\alpha^m)\, F'(p_2) &= 0 \end{aligned}$$

These conditions determine a unique policy outcomes (p_1', p_2'), and it is straightforward to do comparative statics. For example, an increase in α^m, which corresponds to a median voter that prefers the public good p_1 more strongly, will increase p_1 and reduce p_2.

4.6. Preferences with Single Crossing

Unfortunately, the set of situations in which such generalizations are usefulare limited. Instead the *single-crossing property* is often more straightforward and useful. It will enable us to prove a version of Theorem 4.2 under somewhat weaker assumptions.

Definition 4.4. *Consider an ordered policy space $\mathcal{P}$ and also order voters according to their α_i's. Then, the preferences of voters satisfy the* ***single-crossing property*** *over the policy space $\mathcal{P}$ when the following statement is true:*

$$\begin{aligned} \textit{if } p &> p' \textit{ and } \alpha_{i'} > \alpha_i, \textit{ or if } p < p' \textit{ and } \alpha_{i'} < \alpha_i, \textit{ then} \\ U(p;\alpha_i) &> U(p';\alpha_i) \textit{ implies that } U(p;\alpha_{i'}) > U(p';\alpha_{i'}). \end{aligned}$$

EXAMPLE 4.5. To see why single-crossing property is weaker than single-peaked preferences, consider the following example:

$$\begin{array}{ll} 1 & a \succ b \succ c \\ 2 & a \succ c \succ b \\ 3 & c \succ b \succ a \end{array}$$

It can be verified easily that these preferences are not single peaked. The natural ordering is $a > b > c$, but in this case the preferences of player 2 have two peaks, at a and c. To see why these preferences satisfy single crossing, take the same ordering, and also order players as 1, 2, 3. Now,

$$\begin{array}{rcl} \alpha & = & 2\text{: } c \succ b \Longrightarrow \alpha = 3\text{: } c \succ b \\ \alpha & = & 2\text{: } \begin{array}{c} a \succ c \\ a \succ b \end{array} \Longrightarrow \alpha = 1\text{: } \begin{array}{c} a \succ c \\ a \succ b \end{array}. \end{array}$$

Notice that while single peakedness is a property of preferences only, the single-crossing property refers to a set of preferences over a given policy space $\mathcal{P}$. It is therefore a joint property of preferences and choices. The following theorem generalizes Theorem 4.2 to a situation with single crossing.

THEOREM 4.6. ***(Extended Median Voter Theorem)*** *Suppose that A1 and A2 hold and that the preferences of voters satisfy the single-crossing property. Then, a Condorcet winner always exists and coincides with the bliss point of the median voter (voter α_m).*

PROOF. The proof works with exactly the same separation argument as in the proof of Theorem 4.2. Consider the median voter with α_m, and bliss policy p_m. Consider an alternative policy $p' > p_m$. Naturally, $U(p_m; \alpha_m) > U(p'; \alpha_m)$. Then, by the single crossing property, for all $\alpha_i > \alpha_m$, $U(p_m; \alpha_i) > U(p'; \alpha_i)$. Since α_m is the median, this implies that there is a majority in favor of p_m. The same argument for $p' < p_m$ completes the proof □

Given this theorem, the following result is immediate:

THEOREM 4.7. ***(Extended Downsian Policy Convergence)*** *Suppose that there are two parties that first announce a policy platform and commit to it and a set of voters that vote for one of the two parties. Assume that A4 holds and that all voters have preferences that satisfy the single-crossing property and denote the median-ranked bliss point by p_m. Then, both parties will choose p_m as their policy.*

4.7. Application

As an application, consider the canonical redistributive taxation problem (used by Roberts, Romer and Meltzer-Richards). It is a situation with two parties competing to come to power. Suppose that agents have the following preferences

$$u^i\left(c^i, x^i\right) = c^i + h(x^i)$$

where c^i and x^i denote individual consumption and leisure, and $h(\cdot)$ is a well-behaved concave utility function. There are only two policy instruments, linear tax on earnings τ on lump-sum transfers $T \geq 0$. It is a very important assumption of the model (enabling the application of the MVT and the other theorems above) that there are such a limited fiscal instruments.

The budget constraint of each agent is

$$c^i \leq (1-\tau)l^i + T,$$

The real wage is exogenous and normalized to 1. Individual productivity differs, such that the individuals have different amounts of "effective time" available. That is, individuals are subject to the "time constraint"

$$\alpha^i \geq x^i + l^i,$$

Therefore, α^i is a measure of "individual productivity". We assume that α^i is distributed in the population with mean α and median α^m.

Since individual preferences are linear in consumption, optimal labor supply satisfies

$$l^i = L(\tau) + (\alpha^i - \alpha),$$

where $L(\tau) \equiv \alpha - (h')^{-1}(1-\tau)$ is decreasing in τ by the concavity of $h(\cdot)$.

A higher tax rate on labor income distorts the labor-leisure choice and induces the consumer to work less. This will be the cost of redistributive taxation in this model. Let l denote average labor supply. Since the average of α^i is α, we have $l = L(\tau)$. The government budget constraint can therefore be written:

$$T \leq \tau l \equiv \tau L(\tau).$$

Let $U(\tau; \alpha^i)$ be utility of agent of type α^i from tax policy τ with the lump-sum transfer T determined as residual. By straightforward substitution into the individual utility function, we can express the policy preferences of individual i as

$$U(\tau; \alpha^i) \equiv L(\tau) + h(\alpha - L(\tau)) + (1-\tau)(\alpha^i - \alpha). \tag{4.3}$$

Are the preferences represented by (4.3) single-peaked?

The answer depends on the shape of the average labor supply function $L(\tau)$. If this function were concave, $U(\tau; \alpha^i)$ would be everywhere strictly concave, and therefore satisfy the single-peakedness assumption. However, this function could be sufficiently convex that

$U^i(\tau; \alpha^i)$ could have multiple peaks (multiple local maxima). As a result, preferences may not be single peaked.

But it is straightforward to verify that (4.3) satisfies the single-crossing property. Therefore, we can apply the median-voter theorem, and the outcome of party competition between the two parties will be τ^m such that

$$\tau^m = \arg\max_{\tau} U(\tau; \alpha^m)$$

Hence, we have

$$L'(\tau^m)\left[1 - h'(\alpha - L(\tau^m))\right] - (\alpha^m - \alpha) = 0 \tag{4.4}$$

If the mean is greater than the median, as we should have for a skewed distribution of income, it must be the case that $\alpha^m - \alpha < 0$ (that is median productivity must be less than mean productivity). This implies that $\tau^m > 0$—otherwise, (4.4) would be satisfied for a negative tax rate, and we would be at a corner solution with zero taxes (unless negative tax rates, i.e., subsidies, were allowed).

Now imagine a change in the distribution of α such that the difference between the mean and the median widens. From the above first-order condition, this'll imply that the equilibrium tax rate τ^m increases.

This is the foundation of the general presumption that greater inequality (which is generally, but not always, associated with a widening gap between the mean and the median) will lead to greater taxation to ensure greater redistribution away from the mean towards the median. We will see examples below where inequality and the gap between the median in the mean may be unrelated.

Notice also that greater inequality in this model leads to greater "inefficiency" of policy. Why is this? The reason is only weakly related to the logic of redistribution, but more to the technical assumptions that have been made. In order to obtain single-peaked preferences, we had to restrict policy to a single dimensional object, the linear tax rate. Moreover, is this "inefficiency" the same as *Pareto suboptimality?* The answer is clearly no. This is a general feature of static voting models. Equilibria are generally constraints are at the optimal (constrained by the availability of fiscal policies).

Imagine, instead, that different taxes can be applied to different people. Then, redistribution does not necessitate distortionary taxation. But in this case, preferences will clearly be non-single-peaked—agent i particularly dislikes policies that tax him a lot, and likes policies that tax agents j and k a lot, where as agent j likes policies that tax i and k a lot, and so on. This is the reason why these kind of models cannot be analyzed using the MVT. But this obviously does not make the assumptions here “reasonable”.

4.8. Probablilistic Voting

The above analysis discussed situations in which we could apply median voter theorems without problems of cycling. As noted before, many situations do not fall into this category. In these situations, the payoff functions of parties are discontinuous in their policy promises, and a Nash equilibrium often fails to exist in the party competition game.

One way of dealing with the situations is to extend the standard voting models to “probabilistic” voting models, where there is enough individual heterogeneity that the payoff functions of different parties are “smooth” in policy choice, ensuring the existence of an equilibrium.

4.8.1. An example of cycling and nonexistence. Consider redistribution of a fixed amount among three (groups of) voters, $i = 1, 2, 3$, with each group of equal size.Assume that each group has preferences

$$U(p) = u\left(p^i\right),$$

where $p = \left(p^1, p^2, p^3\right)$, and p^i is a nonnegative transfer to group i out of a fixed budget normalized to unity, so that

$$\sum_{i=1}^{3} p^i = 1.$$

Also assume that the common utility function $u(\cdot)$ is strictly monotonic. There are two parties offering policies in order to maximize their probability of coming to power Now it is clear that there will be cycling and nonexistence here. To see this, note the following:

- A policy will be the winner if it gets votes from 2 agents.
- Now take a winning policy (p_1, p_2, p_3) where without any loss of generality suppose that $p_1 > 0$. Then the following policy will always beat this winning policy $(p_1 - 2\varepsilon, p_2 + \varepsilon, p_3 + \varepsilon)$, proving that there will always be cycling.

[Another way of seeing this example is that in this game the core is empty]

4.8.2. Simple Model of Probabilistic Voting. Let the society consist of G distinct groups, with a continuum of voters within each group having the same economic characteristics and preferences. As in the Downsian model, there is electoral competition between two parties, A and B, and let π_P^g be the fraction of voters in group g voting for party P where

$P = A, B$, and let λ^g be the share of voters in group g and naturally $\sum_{g=1}^{G} \lambda^g = 1$. Then, the expected vote share of party P is

$$\pi_P = \sum_{g=1}^{G} \lambda^g \pi_P^g.$$

In our analysis so far, all voters in group g would have cast their votes identically (unless they were indifferent between the two parties). The idea of probabilistic voting is to smooth out this behavior by introducing other considerations in the voting behavior of individuals. Put differently, probabilistic voting models will add "noise" to equilibrium votes, smoothing the behavior relative to models we analyzed so far. In particular, suppose that individual i in group g has the following preferences:

$$\tilde{U}_i^g(p, P) = U^g(p) + \tilde{\sigma}_i^g(P) \tag{4.5}$$

when party P comes to power, where p is the vector of economic policies chosen by the party in power. Suppose that $p \in \mathcal{P} \subset \mathbb{R}^K$, where K is a natural number, possibly greater than 1. Thus $p \equiv \left(p^1, ..., p^K\right)$ is a potentially multidimensional vector of policies. In addition, $U^g(p)$ is the indirect utility of agents in group g as before (previously denoted by $U(p; \alpha_i)$ for individual i) and captures their economic interests. In addition, the term $\tilde{\sigma}_i^g(P)$ captures the non-policy related benefits that the individual will receive if party P comes to power. The most obvious source of these preferences would be ideological. So this model allows individuals within the same economic group to have different ideological preferences.

Let us normalize $\tilde{\sigma}_i^g(A) = 0$, so that

$$\tilde{U}_i^g(p, A) = U^g(p), \text{ and } \tilde{U}_i^g(p, B) = U^g(p) + \tilde{\sigma}_i^g \tag{4.6}$$

In that case, the voting behavior of individual i can be represented as

$$v_i^g(p_A, p_B) = \begin{cases} 1 \text{ if } U^g(p_A) - U^g(p_B) > \tilde{\sigma}_i^g \\ \frac{1}{2} \text{ if } U^g(p_A) - U^g(p_B) = \tilde{\sigma}_i^g \\ 0 \text{ if } U^g(p_A) - U^g(p_B) < \tilde{\sigma}_i^g \end{cases}, \tag{4.7}$$

where $v_i^g(p_A, p_B)$ denotes the probability that the individual will vote for party A, p_A is the platform of party A and p_B is the platform of party B, and if an individual is indifferent between the two parties (inclusive of the ideological benefits), he randomizes his vote.

Let us now assume that the distribution of non-policy related benefits $\tilde{\sigma}_i^g$ for individual i in group g is given by a smooth cumulative distribution function H^g defined over $(-\infty, +\infty)$, with the associated probability density function h^g. The draws of $\tilde{\sigma}_i^g$ across individuals are independent. Consequently, the vote share of party A among members of group g is

$$\pi_A^g = H^g(U^g(p_A) - U^g(p_B)).$$

Furthermore, to simplify the exposition here, suppose that parties maximize their expected vote share. In this case, party A sets this policy platform p_A to maximize:

$$\pi_A = \sum_{g=1}^{G} \lambda^g H^g(U^g(p_A) - U^g(p_B)). \tag{4.8}$$

Party B faces a symmetric problem and maximizes π_B, which is defined similarly. In particular, since $\pi_B = 1 - \pi_A$, party B's problem is exactly the same as minimizing π_A. Equilibrium policies will then be determined as the Nash equilibrium of a (zero-sum) game where both parties make simultaneous policy announcements to maximize their vote share. Let us first look at the first-order condition of party A with respect to its own policy choice, p_A, taking the policy choices of the other party, p_B, as given. This is:

$$\sum_{g=1}^{G} \lambda^g h^g(U^g(p_A) - U^g(p_B))DU^g\,(p_A) = 0,$$

where $DU^g\,(p_A)$ is the gradient of $U^g\,(\cdot)$ given by

$$DU^g\,(p_A) = \left(\frac{\partial U^g\,(p_A)}{\partial p_A^1}, ..., \frac{\partial U^g\,(p_A)}{\partial p_A^K}\right)^T,$$

with p_A^k corresponding to the kth component of the policy vector p_A. Since the problem of party B is symmetric, it is natural to focus on pure strategy symmetric equilibria, where $p_A = p_B = p^*$, and thus $U^g(p_A) = U^g(p_B)$. Consequently, symmetric equilibrium policies, announced by both parties, must be given by

$$\sum_{g=1}^{G} \lambda^g h^g(0)DU^g\,(p^*) = 0. \tag{4.9}$$

It is now straightforward to see that eq. (4.9) also corresponds to the solution to the maximization of the following weighted utilitarian social welfare function:

$$\sum_{g=1}^{G} \chi^g \lambda^g U^g\,(p)\,, \tag{4.10}$$

where $\chi^g \equiv h^g(0)$ are the weights that different groups receive in the social welfare function. This analysis therefore establishes:

THEOREM 4.8. ***(Probabilistic Voting Theorem)*** *Consider a set of policy choices $\mathcal{P}$, let $p \in \mathcal{P} \subset \mathbb{R}^K$ be a policy vector and let preferences be given by (4.6), with the distribution function of $\tilde{\sigma}_i^g$ as H^g. Then, if a pure strategy symmetric equilibrium exists, equilibrium policy is given by p^* that maximizes (4.10).*

The important point to note about this result is its seeming generality: as long as a pure strategy symmetric equilibrium in the party competition game exists, it will correspond to a maximum of some weighted social welfare function. This generality is somewhat exaggerated, however, since such a symmetric equilibrium does not always exists. In fact, conditions to guarantee existence of pure strategy symmetric equilibria are rather restrictive. Let us next turn to a discussion of existence after strategy equilibria in probablilistic voting games.

The first thing to guarantee is that the symmetric pure strategy equilibrium discussed above corresponds to a local maximum of the payoff to each of the parties. As usual, this requires that each party's payoff is locally concave with respect to its own strategy. The condition for this is that the matrix $B(0,p^*)$ given by

$$\begin{aligned} B(0,p^*) \quad \equiv \quad & \sum_{g=1}^{G} \lambda^g h^g(0) D^2 U^g(p^*) \\ & + \sum_{g=1}^{G} \lambda^g \frac{\partial h^g(0)}{\partial x} (DU^g(p^*)) \cdot (DU^g(p^*))^T \end{aligned}$$

is negative semidefinite. Since this is difficult to check without knowing what p^*, the following "sufficient condition" might be useful:

$$B^g(x,p) \equiv h^g(x) D^2 U^g(p) + \left| \frac{\partial h^g(x)}{\partial x} \right| (DU^g(p)) \cdot (DU^g(p))^T \tag{4.11}$$

is negative definite for any x and p, and each g. Here, we require negative definiteness of each part of the sum (since the sum of negative definite matrices are negative definite) and also that the appropriate matrix should be negative definite not only at 0, which corresponds to the point where the two parties choose the same strategy, but also throughout, so that the payoff function of each party is globally concave. This condition is in fact very restrictive and essentially amounts to requiring that the distribution function for "noise" in the voting process should be approximately uniform for each group, that is, H^g is uniform for each g (and that U^g is concave). Naturally, if we know that U^g's is highly concave, then a certain amount of deviation from the H^g's being uniform can be allowed.

THEOREM 4.9. ***(Pure Strategy Existence)*** *Suppose that (4.11) holds. Then in the probabilistic voting game, a pure strategy equilibrium always exists.*

To see why (4.11) is the right type of condition and also to demonstrate in greater detail why it corresponds to H^g's being uniform, let us suppose that p is one dimensional and look at the second-order condition of party A. Recall that their first-order condition written fully in the one-dimensional policy case is

$$\sum_{g=1}^{G} h^g(U^g(p_A) - U^g(p_B)) \frac{\partial U^g(p_A)}{\partial p} = 0$$

The second-order condition is:

$$\sum_{g=1}^{G} h^g(U^g(p_A) - U^g(p_B))\frac{\partial^2 U^g(p_A)}{\partial p^2} + \sum_{g=1}^{G} \frac{\partial h^g(U^g(p_A) - U^g(p_B))}{\partial x}\left(\frac{\partial U^g(p_A)}{\partial p}\right)^2 < 0$$

Again looking at each group's utility separately, this requires

$$-\frac{\partial^2 U^g(p_A)/\partial p^2}{\left(\partial U^g(p_A)/\partial p\right)^2} > \frac{\partial h^g(U^g(p_A) - U^g(p_B))/\partial x}{h^g(U^g(p_A) - U^g(p_B))}$$

for all g. At the same time, this point must also be a best response for party B, so by the same arguments,

$$-\frac{\partial^2 U^g(p_B)/\partial p^2}{\left(\partial U^g(p_B)/\partial p\right)^2} > -\frac{\partial h^g(U^g(p_A) - U^g(p_B))/\partial x}{h^g(U^g(p_A) - U^g(p_B))}.$$

A sufficient condition for both of these inequalities to be satisfied is

$$\sup_x \frac{\left|\partial h^g(x)/\partial x\right|}{h^g(x)} \leq \inf_p \left|\frac{\partial^2 U^g(p)/\partial p}{\partial U^g(p)/\partial p}\right| \text{ for all } g.$$

Naturally, unless the right-hand side is significantly greater than one, the left-hand side has to be equal to 0, thus each H^g has to be uniform.

Naturally, mixed strategy equilibria are easier to guarantee (for example, they are immediate from Glicksberg's Theorem). However, mixed strategy equilibria also exist without probablilistic voting. So probabilistic voting has not really helped too much with the existence problem.

Theorem 4.10. ***(Mixed Strategy Existence)*** *In the probabilistic voting game, a mixed strategy equilibrium always exists.*

So far, the objective of the parties was to maximize vote share. How does the analysis change when their objective is to maximize the probability of coming to power?

4.8.3. Application: the power of the middle class. Let us will now use the probablilistic voting model to show how a unique equilibrium will emerge in the redistribution problem with three groups discussed above, and also show how the "middle class" may get disproportionate redistribution if they are perceived as the "swing voters". Again assume that the population consists of three distinct groups, $g = R, M, P$, representing the rich, the middle class, and the poor, respectively.

Each group has preferences

$$U(p) = u\left(p^i\right),$$

p^i is a nonnegative transfer out of a fixed budget normalized to unity, and $u(\cdot)$ is the strictly monotonic utility function common to all groups. The population share of group g is λ^g, with

$\sum_{g=1}^{3}\lambda^g = 1$. The relevant policy vector is again a vector of redistributions $p = (p^1, p^2, p^3)$, with the budget constraint

$$\sum_{g=1}^{3}\lambda^g p^g = 1.$$

At the time of the elections, voters base their voting decision both on the economic policy announcements and on the two parties' ideologies relative to the realization of their own ideology.

As before, the two parties A and B simultaneously announce their policies, and voters vote between the two parties. At the time of policy announcement, the parties do not know the realization of the stochastic elements. At the time of voting, each individual knows the realization of its own preferences. In particular, voter i in group g prefers party A if

$$U^g(p_A) > U^g(p_B) + \sigma^{ig} + \delta.$$

Here, p_A is a policy vector of party A, and p_B is the policy vector of party B. σ^{ig} is the individual-specific ideology parameter that can take on negative as well as positive values. This parameter measures voter i's individual ideological bias toward candidate B. A positive value of σ^{ig} implies that voter i has a bias in favor of party B, whereas voters with $\sigma^{ig} = 0$ are ideologically neutral, that is, they care only about the economic policy.

Let us assume that this parameter for each group g has group-specific uniform distribution on

$$\left[-\frac{1}{2\phi^g}, \frac{1}{2\phi^g}\right].$$

These distributions thus have density ϕ^g, and none of the three groups has any bias towards one of the candidates/parties. Also assume, for simplicity, that

$$\sum_g \lambda^g \phi^g = 1$$

The parameter δ measures the average (relative) popularity of candidate B in the population as a whole, and also can be positive or negative. Assume that it has a uniform distribution on

$$\left[-\frac{1}{2\psi}, \frac{1}{2\psi}\right].$$

The "indifferent" voter in group g will be a voter whose ideological bias, given the candidates' platforms, makes him indifferent between the two parties. This indifferent voter is defined as

$$\sigma^g = U^g(p_A) - U^g(p_B) - \delta.$$

All voters in group g with $\sigma^{ig} \leq \sigma^g$ prefer party A. Therefore, party A's actual vote share is

$$\pi_A = \sum_g \lambda^g \phi^g \left(\sigma^g + \frac{1}{2\phi^g}\right).$$

Notice that σ^g depends on the realized value of δ, the vote share π_A is also a random variable.

From the perspective of both parties, at the time of policy choice, the electoral outcome is thus a random event, related to the realization of δ. Party A's probability of winning is then

$$\mathbb{P}_A = \operatorname*{Prob}_{\delta}\left[\pi_A \geq \frac{1}{2}\right] = \frac{1}{2} + \psi\left[\sum_{g=1}^{3} \lambda^g \phi^g \left[U^g(p_A) - U^g(p_B)\right]\right], \tag{4.12}$$

Party B wins with probability $1 - \mathbb{P}_A$.

The two parties will choose p^A and p^B to maximize their *probabilities of coming to power.*

It should be clear that the unique equilibrium will involve both parties converging to the same platform, $p^* = (p_1^*, p_2^*, p_3^*)$ (since the two parties are facing exactly the same concave optimization problem—concavity follows from the concavity of the U^g functions).

To characterize this equilibrium policy vector, let us write (4.12) more explicitly, also assuming that party B has announced the equilibrium policy $p_B = p^*$:

$$\mathbb{P}_A = \operatorname*{Prob}_{\delta}\left[\pi_A \geq \frac{1}{2}\right] = \frac{1}{2} + \psi\left[\begin{array}{c} \lambda_1\phi_1\left(u\left(p_{A,1}\right) - u\left(p_1^*\right)\right) \\ +\lambda_2\phi_2\left(u\left(p_{A,2}\right) - u\left(p_2^*\right)\right) \\ +\lambda_3\phi_3\left(u\left(p_{A,3}\right) - u\left(p_3^*\right)\right) \end{array}\right], \tag{4.13}$$

Party A will maximize (4.13) subject to the constraint that

$$\lambda_1 p_{A,1} + \lambda_2 p_{A,2} + \lambda_2 p_{A,3} = 1. \tag{4.14}$$

This maximization problem has the following three first-order conditions

$$\begin{array}{c} \phi_1 u'\left(p_{A,1}\right) = \eta \\ \phi_2 u'\left(p_{A,2}\right) = \eta \\ \phi_3 u'\left(p_{A,3}\right) = \eta \end{array} \tag{4.15}$$

where η is the Lagrangean multiplier on (4.14).

There are two important implications that follow from (4.15):

(1) There now exists a unique equilibrium policy vector that both parties converge to, despite the fact that preferences are not single-peaked.

(2) Groups that have high ϕ, that is groups that are more sensitive to policy and have weaker ideological bias, act as swing voters, and obtain more redistribution. This could be a possible explanation for why middle classes often obtain more redistribution in democracies [...why should middle class voters be more sensitive to economic

policies? Perhaps less party ideology in a world where many parties (used to) represent capital or labor?....]

4.9. Models with Party/Candidate Ideology

Finally, let us discuss models with "partisan politics". So far the parties did not have a preference over policies, or at least they could commit to their policy platform. Another possibility is to have parties that represent existing constituencies, and unable to commit to particular policies, so that their "preferences" will determine post-election policies. Such an approach may be more in line with theories of political economy based on class conflict for example. Moreover, a situation with partisan parties helps us in dealing with situations of nonexistence and cycling as well.

4.9.1. Electoral competition with partisan politics. To analyze the implications of electoral competition, imagine that there is a single dimension of policy, again denoted p from a convex and compact set $\mathcal{P}$, and let there be two parties A and B.

The two parties have the following objective functions

$$\begin{aligned} \text{Party } A &: \max_{p_A} \mathbb{P}(p_A, p_B)\,(Q + W_A(p_A)) + (1 - \mathbb{P}(p_A, p_B))W_A(p_B) \\ \text{Party } B &: \max_{p_B} (1 - \mathbb{P}(p_A, p_B))\,(Q + W_B(p_B)) + \mathbb{P}(p_A, p_B)W_B(p_A), \end{aligned} \tag{4.16}$$

where $W_A(p)$ and $W_B(p)$ denote the 'utility functions' of parties A and B, and $Q \geq 0$ is a rent from being in office. So parties are now maximizing their 'expected utility' taking into account the voting behavior of the citizens as summarized by the function $\mathbb{P}(p_A, p_B)$, which denotes the probability that party A will come to power when the two parties adopt policies p_A, p_B. This expected utility consists of their ideological preferences over policies that are implemented, and the rent from coming to office.

To start with, we consider the case where $\mathbb{P}(p_A, p_B)$ is given by (4.2) above, i.e.,

$$\mathbb{P}(p_A, p_B) = \begin{cases} 1 \text{ if } U^M(p_A) > U^M(p_B) \\ \frac{1}{2} \text{ if } U^M(p_A) = U^M(p_B) \\ 0 \text{ if } U^M(p_A) < U^M(p_B) \end{cases} \tag{4.17}$$

where U^M is the utility of the median voter. This would be the case for example, when preferences are single peaked and there are no ideological considerations on the side of the voters.

Moreover, suppose that the utility functions of the parties are smooth and strictly quasi-concave (i.e., single peaked), with ideal policies p_A^* and p_B^*, i.e.,

$$p_A^* = \arg\max_{p_A} W_A(p_A) \quad \text{and} \quad p_B^* = \arg\max_{p_B} W_B(p_B).$$

In other words, $W_A'(p_A^*) = 0$ and $W_B'(p_B^*) = 0$.

Finally, as before assume that both parties choose their policies (policy platforms) simultaneously. Therefore, the predictions of this model can be summarized by the corresponding *Nash Equilibrium*, where each party chooses the policy which maximizes its utility given the policy of the other party. Nash Equilibrium policy platforms, (p_A^N, p_B^N), will satisfy the following conditions:

$$p_A^N \in \arg\max_{p_A} \; \mathbb{P}(p_A, p_B^N)\left[Q + W_A(p_A)\right] + \left[1 - \mathbb{P}(p_A, p_B^N)\right] W_A(p_B^N),$$

and, simultaneously,

$$p_B^N \in \arg\max_{p_B} \; \left[1 - \mathbb{P}(p_A^N, p_B)\right]\left[Q + W_B(p_B)\right] + \mathbb{P}(p_A^N, p_B) W_B(p_A^N).$$

Intuitively these conditions state that in a Nash Equilibrium, taking p_B^N as given, p_A^N should maximize party A's expected utility. At the same time it must be true that, taking p_A^N as given, p_B^N should maximize B's expected utility.

The problem in characterizing this Nash equilibrium is that the function $\mathbb{P}(p_A, p_B)$ in (4.2), is not differentiable. Nevertheless, it is possible to establish the following proposition, which was first established by Calvert (1985), and shows that even with partisan politics, there will be policy convergence, and this convergence will typically be to the most preferred point of the median voter.

THEOREM 4.11. ***(Policy Convergence with Partisan Politics)*** *Consider the partisan politics model described above, with ideal points of the two parties p_A^* and p_B^*, and the ideal point of the median voter corresponding to p^M. Suppose also that the probability of party A winning the election is given by $\mathbb{P}(p_A, p_B)$, as in (4.2).*

- *If $p_A^* \geq p^M \geq p_B^*$ or if $p_B^* \geq p^M \geq p_A^*$, then the unique equilibrium involves $p_A = p_B = p^M$ and each party wins the election with probability one half.*
- *Otherwise, the unique equilibrium involves $p_A = p_B = p^M$ if $Q > 0$, and if $Q = 0$, then $p_A = p_B = p_A^*$ when $U^M(p_A^*) > U^M(p_B^*)$ and $p_A = p_B = p_B^*$ when $U^M(p_A^*) < U^M(p_B^*)$.*

PROOF. **(Sketch):** Consider the first case where the preferences of the median voter are intermediate with respect to the ideal points of the two parties. Consider first the situation in which $p_A = p^M \neq p_B$. Then, we have that $\mathbb{P}(p_A, p_B) = 1$, and party A is winning for sure. The utility of party B is given by: $W_B(p^M)$. Now imagine a deviation by party B to $p_B = p^M$. Now we will have that $\mathbb{P}(p_A, p_B) = 1/2$, so the utility of party B changes to $Q/2 + W_B(p^M) > W_B(p^M)$, hence the deviation is profitable, and $p_A = p^M \neq p_B$ cannot be an equilibrium (in the case where $Q = 0$, the argument is a little different, and now party A can change its policy to something slightly away from p^M towards its ideal point p_A^* and still win the election and now implement a policy closer to its preferences).

Similarly, consider now a situation where $p_A \neq p^M \neq p_B$, and suppose without loss of any generality that $p_A^* > p^M > p_B^*$ and $U^M(p_A) > U^M(p_B)$, so that we again have $\mathbb{P}(p_A, p_B) = 1$. It is clear that we must have $p_A \geq p^M$, otherwise, party A could find a policy p_A' such that $U^M(p_A') > U^M(p_B)$ and $p_A' \geq p^M$ preferable to any $p_A \in (p^M, p_B)$. But then party B is obtaining utility $W_B(p_A)$, and by changing its policy to $p_B = p^M$ it will obtain utility $Q + W_B(p^M)$ if $p_A > p^M$ and $Q/2 + W_B(p^M)$ if $p_A = p^M$. By the fact that $p_A \geq p^M$ both of these are greater than its initial utility, $W_B(p_A)$, hence, no policy announcements with $p_A \neq p^M \neq p_B$ can be an equilibrium. Therefore, the equilibrium must have $p_A = p_B = p^M$, i.e., convergence to the median. Intuitively, the median voter's ideal point is preferable to each party relative to the other party's ideal point, and moreover increases their likelihood of coming to power. Therefore, no policy other than the median voter's ideal point can ever be implemented in equilibrium.

Next let us consider the case where $p_B^* > p_A^* > p^M$ (other configurations give analogous results) Now, suppose that we have $p_A = p_A^*$. What should party B do? Clearly, any policy $p_B > p_A^*$ will lose the election. $p_B = p_A^*$ will win the election with probability 1/2 and is preferable. But in fact party B can do better. It can set $p_B = p_A^* - \varepsilon$ which is closer to the median voter's preferences, and by the fact that voters' preferences are single peaked, this is preferable to p_A^*, and therefore will win the election for party B. Although this policy is worse for party B than p_A^* (since $p_B^* > p_A^*$), for ε small enough, the difference is minuscule, whereas the gain in terms of the rent from coming to power is first-order. This argument only breaks down when $Q = 0$, and in this case, the best thing that party B can do is to offer $p_B = p_A^*$ (or any other policy $p_B > p_A^*$ for that matter, since it does not care about coming to power, and in either case, p_A^* will be the equilibrium policy). □

Therefore, the basic result is that although there can be exceptions when there are no rents from coming to office and both parties have the same type of ideological bias, there are very strong forces towards policy convergence. As the discussion will illustrate, the source of these powerful forces is equation (4.17), which implies that the policy that comes closer to the median voter's preferences will win relative to another policy.

We therefore see that policy convergence to the median is a rather strong force, but there can be exceptions especially when rents from coming to power are nonexistent.

Nevertheless, the above results depend crucially on the form of the $\mathbb{P}(p_A, p_B)$ function, which created very strong returns to being close to the most preferred point of the median voter. We saw in the last section how in the presence of ideological considerations on the side of the voters, $\mathbb{P}(p_A, p_B)$ can become a continuous function. If that's the case, then policy convergence will break down. To see this, suppose that $\mathbb{P}(p_A, p_B)$ is a continuous and differentiable function, and suppose that it reaches its maximum for each party at p^M (i.e., being closer to the median voter's preferences is still beneficial in terms of the probability of being elected—that we make this point which maximizes winning probabilities the median voter's ideal point is simply a normalization without any consequences). In that case, the Nash equilibrium of the policy competition game between the two parties will be a pair of policies p_A^N, p_B^N such that the following first-order conditions hold:

$$\begin{aligned}\frac{\partial \mathbb{P}(p_A^N, p_B^N)}{\partial p_A}\left(W_A(p_A^N) + Q - W_A(p_B^N)\right) + \mathbb{P}(p_A^N, p_B^N) W_A'(p_A^N) &= 0, \\ -\frac{\partial \mathbb{P}(p_A^N, p_B^N)}{\partial p_B}\left(W_B(p_B^N) + Q - W_B(p_A^N)\right) - \mathbb{P}(p_A^N, p_B^N) W_B'(p_B^N) &= 0.\end{aligned} \tag{4.18}$$

The first term on both lines is the gain in terms of the utility of winning times the change in the probability of winning in response to a policy change, and the second term is the product of the current probability of winning times the gain in terms of improvements in the party's utility because of the policy change. When these two marginal effects are equal to each other, each party is playing its best response. When both parties are playing their best responses, we have the Nash equilibrium.

Inspection of these first-order conditions will reveal that both parties offering the ideal point of the median voter, i.e., $p_A = p_B = p^M$, is typically not an equilibrium, despite the fact that the probability of winning is higher for both parties at this point. The reason is that now parties are trading off ideological benefits coming from their partisan views against the probability of winning. To see this, suppose that $p_A = p_B = p^M$, and party A deviates and offers a policy slightly away from p^M and closer to its ideal point, p_A^*. The first effect of this on party A's utility is a loss

$$\frac{\partial \mathbb{P}(p^M, p^M)}{\partial p_A} Q < 0,$$

where the negative sign follows from the fact that any move away from p^M necessarily reduces the probability of winning by definition. But nevertheless, as long as $\mathbb{P}(p^M, p^M) W_A'(p^M)$ is large, i.e., party A is sufficiently ideological, the deviation can be profitable, i.e., we can have

$$\mathbb{P}(p^M, p^M) W_A'(p^M) > -\frac{\partial \mathbb{P}(p^M, p^M)}{\partial p_A} Q.$$

In this case, we will get non-convergence in the two parties' platforms, and equilibrium policy will be different from the preferences of the median voter, and will depend on the preferences of the parties' ideologies.

Moreover, suppose that both parties are ideologically biased in one direction relative to the populace. In particular, suppose that $p_B^* > p_A^* > p^M$. Now it is easy to see that $p_A = p_B = p_A^*$ can be an equilibrium because by shifting its platform to $p_B = p_A^* - \varepsilon$, party will only increase its chance of winning the election continuously, since $\mathbb{P}(p_A, p_B)$ is a continuous and smooth function. Then we have:

Theorem 4.12. ***(Policy Non-Convergence with Partisan Politics and Probabilistic Voting)*** *Suppose that $\mathbb{P}(p_A, p_B)$ is a continuous and smooth function because of probabilistic voting (or other reasons). Then there can exist an equilibrium where $p_A \neq p_B \neq p^M$ even if $Q > 0$.*

The reason why this proposition is important for us is that it suggests that certain groups can be quite powerful in democratic politics if they can manage to control the ideological leanings of the parties.

4.10. Commitment and Convergence

An important assumption so far is that parties announce policy platforms and then they can commit to the policies that they have announced in those platforms. This way, parties could basically compete by varying the policies that they will implement when in office. However, as emphasized by Alesina (1988), the assumption of commitment is not necessarily plausible. In these one-shot models, what is there to stop the politicians from changing policies to their ideal point once they come the power? Nothing. There is no potential punishment (there would have been some punishment if we were in the world with repeated elections, but this is beyond the scope of our treatment here).

Therefore, it is important to see what happens when we remove this commitment assumption. So consider the model of the last section, but assume that parties can choose whichever policy they like when they come to office. Suppose also that $\mathbb{P}(p_A, p_B)$ is given by (4.2). This means that announcements before the election are nothing other than cheap talk, and in the subgame perfect equilibrium, voters will realize that once they come to power, parties will implement their ideal points. Therefore, they will simply compare $U^i(p_A^*)$ and $U^i(p_B^*)$, and vote for whichever party has an ideal point closer to their ideal point. The result will be that the party with an ideal point closer to that of the median voter will win. We therefore have that

Theorem 4.13. ***(Policy Non-Convergence With Partisan Politics and No Commitment)*** *Suppose that there is no commitment to policy platforms in the above model of partisan politics. Then in the unique equilibrium, we have that: if $U^M(p_A^*) > U^M(p_B^*)$ party A comes to power with probability 1 and the equilibrium policy is p_A^*; if $U^M(p_B^*) > U^M(p_A^*)$ party B comes to power with probability 1 and the equilibrium policy is p_B^*; and if $U^M(p_A^*) = U^M(p_B^*)$ each party comes the power with probability 1/2 and the equilibrium policy is $p_A^* = p_B^*$.*

In this model of partisan politics without commitment, we see that parties' policy preferences matter more. This implies that control of the political agenda and parties internal structures may now be much more important in determining equilibrium policies.

Even though the analysis in this section shows that the way that some groups, in particular, richer segments of the society, may influence equilibrium policies through political capture is quite different from their effects in the probabilistic voting model and lobbying model, the overall result is the same: certain groups can be more powerful in democratic politics than suggested by the basic model of Downsian political competition. This result is important for our analysis of the emergence and consolidation of democracy below, but what matters is this general qualitative tendency. Therefore, in the rest of the lectures, we will use the formalization of the probabilistic voting and lobbying models, which can be summarized by saying that equilibrium policies will be such that they maximize a weighted utilitarian social welfare function

$$\sum_{i=1}^{I} \chi^i U^i(p),$$

and in this formulation, the parameters χ^i's denote the political power of the groups. Given the analysis in this section, however, we will think of these parameters more generally as resulting from political capture as well as lobbying and probabilistic voting.

4.10.1. Entry of parties. Besley and Coate study an extended form of these partisan politics models with entry decisions of parties endogenized. For example, imagine that in the above game each of the three groups can organize their own party at some (small) cost ε. What is the equilibrium of the game?

Besley and Coate show that this class of games always has an equilibrium (though the equilibria may be a mixed strategy one). Typically, however, there are many equilibria. For example, suppose that if there is a tie in votes, there will be randomization over all the parties with equal votes. Then, in the above model, there is a unique equilibrium in which all three groups organize their own parties, and each agent votes for its own party.

Alternatively, consider a situation in which group 1 is larger than the other two groups, which are of equal size, and group 1 is less than half of the population. Now there exists an equilibrium in which only group 1 organizes as a party, and group 1 agents obtain all the redistribution. If group 2 deviates and organizes a party, in the continuation game (in the subgame following group 2's entry) group 3 members randomize between the two parties with equal probability (this is optimal for them, since both parties will give them zero redistribution). Therefore, since there are many agents within each group, the party representing group 1 will obtain half of the votes from group 3 and win the election. Therefore, it is a best response for group 2 not to enter.

However, there is also another equilibrium in which groups 2 and 3 organize, and group 1 does not enter. Members of group 2 and 3 vote for their own party, and members of group 1 randomize between the two parties. A deviation by group 1 is not profitable because in the continuation game following group 1's entry, all group 2 agents vote for the party of group 3. This is optimal for them since in either case they're getting zero redistribution.

4.10.2. Repeated interactions. Another possible set of issues arises when parties have ideologies, and are competing dynamically. To discuss the set of issues here suppose that two parties A and B are competing for election every period, and these parties have utilities given by

$$-\sum_{t=0}^{\infty}\beta^{t}\left(p_{t}-p^{A}\right)^{2} \text{ and } -\sum_{t=0}^{\infty}\beta^{t}\left(p_{t}+p^{B}\right)^{2}$$

where p_t is the policy choice at time t.

Suppose that $p=0$ is the Condorcet winner in every period, and assume that

$$p^{A}>p^{B}>0$$

Can we design some type of trigger strategies such that party A is convinced to choose $p=0$ all the time? Consider the following voters strategies: All voters with $p\left(\alpha^{i}\right)\geq 0$, which form a majority, vote for party A now at time $t=0$, and keep on voting for party A at time $t=k$ as long as $p_{k-j}=0$ for all $j\leq k$. If $p_{k-j}\neq 0$ for some $j\leq k$, all voters with bliss point $p\left(\alpha^{i}\right)<\varepsilon$ for some small ε vote for party B in all future elections.

First note that once it comes to power, party B will always choose its most preferred policy, $p=-p^{B}$, since it is under no constraints. Moreover, since $p^{A}>p^{B}>0$, more than half of the voters will support party B against party A when both of them are playing their most preferred policy.

Then, if party A adopts the policy of $p = 0$ in all periods, its utility is

$$U_C^A = -\sum_{t=0}^{\infty} \beta^t \left(p^A\right)^2 = \frac{-\left(p^A\right)^2}{1-\beta}$$

If it deviates to its most preferred policy, its utility this period is 0, but from the next period onwards, the equilibrium policy will be $p = -p^B$, so the utility to deviating is

$$U_D^A = -\sum_{t=1}^{\infty} \beta^t \left(p^B + p^A\right)^2 = \frac{-\beta\left(p^B + p^A\right)^2}{1-\beta}$$

Therefore, despite party ideologies, the Condorcet winner will be implemented as long as

$$U_C^A \geq U_D^A,$$

that is, as long as

$$\beta\left(p^B + p^A\right)^2 \geq \left(p^A\right)^2$$

This condition will be satisfied if β is high enough, that is, if political parties are patient, or if p^B is sufficiently large. This last condition is interesting, since it emphasizes that greater disagreement among the parties may be useful in forcing one of the parties to adopt the policies desired by the voters.

Question: why did I choose party A as the one to been power and implement the Condorcet winner policy of voters?

Question: how would you support an equilibrium in which party B is induced to choose the Condorcet winner policy?

This discussion is more like a sketch. It looks for rather simple strategies for different parties and stationary equilibria. Later, we will analyze a model with a similar flavor more fully, allowing for unrestricted strategies and focusing on subgame perfect equilibria along the Pareto frontier.

4.11. References

(1) Acemoglu, Daron and Robinson, James (2006) "Chapter 4: Democratic Politics" and "Appendix to Chapter 4: Distribution of Power in Democracy" in *Economic Origins of Dictatorship and Democracy*, Acemoglu, Daron and Robinson, James, eds. Cambridge University Press.

(2) Alesina, Alberto. (1988). "Credibility and Policy Convergence in a Two-Party System with Rational Voters." American Economic Review 78: 796-805.

(3) Alesina, Alberto. (1987). "Macroeconomic Policy in a Two-Party System as a Repeated Game." Quarterly Journal of Economics 102: 651-678.

(4) Austen-Smith, David and Banks, Jeffrey (1999) *Positive Political Theory: Collective Preference*, Ann Arbor; University of Michigan Press.

(5) Besley, Timothy, and Coate, Stephen. (1997). "An Economic Model of Representative Democracy." Quarterly Journal of Economics 112: 85-114.

(6) Lindbeck, Assar, and Weibull, Jorgen W. (1987) "Balanced-Budget Redistribution as the Outcome of Political Competition." *Public Choice* 52: 273-297.

(7) Meltzer, Allan H. and Scott Richard (1981) "A Rational Theory of the Size of Government" *Journal of Political Economy* volume 89, #5, 914-927.

(8) Person, Torsten and Tabellini, Guido (2000) *Political Economics: Explaining Economic Policy*, Cambridge; The MIT Press, Chapters 2, 3 and 6.

(9) Roberts, Kevin (1977) "Voting Over Income Tax Schedules" *Journal of Public Economics*, 8,329-40.

(10) Romer, Thomas (1975) "Individual Welfare, Majority Voting and the Properties of a Linear Income Tax" *Journal of Public Economics*, 7, 163-68.

CHAPTER 5

Dynamic Voting with Given Constituencies

5.1. Myopic Dynamic Voting

The famous application of the median voter theorem to taxation is in Romer (1975), Roberts (1977) and Meltzer and Richard (1981). These authors assumed linear (distortionary) taxes, with proceeds either redistributed lumpsum or invested in public goods, and showed how redistributive taxation can arise as an equilibrium outcome from median-voter politics in static models.

Two papers, Alesina and Rodrik (1994) and Persson and Tabellini (1994), applied this same model, with linear taxes, to an economy with endogenous growth to analyze the relationship between redistribution and growth. However, despite dealing with dynamic situations, these papers focused on static voting, in particular, they basically applied the static median voter theorem. In fact, even though there are many agents in these economies (possibly a continuum), the median voter is *pivotal*, and thus as in the dynamic games we studied above, he or she should recognize how current policy will affect future policies via its effect on the capital stock. However, the static voting framework of these papers ignored this interaction. Here I give a brief overview of their results, and then move to a simplified version of such a model to analyze Markov Perfect Equilibria in a dynamic voting game.

5.1.1. Environment in Myopic Models. Individual preferences are given by:

$$\sum_{t=0}^{\infty} \beta^t \ln c_t$$

I suppress time dependence throughout to save on notation.

Output is given by the aggregate production function

$$y = Ak^{1-\alpha} g^{\alpha} l^{\alpha}$$

where k is capital and g is government investment in infrastructure, and labor l is normalized 1. Government investment g is financed by linear capital taxation at the rate τ. So

$$g = \tau \bar{k} \tag{5.1}$$

where $\bar{k}$ is the average capital stock in the economy. Hence, government's provision of infrastructure creates a Romer-type externality. Assume that capital depreciates fully and that all factor markets are competitive. Then, after substituting from (5.1), the before-tax gross return on capital and the equilibrium wage rate are given as

$$\begin{aligned} r &= (1-\alpha) A\tau^{\alpha} \\ w &= \alpha A\tau^{\alpha}\bar{k} \end{aligned}$$

There is no tax on labor, but the take-home pay per unit of capital is

$$r - \tau = A(1-\alpha)\tau^{\alpha} - \tau$$

Agents are heterogeneous in their holdings of capital (and all have labor $l = 1$).Then the earnings of individual i can be written as

$$y_i = \alpha\tau^{\alpha}\bar{k} + [A(1-\alpha)\tau^{\alpha} - \tau] k_i$$

The first term is labor earnings, while the second term is the capital income of individual i. To make more progress, let θ_i be the the relative capital holding of individual i,

$$\theta_i = k_i/\bar{k},$$

so

$$y_i = [\alpha\tau^{\alpha} + [A(1-\alpha)\tau^{\alpha} - \tau]\theta_i]\bar{k}$$

Logarithmic preferences and capital taxation means that the growth rate of consumption for all agents between two dates is given by

$$\frac{c_{t+1}}{c_t} = \gamma(\tau_t) \equiv \beta[A(1-\alpha)\tau_t^{\alpha} - \tau_t], \tag{5.2}$$

which is only a function of the tax rate. Moreover, this also implies that all individuals accumulate income at the same rate.

How can we think about the preferences of individuals over taxes? It turns out that this is a difficult question. The literature typically takes a very major shortcut. First, they solve for the equilibrium given a constant tax rate τ that applies in all dates. Then they let agents vote once and for all over this tax rate.

Given a constant tax rate τ at all dates, and using the lifetime budget constraint, we obtain that time $t = 0$ consumption level of individual i is given by

$$c_{i0}(\tau) = [\alpha\tau^{\alpha} + (1-\beta)(A(1-\alpha)\tau^{\alpha} - \tau)\theta_i] k_0,$$

where we have included the time subscript to denote the period zero allocation. From (5.2), consumption of each individual grows at the constant rate $\gamma(\tau)$ per period, so the lifetime utility of individual i can be written as:

$$\sum_{t=0}^{\infty}\beta^t \ln c_{it} = \sum_{t=0}^{\infty}\beta^t \ln\left[(\gamma(\tau))^t c_{i0}(\tau)\right] = \frac{\ln c_{i0}(\tau)}{1-\beta}+\frac{\ln\gamma(\tau)}{(1-\beta)^2}. \tag{5.3}$$

Since $c_{i0}(\tau)$ and $\gamma(\tau)$ are concave, each voter's utility is concave, and thus each voter has single-peaked preferences with a unique (political) bliss point. The bliss point of voter i, $\tau(\theta_i)$, is the maximizer of (5.3) and is given as the implicit solution to

$$\frac{c_{i0}'(\tau)}{c_{i0}(\tau)}+\frac{\gamma'(\tau)}{(1-\beta)\gamma(\tau)}=0. \tag{5.4}$$

It can easily be verified by implicit differentiation that $\tau'(\theta_i)<0$, so that the preferred tax rate is decreasing in θ_i. This implies that voters with those with greater capital holdings preferring lower taxes.

Moreover, it can be verified that equation (5.4) can hold only if $\gamma'(\tau)/\gamma(\tau)<0$ and $c_{i0}'(\tau)/c_{i0}(\tau)>0$. Therefore, the tax rate will be above the growth maximizing tax rate τ^* which sets $\gamma'(\tau^*)=0$, which means that a lower value of θ_i corresponds to a higher prefered tax rate and a lower rate of economic growth.

The literature, then, applies the median voter theorem and concludes that the equilibrium tax rate will be the bliss point of the median-ranked voter.

However, this is not very compelling, since we should expect individuals to vote at every instant, and then they may anticipate that their current votes will influence future votes et cetera. This takes us to models of Markov perfect political equilibria.

Before this, we can also mention some important comparative statics. As in the Meltzer-Richards model, the literature typically assumes that the distribution of capital ownership is skewed, so that average income is greater than the income of the median. From this, it concludes that greater inequality will lead to greater taxes. This literature then links greater inequality to lower growth because of its implications for capital taxation.

5.1.2. Inequality and Redistribution. As noted above, the analysis with myopic voting is difficult to interpret. Nevertheless, the insights of these myopic models have been used extensively in the literature. One of the key claims is that these models predict that greater inequality should lead to greater redistribution. Despite these claims in the literature, however, there is no such unambiguous prediction.

More importantly, there is no empirical evidence that greater inequality leads to more distribution. In fact, a major puzzle is why many highly unequal societies do not adopt more redistributive policies. For now, let us delay a discussion on this issue and understand the relationship between inequality and redistribution.

Suppose the economy consists of three groups, upper class, middle class and lower class. All agents within a class have the same income level.

A middle class agent is the median voter, and decides the linear tax rate on incomes. Tax revenues are redistributed lump sum. Let y_m be the income of a middle-class agent, y_l be the income of the lower class agent, and $\bar{y}$ be the average income. Assume that $\bar{y} > y_m$.

Also assume that redistributive taxation at the rate τ has a cost $c(\tau)$ per unit of income. Then, the median voter will maximize

$$(1-\tau) y_m + (\tau - c(\tau)) \bar{y}$$

The first-order condition is:

$$\frac{\bar{y} - y_m}{\bar{y}} = c'(\tau)$$

The left-hand side of this expression, $(\bar{y} - y_m)/\bar{y}$ corresponds to 1 minus the income share of the middle class, and some empirical work, e.g., Perotti, has looked explicitly at the relationship between policy and income share of the middle class.

Now imagine a reduction in y_l and a corresponding increase in y_m such that average income, $\bar{y}$, remains unchanged. This increase in the income share of the middle class will reduce the desired tax rate of the median voter. But in this example, this change in the income distribution corresponds to *greater inequality*. So we have a situation in which greater inequality reduces taxes.

5.2. Dynamic Voting and Markov Perfect Equilibria

The above discussion illustrates why we have to go beyond myopic voting. It turns out that the infinite-horizon model with arbitrary distribution of initial income (or skills) leads to a sufficiently complicated problem when we look at the dynamic linkages. So here I will look at the model of dynamic voting by Hassler, Rodriguez Mora, Storesletten and Zilibotti, which is in a simpler two-period overlapping generation setting, with the main redistributive policy corresponding to the "welfare state". In this model, consistent with the notion of an MPE, agents take into account the implications of current policies on future policies. The simplicity of the framework enables them to derive analytical solutions.

The substantive question motivating the analysis is also interesting: as is well known, the welfare state (WS) was initiated in many countries after the Great Depression of 1929. Why this crisis led to the emergence of the welfare state is a very interesting political economy question and our analysis of endogenous political institutions below will give some clues about this. For now, the question is why, once in place, did the welfare state survive the end of the crisis.

To answer this question in the starkest possible way, the authors assume that the welfare state is "inefficient" in aggregate sense, that is, it creates the standard distortionary effects of redistributive taxation. Then the central question they address are: (1) Why are redistributive institutions so persistent? (2) Will the support for an "inefficient redistribution" driven by ex-post interests (not insurance) be sustained over time, when voters are rational and forward-looking?

The key findings are as follows:

- The welfare state can regenerate its own political constituency, leading to a pattern of persistence, because of dynamic linkages. In particular, redistributive policies create their own constituency.
- But, the welfare state can also breakdown as an equilibrium phenomenon, thus an equilibrium with the welfare state is potentially fragile.

An auxiliary implication:

- An increase in (technological) wage inequality can undermine the support for the welfare state.

5.2.1. The model. 2-period OLG, risk neutral agents, work in both periods of life.

To focus sharply, we assume that individuals are born identical but become "successful" or "unsuccessful".

Young individuals can affect the probability of becoming "successful" by an investment e, at the cost e^2,

- With prob. e the agent becomes type S (Successful) and earns 1 in both periods of her life.
- With prob. $1-e$ the agent becomes type U (Unsuccessful) and earns 0 in both periods of her life.

Timing

Political decision: set, each period, transfer $b_t \in [0,1]$ to unsuccessful, financed by lump-sum tax τ_t under budget balance.

(1) Either: agents vote at the end-of-period for next period's benefits;
(2) Or: agents vote at the beginning of period but only the old are entitled to vote (extension: young vote but lower turnout).

Young make private investment (e_t),

Realization of uncertainty.

The utilities (net income) of agents alive at t are:

$$V_t^{os} = 1 - \tau_t, \; V_t^{ou} = b_t - \tau_t,$$
$$V_t^{y} = e_t (1 + \beta) + (1 - e_t)(b_t + \beta b_{t+1}) - e_t^2 - \tau_t - \beta \tau_{t+1}.$$

Optimal choice of investment gives

$$e_t = 1 - \frac{1 - \beta + (b_t + \beta b_{t+1})}{2}$$

Let u_t denote the proportion of old U at t. Then,

$$u_{t+1} = (1 - e_t),$$

since all young are identical. Budget balance imposed each period requires

$$2\tau_t = (u_{t+1} + u_t) b_t$$

which implies

$$\tau_t = \frac{(1 - \beta) + (b_t + \beta b_{t+1}) + 2u_t}{4} b_t.$$

Politically decisive (median) voter chooses b to maximize her indirect utility. For simplicity, let us assume that only those who know their type, i.e. the old, participate in the political process. (Otherwise, the young would always form a majority). An alternative, pursued in the paper, is to have both the old and the young participate in the political process, but give greater weight to the old (why might this be?).

Agents are rational and forward-looking. In particular:

- The old at t care about b_{t+1} since this affects the incentives of the young to invest (and the taxbase of current redistribution).
- The old realize that their political choice affects future distribution of types, and, hence, b_{t+1} and their utility.

$$V^{os}(b_t, b_{t+1}, u_t) = 1 - \tau_t = 1 - \frac{(1 - \beta) + (b_t + \beta b_{t+1}) + 2u_t}{4} b_t,$$
$$V^{ou}(b_t, b_{t+1}, u_t) = b_t - \tau_t = b_t - \frac{(1 - \beta) + (b_t + \beta b_{t+1}) + 2u_t}{4} b_t.$$

5.2.2. Definition of equilibrium. *Markov Perfect Equilibria*: take strategies conditional only on the current state of the world. Then, a fixed point in the mapping from expectations about future redistribution.

A (Markov perfect) political equilibrium is defined as a pair of functions $\langle B, U \rangle$, where $B : [0,1] \rightarrow [0,1]$ is a public policy rule, $b_t = B(u_t)$, and $U : [0,1] \rightarrow [0,1]$ is a private decision rule, $u_{t+1} = 1 - e_t = U(b_t)$, such that the following functional equations hold:

- $B(u_t) = \arg\max_{b_t} V(b_t, b_{t+1}, u_t)$ subject to $b_{t+1} = B(U(b_t))$, and $b_t \in [0,1]$, and $V(b_t, b_{t+1}, u_t)$ is defined as the indirect utility of the current decisive voter.
- $U(b_t) = (1 - \beta + b_t + \beta b_{t+1})/2$, with $b_{t+1} = B(U(b_t))$.

5.2.3. Dictatorship. Let us refer to a situation in which only one type of old agents have political power as a "dictatorship". Therefore, there can be two types of dictatorships:

- Dictatorship of proletariat (DP), where the unsuccessful agents have all the political power.
- Plutocracy (PL), where the successful agents have old optical power.

The characterization of equilibrium under PL is straightforward, since it will never involve any redistribution. Consequently, we have a unique equilibrium under PL where:

$$u_t = u^{pl} = (1 - \beta)/2.$$

The equilibrium under DP is more complicated. Now, a representative unsuccessful old agent will choose b_t to maximize:

$$V^{ou}(b_t, b_{t+1}, u_t) = b_t - \frac{(1-\beta) + (b_t + \beta B(U(b_t))) + 2u_t}{4} b_t$$

Equilibrium first-order condition is:

$$2 - \left(u_t + \left(\frac{1-\beta}{2} \right) + b_t + \frac{\beta}{2} B(U(b_t)) + \frac{\beta}{2} b_t B'\left(U'(b_t)\right) \right) = 0$$

The last term from the non-myopic political behavior. This is the equivalent of the dynamic linkage terms we so in the analysis of MPE in the previous lecture.

To characterize the equilibrium, let us guess the form of the solution and then verify. Guess the form of the value function as

$$B(u_t) = a_1 + a_2 u_t,$$

which implies

$$B' = a_2.$$

Therefore, we have

$$U(b_t) = (1 - \beta + b_t + \beta B(U(b_t)))/2 \rightarrow$$

$$U(b_t) = \frac{1 - \beta(1 - a_1) + b_t}{2 - \beta a_2}, \quad U'(b_t) = \frac{1}{2 - \beta a_2}$$

Substituting B and U into the first-order condition and solving for b_t, we obtain

$$b_t = \left(\frac{3}{2} - \beta a_2 + \frac{1}{2}\beta\left(1 - a_1\right)\right) - \left(1 - \frac{1}{2}\beta a_2\right) u_t$$

Now verifying that this is a solution involves making sure that the following equality holds:

$$b_t = \left(\frac{3}{2} - \beta a_2 + \frac{1}{2}\beta\left(1 - a_1\right)\right) - \left(1 - \frac{1}{2}\beta a_2\right) u_t = B\left(u_t\right) = a_1 + a_2 u_t.$$

This will be the case as long as

$$a_1 = \frac{3\left(2+\beta\right) - \beta^2}{4 - \beta^2} \text{ and } a_2 = -\frac{2}{2-\beta}.$$

Thus the solution is:

$$\begin{aligned} B\left(u_t\right) &= \max\left[\frac{3\left(2+\beta\right)-\beta^2}{4-\beta^2} - \frac{2}{2-\beta}u_t, 1\right] \\ U\left(b_t\right) &= \frac{\beta\left(1+\beta\right)+2}{2\left(2+\beta\right)} + \frac{2-\beta}{4}b_t \end{aligned}$$

(though we have to make sure that the constraint $b \in [0,1]$ and additional boundary conditions hold):

5.2.4. Majority voting. Let us now analyze the equilibrium under majority voting. Majority voting implies the following pattern:

If $u_t \leq 1/2$, the successful agents decide. If $u_t > 1/2$, the unsuccessful agents decide.

Therefore, we have the following pattern of results:

(1) If $u_0 \leq 1/2$, no welfare state ever arises.
(2) If $u_0 > 1/2$, two possible equilibria depending on expectations;
   (a) Perpetual survival of the welfare state ("pro-welfare" expectations), and
   (b) The welfare state is (strategically) terminated in, at most, two periods ("anti-welfare" expectations).

The equilibrium (a) can be sustained for any parameter, but (b) only sustained if agents are sufficiently forward-looking (i.e., patient).

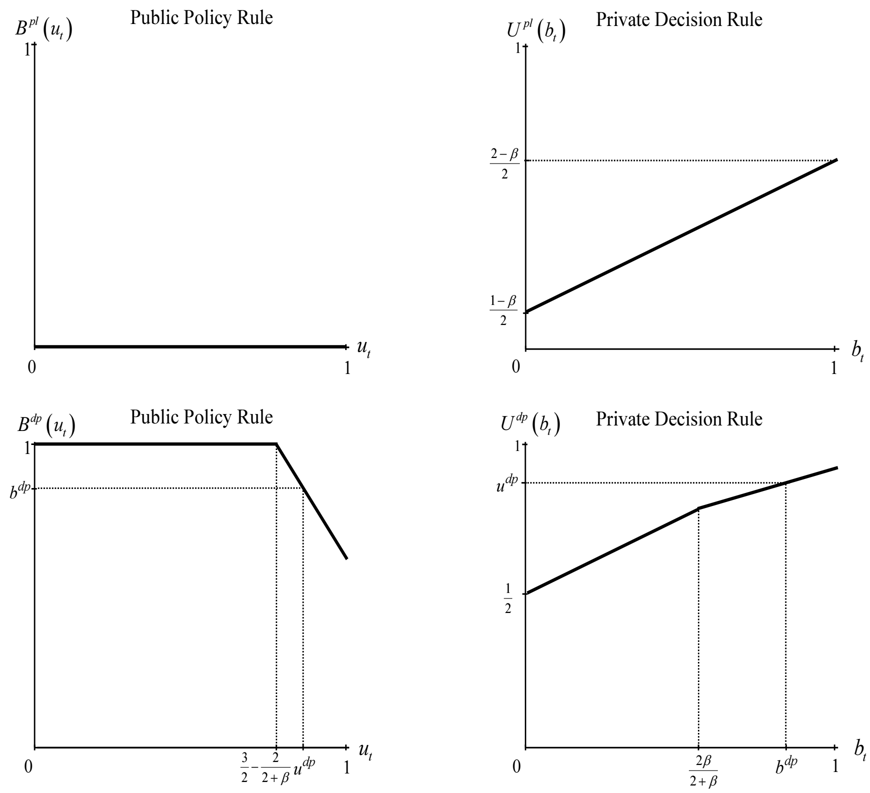


Figure 1. Public policy rule and private decision rule under Plutocracy (upper panels) and Dictatorship of Proletariat (lower panels)

5.2.5. Equilibrium survival of the welfare state. Intuition: an existing welfare state can regenerate its own political support, at least as long as the young have faith in its survival (i.e., as long as they have the right expectations).

In particular: the existence of the welfare state implies low investment by young agents and a large future constituency for the welfare state.

No welfare state, in turn, implies high investment and a small future constituency for the welfare state.

How does the welfare state start? Initially, a big shock, (depression, democratization, rise of labor movement) could start the welfare state, then it regenerates its support.

5.2.6. Equilibrium with the break-down of the welfare state. The old unsuccessful agents want to be the last generation living in a welfare state, since their tax burden depends positively on b_{t+1}.

Thus, if the young believe the welfare state to be fragile, the old can induce its breakdown by voting for sufficiently low b_t.

The young work (invest) hard and $u_{t+1} \leq 0.5$

The termination of the welfare state in finite time can be an equilibrium if $\beta > \beta_M \approx 0.555$.

Intuitively, to induce young to rationally believe that the welfare state is about to breakdown, the old unsuccessful must set b sufficiently low.

How low is low depends on the young's expectations.

Expectations must be consistent with

$$U(b_t) = (1 - \beta + b_t + \beta B(U(b_t)))/2.$$

In other words, the condition that

$$B(U(b_t)) = 0$$

requires $U(b_t) \leq 0.5$, and this can be sustained as rational expectations as long as $b_t \leq \beta$ ("anti-welfare" expections).

But the belief that $B(U(b_t)) > 0$ is also self-fulfilling under any b_t except $b_t = 0$.

Break-down requires forward looking behavior, cannot happen if β is low.

5.2.7. Wage inequality and political equilibrium. Parameterize inequality by assuming that the successful agents earn $w \neq 1$.

$$u_{t+1} = 1 - e_t^* = \frac{2 - (1+\beta)w + b_t w + \beta b_{t+1} w}{2}.$$

If $w < 1/(1+\beta)$, the welfare state is the unique outcome ($u_{t+1} > 1/2$, for any non-negative sequence of b's).

Intermediate w's: multiple equilibria (as before).

Large w's, on the other hand, implied that there is no welfare state equilibrium.

Consequently, a shock, such as skill-biased technological change or globalization, that increases equilibrium wage premium may undermine the put it will support for the welfare state.

5.3. References

(1) Alesina, Alberto and Dani Rodrik (1994) "Distributive Politics and Economic Growth" *Quarterly Journal of Economics*, 109, 465-490.

(2) Banks, Jeffrey S. and Rangarajan K. Sundaram (1998) "Optimal Retention in Agency Models," *Journal of Economic Theory*, 82, 293-323.

(3) Hassler Jon, Sevi Mora, Kjandetil Storlesseten and Fabrizio Zilibotti (2003) "Survival of the Welfare State," *American Economic Review*, 93, 87-112.

(4) Meltzer, Allan H. and Scott Richard (1981) "A Rational Theory of the Size of Government" *Journal of Political Economy* volume 89, #5, 914-927.

(5) Persson, Torsten and Guido Tabellini (1994) "Is Inequality Harmful for Growth? Theory and Evidence," *American Economic Review*, 84, 600-621.

(6) Roberts, Kevin (1977) "Voting Over Income Tax Schedules" *Journal of Public Economics*, 8,329-40.

(7) Romer, Thomas (1975) "Individual Welfare, Majority Voting and the Properties of a Linear Income Tax" *Journal of Public Economics*, 7, 163-68.

CHAPTER 6

Dynamic Voting with Changing Constituencies

The issues of dynamic voting become more interesting and challenging when votes affect the future also by influencing the distribution of power or the "constituency" taking part in elections. The simplest case of this is when a particular club votes over its own membership. Another case is when a society decides who will have the right to vote and what voting procedure will be used in the future. Some of these issues will be discussed in greater detail when we turn to endogenous changes institutions and the meaning and analysis of constitutional rules. Nevertheless, it is useful to start now with two models that illustrate the major problems.

6.1. Dynamic Voting in Clubs

We now study a model of dynamic voting, where the votes are directly over the size of a club. This will have a natural affinity to models of franchise extension and constitutional stability we will study below. This model is due to Roberts (1999) and allows changes in the size of the voting population (the club size) at every date. It also provides us with some tools for analysis of dynamic voting problems more generally.

To simplify the analysis, throughout this section let us focus on *Markov Perfect Equilibria.*

Consider an economy consisting of a finite group $X = \{1, 2...\bar{x}\}$. Each individual is assumed to have preferences such that they always want to be part of the voting "club", though a given individual may want others to be excluded. To make the model tractable, it is assumed that there is an actual seniority system whereby if the voting population is of size x, it includes individuals $\{1, 2, ..., x\}$, i.e., lower index individuals are always included before higher index individuals. Let the set of potential franchises be denoted by $\mathcal{X}$ (these are sets of the form $\{1\}$, $\{1, 2\}$, etc.). But in fact, since $\mathcal{X}$ consist so ordered pairs, it is sufficient to represented by its highest element, thus the set X is sufficient for our purposes.

Let us denote the size of the voting population at time t by x_t and assume that the instantaneous utility of individual ξ when the size of the voting club is x is given by $u(x, \xi)$. In terms of more micro models, this instantaneous utility function incorporates what the utility of individual ξ will be when tax policies are determined by a franchise of x individuals.

Given this instantaneous utility function, the expected utility of individual ξ at time $t = 0$ is given by:

$$U_0 = \mathbb{E}_0 \sum_{t=0}^{\infty} \delta^t u(x_t, \xi), \tag{6.1}$$

where $0 < \delta < 1$ is the discount factor. The key assumption that will simplify the analysis is the following.

ASSUMPTION 6.1. ***(Increasing Differences)*** *For all $x > x'$, $\xi > \xi'$, we have*

$$u(x, \xi) - u(x', \xi) \geq u(x, \xi') - u(x', \xi').$$

This assumption implies that not only are higher ranked individuals included later in the club than lower-ranked individuals, but they also have preference towards larger groups. This would make sense, for example, when we think of larger franchises as leading to higher taxes, and higher taxes being more damaging to richer individuals.

Two points are noteworthy:

(1) Increasing Differences has an obvious parallel to the single-crossing property we introduced earler. You should verify that Increasing Differences is a form of a single-crossing assumption. We will see that this assumption will play a similar role here.
(2) Note that this assumption does not mean that individuals like or dislike larger voting franchises. It is possible, for example, for all individuals to dislike larger voting franchises. The assumption simply compares the relative preferences of individuals of different ranks.

For the analysis, is more convenient to have the following stronger version:

ASSUMPTION 6.2. ***(Strict Increasing Differences)*** *For all $x > x'$, $\xi > \xi'$, we have*

$$u(x, \xi) - u(x', \xi) > u(x, \xi') - u(x', \xi').$$

Let us consider Markov transition rules for analyzing how the size of the club changes over time. A Markov transition rule is denoted by y such that

$$y : X \to X.$$

A transition rule is useful because it defines the path of x recursively such that for all t, i.e.,

$$x_{t+1} = y(x_t).$$

If there is an x such that $x = y(x)$, then x is a *steady state* of the system.

In general, we can consider both deterministic and stochastic transition rules $y(\cdot)$. Later to establish existence of equilibrium, it will be important to allow for mixed strategies and thus stochastic rules.

Let us now represent individual utility recursively. In particular the utility of individual ξ under Markov transition rule $y(\cdot)$, starting at x, is given by $V(x, \xi, y(\cdot))$ such that

$$V(x, \xi, y(\cdot)) = u(x, \xi) + \delta V(y(x), \xi, y(\cdot)), \tag{6.2}$$

where it is important to note that $y(\cdot)$ denotes the entire function. Therefore, denoting the set of potential transition functions by $\mathcal{Y}$, the value function is:

$$V : X \times X \times \mathcal{Y} \rightarrow \mathbb{R}$$

In the case where $y(\cdot)$ is stochastic, then (6.1) is assumed to be the von Neumann-Morgenstern utility function, and (6.2) should be written as:

$$V(x, \xi, y(\cdot)) = u(x, \xi) + \delta \mathbb{E} V(y(x), \xi, y(\cdot)),$$

An individual's preferences over club size will be given by V and will be conditional on the transition rule $y(\cdot)$.

Let us now define *Markov Voting Equilibrium (MVE).*

DEFINITION 6.1. *Given any transition rule $y^*(\cdot)$, for each $x \in X$, let $Y^*(x)$ be the set of y such that for all $z \in X$:*

$$\begin{aligned} & \# \{\xi : \xi \leq x \ \& \ V(y, \xi, y^*) > V(z, \xi, y^*)\} \\ \geq \ & \# \{\xi, \xi \leq x \ \& \ V(y, \xi, y^*) < V(z, \xi, y^*)\} \end{aligned} \tag{6.3}$$

If $y^*(x) \in Y^*(x)$ for all x then $y^*(\cdot)$ is an ***MVE***.

This condition says that, given that $y^*(\cdot)$ is followed in the future, no club size defeats $y^*(x)$ in pairwise majority voting as the choice for the next period—$y^*(x)$ is the Condorcet winner for the club of size x. (If $y^*(\cdot)$ is stochastic then (6.3) must hold for all realizations of $y^*(\cdot)$ that occur with non-zero probability).

Voting for the state that maximizes V is a weakly dominant strategy for each member [as usual, there are other, degenerate Nash equilibria where, for instance, everybody votes for the *status quo* because no single member can disrupt such an outcome, but such possibilities are ruled out by assumption].

Note that as the electorate at $t+1$ will be $\{1, ...y^*(x_t)\}$, the majority winner chosen by this group will in general be different to that chosen by $\{1, ...x_t\}$ and so $y^*(y^*(x_t))$ may differ from $y^*(x_t)$.

One of the main results will concern the comparison of MVE to a median-voter rule where, for every club size, an individual with median seniority chooses the club size for the next period. A *median voter* is a function $m(\cdot)$ such that, for voting club x, $m(x)$ is $(x+1)/2$ when x is odd and either $x/2$ or $(x/2)+1$ when x is even. The group of individuals both below and above the median voter, including the median voter, constitutes a weak majority.

DEFINITION 6.2. *Given any transition rule $\tilde{y}(\cdot)$, for each $x \in X$, let $\tilde{Y}(x)$ be the set of y such that for all $z \in X$:*

$$\begin{array}{lll} x \text{ odd:} & & V(y, m(x), \tilde{y}(\cdot)) \geq V(z, m(x), \tilde{y}(\cdot)) \\ x \text{ even:} & \text{either} & V(x, m(x), \tilde{y}(\cdot)) \geq V(z, m(x), \tilde{y}(\cdot)) \text{ for each } m(x) \\ & \text{or} & V(x, m(x), \tilde{y}(\cdot)) > V(z, m(x), \tilde{y}(\cdot)) \text{ for some } m(x) \end{array} \tag{6.4}$$

*If $\tilde{y}(x) \in \tilde{Y}(x)$ for all x then $\tilde{y}(\cdot)$ is a **Median Voter Rule**.*

6.1.1. Equilibrium. Equilibrium analysis will characterize both Markov voting equilibria and median voter equilibria. Both of these have the difficulty that they involve the value functions of voters, and the value functions are endogenously determined as a function of future votes (equilibria). In the absence of this dynamic linkage, characterization of the equilibrium would be straightforward. For example, in the hypothetical case where a club of size x would vote once for a change in the size of the club and there will be no future vote, the median voter choice $\tilde{y}$ could be easily determined as:

$$\begin{array}{lll} x \text{ odd:} & & u(\tilde{y}, m(x)) \geq u(z, m(x)) \\ x \text{ even:} & \textit{either} & u(\tilde{y}, m(x)) \geq u(z, m(x)) \text{ for each } m \\ & \textit{or} & u(\tilde{y}, m(x)) > u(z, m(x)) \text{ for some } m. \end{array} \tag{6.5}$$

Such a $\tilde{y}$ exists since X is finite: trivially when x is odd, and when x is even, let $\tilde{y}$ be a best choice for $m(x) = (x/2)+1$ from among the best outcomes for $m(x) = x/2$.

Assume now that $\tilde{y} > z$ for some z (the alternative case $z > \tilde{y}$ is treated similarly). Applying the strict increasing differences condition, we have

$$u(\tilde{y}, \xi) > u(z, \xi) \tag{6.6}$$

for all $\xi > (x+1)/2$, x odd, and $\xi > x/2$, x even—both median voters cannot be indifferent when x is even. Thus

$$\#\{\xi : \xi \leq x \ \& \ u(\tilde{y}, \xi) > u(z, \xi)\} \geq \#\xi, \xi \leq x \ \& \ u(\tilde{y}, \xi) < u(z, \xi)\}, \tag{6.7}$$

and $\tilde{y}$ would indeed be a majority (median voter) solution.

The idea of the approach pursued here is to try to extend these notions to the dynamic case. In particular, we would like to extend the single crossing features to the dynamic voting situations. The following lemma is an important step in this direction:

LEMMA 6.1. *Given any $x \in X$, let $(y_0, ..., y_t, ...)$ be a sequence such that $y_t \in X$ and $y_t \geq x$ for all t, $y_t > x$ for some t. If an individual ξ weakly prefers a constant x to the stream $\{y_t\}$, then there is strict preference for all ξ', such that $\xi' < \xi$*

$$\sum_{t=0}^{\infty} \delta^t u(x, \xi) \geq \sum_{t=0}^{\infty} \delta^t u(y_t, \xi) \tag{6.8}$$
$$\implies \sum_{t=0}^{\infty} \delta^t u(x, \xi') > \sum_{t=0}^{\infty} \delta^t u(y_t, \xi')$$

The same conclusion also follows for any $\xi' > \xi$ if $y_t \in X$ and $y_t < x$ for all t, $y_t < x$ for some t.

PROOF. Consider the case where $y_t \geq x$ and $\xi' < \xi$. Using the assumption on the strict increasing differences, this implies

$$u(y_t, \xi) - u(x, \xi) \geq u(y_t, \xi') - u(x, \xi')$$

with strict inequality for some t. Weighting by δ^t and summing over all t gives the street inequality in the second line of (6.8). The proof of the converse case is analogous. □

An immediate strengthening of Lemma 6.1 is:

LEMMA 6.2. *Let $(y_0, y_1, ...)$ and $(y_0', y_1', ...)$ be two sequences such that $y_t' \geq y_t$ for all t with the inequality strict from some t. If an individual weakly prefers the stream $\{y_t\}$ to the stream $\{y_t'\}$, then there is strict preference of $\{y_t\}$ over the stream $\{y_t'\}$ for all ξ', ξ such that $\xi' < \xi$. At the same conclusion also follows for $\xi' > \xi$ and $y_t' \leq y_t$.*

We also have:

THEOREM 6.1. *Consider a dynamic path $\{y_t\}$ generated by a non-stochastic MVE, i.e., $y^* : y_{t+1} = y^*(y_t)$ for all $, t \geq 0$ or by a median voter rule $\tilde{y} : y_{t+1} = \tilde{y}(y_t)$ for all $t \geq 0$. Then the following types of cycles are not possible: (i) $y_0 > y_1 \leq y_t$, for all $t \geq 2$ and with $y_1 < y_\tau$ for some $\tau \geq 2$, or (ii) $y_0 < y_1 \geq y_t$, for all $t \geq 2$ with $y_1 > y_t$ and for some $\tau \geq 2$.*

The same conclusions hold for stochastic MVE and median voter rules, i.e., if y^* is stochastic, then a sample path with $y_0 > y_1 \leq y_t$, for all $t \geq 2$ and with $y_1 < y_\tau$ for some $\tau \geq 2$ has zero probability.

PROOF. Suppose, to obtain a contradiction, that under some MVE, y^*, there is a dynamic path with $y_0 > y_1 \leq y_t, t \geq 2$ and $y_1 < y_\tau$ for some $\tau \geq 2$ (this covers the "*and*" case; the "*or*" case is treated similarly). Now

$$V(y_1, \xi, y^*) = u(y_1, \xi) + \delta V(y^*(y_1), \xi, y^*) \tag{6.9}$$

so that

$$V(y_1,\xi,y^*) \begin{matrix}\le\\ \ge\end{matrix} V(y_2,\xi,y^*) \tag{6.10}$$
$$\implies \sum_{t=2}^{\infty}\delta^t u(y_1,\xi) \begin{matrix}\le\\ \ge\end{matrix} \sum_{t=2}^{\infty}\delta^t u(y_t,\xi)$$

Now if $y_t \ge y_1$ for all $t \ge 2$, with the inequality being strict for some t, Lemma 6.1 implies that there must exist $\xi^*, \xi^{**}, \xi^* \le \xi^{**}$ such that

$$\begin{array}{lll} \xi \le \xi^* & \iff & V(y_1,\xi,y^*) > V(y_2,\xi,y^*) \\ \xi^* < \xi \le \xi^{**} & \iff & V(y_1,\xi,y^*) = V(y_2,\xi,y^*) \\ \xi^{**} < \xi & \iff & V(y_1,\xi,y^*) < (y_2,\xi,y^*) \end{array} \tag{6.11}$$

Moreover, since the increasing difference condition is strict, it must in fact be the case that either $\xi^{**} = \xi^*$ or $\xi^{**} = \xi^* + 1$. Now using Definition 6.1, in particular, equation (6.3), yields that since y_1 is chosen at y_0, we must have:

$$\xi^* \ge y_0 - \xi^{**} \tag{6.12}$$

and y_2 is chosen at y_1 so:

$$y_1 - \xi^{**} \ge \xi^*$$

which combine to give

$$y_1 \ge \xi^* + \xi^{**} \ge y_0,$$

yielding a contradiction.

To prove the same result for median voter rules, the proof can be replicated with y^* replaced by $\tilde{y}$. Instead of (6.12), we applied the definition in equation (6.4), which implies that as y_1 is a median voter choice at y_0, it must satisfy

$$\begin{array}{ll} y_0 \text{ odd:} & m(y_0) = (y_0+1)/2 \le \xi^{**} \\ y_0 \text{ even:} & m(y_0) = y_0/2 \le \xi^* \end{array} \tag{6.13}$$

which exploits the fact that both $y_0/2$ and $(y_0/2)+1$ cannot be indifferent. Also as y_2 is the median voter choice at y_1, we have

$$\begin{array}{ll} y_1 \text{ odd:} & m(y_1) = (y_1+1)/2 > \xi^* \\ y_1 \text{ odd:} & m(y_1) = (y_1+1)/2 > \xi^{**} \end{array} \tag{6.14}$$

again yielding a contradiction for all $\xi^{**} = \xi^*$ or $\xi^* + 1$, and completing the proof. □

This theorem rules out extreme turning points in the size of club membership. We use it first to investigate the occurrence of cycles. A transition rule $y(\cdot)$ generates a *cycle* if there is a dynamic path $\{y_t\}$ such that $y_{t+1} = y(y_t)$ for all $t \ge 0$, $y_s = y_t,\ s < t$ and $y_\tau \ne y_s, y_t$ for some $\tau, s < \tau < t$. A stochastic rule generates cycles if a cycle occurs with strictly positive probability.

THEOREM 6.2. *An MVE transition rule y^* generates no cycles and a median voter rule $\tilde{y}$ generates no cycles.*

This result follows from Theorem 6.1. If there is a cycle then let $\underline{x}$ be the *lowest* membership size belonging to the cycle. If $y(\cdot)$ is the transition rule then $\underline{x} = y(x')$ for some $x' > \underline{x}$ (equality is ruled out because $\underline{x}$ would then be a steady state ($\underline{x} = y(\underline{x})$) which rules out a cycle). If $y(\cdot)$ is stochastic then there is a strictly positive realization with these properties. From the definition of $\underline{x}$, the transition path (y_t) starting at $x'(y_0 = x')$ satisfies the property that $y_0 > y_t, t \geq 2$ with $y_1 < y_t$ for some t, e.g. $t = 2$. Using Theorem 6.1, cycles are not possible if the transition rule is an MVE or a median voter rule.

Theorem 6.2 shows that the transition rules under consideration do not give rise to a perpetual cycle and, as x is finite, the dynamic paths generated by y^* and $\tilde{y}$ must, in a finite time, reach a steady state $\underline{x}$, $\underline{x} = y(\underline{x})$ where $y(\cdot)$ is $y^*(\cdot)$ or $\tilde{y}(\cdot)$. An induction argument, moving backwards from a steady state, allows us to apply Theorem 6.1 to show that transitions to a steady state must involve a degree of monotonicity.

THEOREM 6.3. *Let (y_t) be a dynamic path generated by an MVE or a median voter rule. Then the path is monotonic:*

$$y_{t+1} \gtrless y_t \Longrightarrow y_\tau \begin{smallmatrix} \geq \\ \leq \end{smallmatrix} y_t \tag{6.15}$$

for all t, τ, $\tau > t$.

Theorem 6.3 does not show that all dynamic paths are monotonic in the same direction and we will see below that commonly there will be a mixture of monotonically increasing and decreasing paths. What can be shown is that paths do not cross, and this is shown by a similar induction argument to that used to prove Theorem 6.3:

THEOREM 6.4. *If $\{y_t\}$ and $\{y'_t\}$ are dynamic paths generated by an MVE or a median voter rule then $y_0 \geq y'_0 \Longrightarrow y_t \geq y'_t$ for all $t \geq 0$.*

THEOREM 6.5. *An MVE transition rule y^* is a median voter rule $\tilde{y}$ and vice versa.*

PROOF. Let $\tilde{y}(\cdot)$ be any transition rule satisfying the ordering condition of Theorem 6.4. The rules y^* and $\tilde{y}$ are two examples. It is sufficient to show that $Y^*(x) = Y(x)$ for all x, where the sets are defined by (6.3) and (6.4) under the rule $\tilde{y}$. Take $Y \in Y^*(x)$. Then, for any z:

$$\begin{aligned} & \#\{\xi : \xi \leq x \& V(y, \xi, \tilde{y}) > V(z, \xi, \tilde{y})\} \\ \geq \quad & \#\{\xi : \xi \leq x \& V(y, \xi, \tilde{y}) < V(z, \xi, \tilde{y})\} \end{aligned} \tag{6.16}$$

As $\tilde{y}$ satisfies the ordering condition, Lemma 6.2 implies that the two sets in (6.16) take the form $(1, ..., \xi^*)$ and $(\xi^{**} + 1, ..., x)$, with $\xi^{**} = \xi^*$ or $\xi^* + 1$.

Now if $y > z$, then (6.16) gives

$$x - \xi^{**} \geq \xi^*$$

If x is odd, then $m(x) > \xi^*$; if x is even then $x - \xi^{**} \geq x/2$ so $m(x) = x/2 + 1 \geq \xi^{**} + 1$.

In both cases (6.4) is implied. Repeating the exercise for $z > y$ gives $y \in Y(x)$. Now take any $y \in Y(x)$. Assume that x is odd (the case of x even is dealt with similarly): then, using (6.4), we have for all z

$$V\left(y, \frac{x+1}{2}, \bar{y}\right) \geq V\left(z, \frac{x+1}{2}, \bar{y}\right) \tag{6.17}$$

If $y > z$, then Lemma 6.2 implies (given that $\tilde{y}$ satisfies the ordering condition):

$$\begin{array}{ll} \xi \leq \xi^* & \Longrightarrow V(y, \xi, \tilde{y}) < V(z, \xi, \tilde{y}) \\ \xi^* < \xi \leq \xi^{**} & \Longrightarrow V(y, \xi, \tilde{y}) = V(z, \xi, \tilde{y}) \\ \xi > \xi^{**} & \Longrightarrow V(y, \xi, \tilde{y}) > V(z, \xi, \tilde{y}) \end{array}$$

where $\xi^{**} = \xi^*$ or $\xi^* + 1$. Equation (6.17) then implies that $\frac{x+1}{2} > \xi^*$ which gives $x - \xi^{**} \geq \xi^*$, or

$$\begin{array}{ll} & \#\{\xi : \xi \leq x \ \& \ V(y, \xi, \tilde{y}) > V(z, \xi, \tilde{y})\} \\ \geq & \#\{\xi : \xi \leq x \ \& \ V(y, \xi, \tilde{y}) < V(z, \xi, \tilde{y})\} \end{array}$$

Repeating the exercise for $z > y$ gives the same conclusion so that $y \in Y^*(x)$. We have thus shown that $Y^*(x) = Y(x)$ which implies that y^* is a median voter rule under transition rule y^* and $\tilde{y}$ is an MVE under the transition rule $\tilde{y}$, i.e. y^* is a median voter rule and $\tilde{y}$ is an MVE. □

THEOREM 6.6. *A (possibly stochastic) median voter rule $\tilde{y}(\cdot)$ exists.*

PROOF. The proof is similar to that of existence of Markov Perfect Equilibria in stochastic games we saw earlier in the lectures. It involves constructing a normal form game G with a finite number of players and (pure) strategies, showing the existence of a Nash equilibrium in mixed strategies, and then establishing that this corresponds to the Markov Perfect Equilibrium of the dynamic game. □

A similar argument also establishes:

THEOREM 6.7. *A Markov Voting Equilibrium MVE exists.*

6.1.2. Steady States. To understand the structure of MVEs, let us look at median voter rules. Given an x let $\mu(x)$ be the "*bliss point*" of this voter, i.e., the club size that would be optimal for a median voter who could commit to no further changes in the future. Then we have:

$$x \text{ odd} \quad : \quad \mu(x) = \arg\max u\left(\cdot, \frac{x+1}{2}\right) \tag{6.18}$$

$$x \text{ even} \quad : \quad \mu(x) \in [\mu^*, \mu^{**}] \tag{6.19}$$

where

$$\mu^* = \arg\max u\left(\cdot, \frac{x}{2}\right) \qquad \mu^{**} = \arg\max u\left(\cdot, \left(\frac{x}{2}\right) + 1\right).$$

Now using these definitions, we can look at the characterization of city states.

Consider an electorate of x^* with μ unique. By definition,

$$u(x^*, m(x^*)) > u(x, m(x^*)) \qquad \text{for all } x, x \neq x^* \tag{6.20}$$

Let $\tilde{y}(\cdot)$ be the median voter transition rule. Now, if $\tilde{y}(x^*) \neq x^*$ then

$$V(x^*, m(x^*), \tilde{y}) = u(x^*, m(x^*)) + \delta V(\tilde{y}(x^*), m(x^*), \tilde{y}) > V(\tilde{y}(x^*), m(x^*), \tilde{y}) \tag{6.21}$$

as from (6.20)

$$V(\tilde{y}(x^*), m(x^*), \tilde{y}) = \sum_{t=0}^{\infty} \delta^t u(y_t, m(x^*)) < \left(\sum_{t=0}^{\infty} \delta^t\right) u(x^*, m(x^*)). \tag{6.22}$$

As (6.21) violates (6.4), we have:

THEOREM 6.8. *If there is a club size x^* where the unique value of $\mu(x^*)$ is x^*, then x^* is a steady state of an MVE.*

The intuition is straightforward – if a club size is reached which the median voter views as optimal then he will not wish to vote for a change in its size and, as he is a median voter, he can always enlist a majority in ensuring no change. We call such a situation an *extrinsic steady state* because it is a steady state irrespective of the transition rule adopted away from this state. In other words, extrinsic steady states are equilibria because the current club does not wish to change the political institutions.

Interestingly, there can be other types of steady states. Consider a situation (with weak increasing differences) where the bliss points are as follows:

$$\begin{aligned} \mu(x^*) &= \mu(x^*+1) = \mu(x^*+2) = x^* \\ \mu(x^*+3) &= x^*+2 \\ \mu(x^*+k) &= x^*+5, \ k \geq 4 \end{aligned} \tag{6.23}$$

Inspection shows that x^* is an extrinsic steady state. When the club size is x^*+1 or x^*+2, the median voter will choose a club size of x^* in the knowledge that x^* will be chosen forever and this must dominate any other choice. At a club size of x^*+3, a decision to increase the size to x^*+4 will lead the then median voter to raise it to x^*+5 in the knowledge that it will remain at that level forever; on the other hand, a decision to reduce the size of the club to x^*+2 say, will then lead to it falling to x^*. Let the median voter at x^*+3 have preferences of the form

$$u(x, m(x^*+3)) = -(x-(x^*+2))^2 \tag{6.24}$$

The future discounted utility of $m(x^*+3)$ from changing the size of the club is:

$$\begin{array}{ll} \text{increase:} & -2^2 - \frac{\delta}{1-\delta^2}3 = -\left(\frac{4+5\delta}{1-\delta}\right) \\ \text{no change:} & -\frac{1}{1-\delta} \\ \text{reduce:} & -\frac{\delta}{1-\delta}2^2 = -\frac{4\delta}{1-\delta} \end{array} \tag{6.25}$$

The optimal rule is

(1) No change if $\delta \geq 1/4$
(2) Reduce to x^*+2 if $\delta < 1/4$

If discounting is not too high, the voting club of x^*+3 is a steady state even though the median voter would prefer a smaller club and this proves

Theorem 6.9. *There can be steady states of an MVE which involve voting franchises that are sub-optimal for the median voter at that club size.*

Such steady states will be referred to as *intrinsic steady states* because they are sustained as steady states by the transition rule operated away from the steady state rather than by the preferences of the median voter for that steady state. Intuitively, such steady states exist because the current club is afraid of changing its size because of future induced changes. This intuition (as well as the above example) immediately show that the role of discounting is central. If the future does not matter too much, the fear of future changes in club size will not be that important.

Theorem 6.10. *If δ is sufficiently small then there are no intrinsic steady states.*

The proof essentially follows from the observation that with δ small, the choice by $m(x)$ of a utility maximizing club size followed by any dynamic path dominates any other possibility.

As steady states are reached in finite time, low discounting implies that the value of a dynamic path is dominated by the value of the steady state that will be reached. We thus have:

Theorem 6.11. *If S is the set of steady states for all values of the discount factor close to unity, then*

$$(6.26)\qquad \begin{array}{lll} (i) & x \in S & \Longrightarrow u(x,m(x)) \geq u(z,m(x))\,\forall\, z \in S \\ (ii) & x \in X/S & \Longrightarrow u(x,m(x)) \leq u(z,m(x)) \text{ for some } z \in S. \end{array}$$

Proof. To show (i), assume that $u(x,m(x)) < u(z,m(x))$ for some $x \in S$. Then at club size of x, the median voter would gain from changing the size of the club to z rather than remaining at x. This is true whatever the discount factor. To show (ii), note that if $u(z,m(x)) < u(x,m(x))$ for all $z \in S$ then with δ close to 1, all dynamic paths will be inferior to remaining at x. □

6.1.3. Further Analysis of Dynamics. To make more progress with the analysis of dynamics, let us introduce an additional assumption:

Assumption 6.3. ***(Strict Quasi-Concavity, SQC)*** *Each individual has a strictly quasi-concave utility function $u(\cdot,\xi)$.*

Notice that so far we have assumed "single crossing," which is weaker than the utility function of each individual being strictly quasi-concave. In fact, recall that strict quasi-concavity is equivalent to single-peakedness. Single peaked preferences generally play the same role as preferences that satisfy single crossing. But in this dynamic context, they have additional bite. Notice also that the assumption is on the instantaneous utility function, u, even with this assumption, V may not be single-peaked in its first argument.

Consider, first, an extrinsic steady state x^* and consider some x' above x^*. Assume that for all x, $x^* < x \leq x'$:

$$(6.27)\qquad \mu(x) < x$$

(recall that $\mu(x^*) = x^*$). Given (6.27), the median voter at (x^*+1) most prefers a club size of x^* and this choice, as it involves x^* then being chosen forever must dominate any other dynamic path. At (x^*+2), the median voter most prefers a club size of x^* or x^*+1. Given SQC, any dynamic path starting above (x^*+2) is dominated by choosing (x^*+2) until the path drops below (x^*+2) and then replicating it. The same applies to paths starting below x^*. Thus, the median voter at (x^*+2) always chooses a club size between x^* and (x^*+2). the same argument applies inductively for all x up to x' and a similar argument then applies below x^*.

Theorem 6.12. *Assume SQC and let x^* be an extrinsic steady state. If $\mu(x) < x$ for all x, $x^* < x \leq x'$ then $y^*(x') \in [x^*, x']$. Similarly, if $\mu(x) > x$ for all x, $x'' \leq x < x^*$ then $y^*(x'') \in [x'', x^*]$. Under these conditions, the steady state is an attractor.*

Theorem 6.13. *Assume SQC and let x^* be an extrinsic steady state. If $\mu(x) > x$ for all x, $x^* < x \le x'$ then $y^*(x') \in [x', \bar{x}]$. Similarly, if $\mu(x) < x$ for all x, $x'' \le x < x^*$ then $y^*(x') \in [0, x'']$. Under these conditions, the steady state is a repeller.*

6.1.4. Example. Some of these ideas can be seen more clearly in the context of an example, which exhibits "gradualism" meaning a step-by-stepped change in the voting club. Consider the following simple situation. Let utilities be of the form

$$u(x', m(x)) = (x' - (x^* + \nu(x - x^*)))^2 \tag{6.28}$$

so that

$$\mu(x) = x^* + \nu(x - x^*) \tag{6.29}$$

The preferences in (6.28) may be viewed as an approximation to more general preferences around the extrinsic steady state x^*. It is assumed that $\nu < 1$ so that x^* is an attractor.

To see when gradualism is likely, assume that $\delta \to 0$. A median voter will always wish to choose the club size given according to instantaneous utility and the median voter at $x^* + k$ will most prefer $x^* + k - 1$ if

$$1 - \frac{1}{2k} > \nu > 1 - \frac{3}{2k}. \tag{6.30}$$

6.1.5. Further Results on Dynamics. We conclude by noting two more results about the dynamics of the voting:

Theorem 6.14. *Assume SQC and let x^* be an intrinsic steady state with $\mu(x^*) > x^*$ (the case $\mu(x^*) < x^*$ is symmetrically treated). If the initial club size is above x^* then it never falls to x^*. If the initial club size x is below x^* but $\mu(x^*) \ge x^*$ then members will vote for a club size of x^*.*

Theorem 6.15. *Assume SQC. Then, a Markov Voting Equilibrium exists in pure strategies.*

6.2. Voting over Coalitions

A similar set of issues arises when we look at a voting over coalitions. While the focus here is on voting, these issues are best illustrated by looking at voting in a society under "weak institutions," where institutions do not place to many checks on behavior. The model is based on Acemoglu, Egorov and Sonin (2007).

6.2.1. The Political Game. Let $\mathcal{I}$ denote the collection of all individuals, which is assumed to be finite. The non-empty subsets of $\mathcal{I}$ are *coalitions* and the set of coalitions is denoted by $\mathcal{C}$. In addition, for any $X \subset \mathcal{I}$, $\mathcal{C}_X$ denotes the set of coalitions that are subsets of X and $|X|$ is the number of members in X. In each period there is a designated *ruling coalition*, which can change over time. The game starts with ruling coalition N, and eventually the *ultimate ruling coalition* (URC) forms. We assume that if the URC is X, then player i obtains *baseline* utility $w_i(X) \in \mathbb{R}$. We denote $w(\cdot) \equiv \{w_i(\cdot)\}_{i \in \mathcal{I}}$.

Our focus is on how differences in the powers of individuals map into political decisions. We define a *power* mapping to summarize the powers of different individuals in $\mathcal{I}$:

$$\gamma : \mathcal{I} \to \mathbb{R}_{++},$$

where $\mathbb{R}_{++} = \mathbb{R}_+ \setminus \{0\}$. We refer to $\gamma_i \equiv \gamma(i)$ as the political *power* of individual $i \in \mathcal{I}$. In addition, we denote the set of all possible power mappings by $\mathcal{R}$ and a power mapping γ restricted to some coalition $N \subset \mathcal{I}$ by $\gamma|_N$ (or by γ when the reference to N is clear). The power of a coalition X is $\gamma_X \equiv \sum_{i \in X} \gamma_i$.

Coalition $Y \subset X$ is *winning* within coalition X if and only if $\gamma_Y > \alpha\gamma_X$, where $\alpha \in [1/2, 1)$ is a fixed parameter referring to the degree of (weighted) supermajority. Naturally, $\alpha = 1/2$ corresponds to majority rule. Moreover, since $\mathcal{I}$ is finite, there exists a large enough α (still less than 1) that corresponds to unanimity rule. We denote the set of coalitions that are winning within X by $\mathcal{W}_X$. Since $\alpha \geq 1/2$, if $Y, Z \in \mathcal{W}_X$, then $Y \cap Z \neq \varnothing$.

The assumption that payoffs are given by the mapping $w(\cdot)$ implies that a coalition cannot commit to a redistribution of resources or payoffs among its members (for example, a coalition consisting of two individuals with powers 1 and 10 cannot commit to share the resource equally if it becomes the URC). We assume that the *baseline* payoff functions, $w_i(X) : \mathcal{I} \times \mathcal{C} \to \mathbb{R}$ for any $i \in N$, satisfy the following properties.

Assumption 6.4. *Let $i \in \mathcal{I}$ and $X, Y \in \mathcal{C}$. Then:*

(1) *If $i \in X$ and $i \notin Y$, then $w_i(X) > w_i(Y)$ [i.e., each player prefers to be part of the URC].*

(2) *For $i \in X$ and $i \in Y$, $w_i(X) > w_i(Y) \iff \gamma_i/\gamma_X > \gamma_i/\gamma_Y$ ($\iff \gamma_X < \gamma_Y$) [i.e., for any two URCs that he is part of, each player prefers the one where his relative power is greater].*

(3) *If $i \notin X$ and $i \notin Y$, then $w_i(X) = w_i(Y) \equiv w_i^-$ [i.e., a player is indifferent between URCs he is not part of].*

This assumption is natural and captures the idea that each player's payoff depends positively on his relative strength in the URC. A specific example of function $w(\cdot)$ that satisfies these requirements is sharing of a pie between members of the ultimate ruling coalition proportional to their power:

$$w_i(X) = \frac{\gamma_{X\cap\{i\}}}{\gamma_X} = \begin{cases} \gamma_i/\gamma_X & \text{if } i \in X \\ 0 & \text{if } i \notin X \end{cases}. \tag{6.31}$$

The reader may want to assume (6.31) throughout the text for interpretation purposes, though this specific functional form is not used in any of our results or proofs.

We next define the extensive-form complete information game $\Gamma = (N, \gamma|_N, w(\cdot), \alpha)$, where $N \in \mathcal{C}$ is the initial coalition, γ is the power mapping, $w(\cdot)$ is a payoff mapping that satisfies Assumption 10.1, and $\alpha \in [1/2, 1)$ is the degree of supermajority; denote the collection of such games by $\mathcal{G}$. Also, let $\varepsilon > 0$ be sufficiently small such that for any $i \in N$ and any $X, Y \in \mathcal{C}$, we have

$$w_i(X) > w_i(Y) \implies w_i(X) > w_i(Y) + 2\varepsilon \tag{6.32}$$

(this holds for sufficiently small $\varepsilon > 0$ since $\mathcal{I}$ is a finite set). This immediately implies that for any $X \in \mathcal{C}$ with $i \in X$, we have

$$w_i(X) - w_i^- > \varepsilon. \tag{6.33}$$

The extensive form of the game $\Gamma = (N, \gamma|_N, w(\cdot), \alpha)$ is as follows. Each *stage* j of the game starts with some ruling coalition N_j (at the beginning of the game $N_0 = N$). Then the *stage game* proceeds with the following steps:

1. Nature randomly picks agenda setter $a_{j,q} \in N_j$ for $q = 1$.

2. [Agenda-setting step] Agenda setter $a_{j,q}$ makes proposal $P_{j,q} \in \mathcal{C}_{N_j}$, which is a subcoalition of N_j such that $a_{j,q} \in P_{j,q}$ (for simplicity, we assume that a player cannot propose to eliminate himself).

3. [Voting step] Players in $P_{j,q}$ vote sequentially over the proposal (we assume that players in $N_j \setminus P_{j,q}$ automatically vote against this proposal). More specifically, Nature randomly chooses the first voter, $v_{j,q,1}$, who then casts his vote vote $\tilde{v}(v_{j,q,1}) \in \{\tilde{y}, \tilde{n}\}$ (Yes or No), then Nature chooses the second voter $v_{j,q,2} \neq v_{j,q,1}$ etc. After all $|P_{j,q}|$ players have voted, the game proceeds to step 4 if players who supported the proposal form a winning coalition within N_j (i.e., if $\{i \in P_{j,q} : \tilde{v}(i) = \tilde{y}\} \in \mathcal{W}_{N_j}$), and otherwise it proceeds to step 5.

4. If $P_{j,q} = N_j$, then the game proceeds to step 6. Otherwise, players from $N_j \setminus P_{j,q}$ are eliminated and the game proceeds to step 1 with $N_{j+1} = P_{j,q}$ (and j increases by 1 as a new transition has taken place).

5. If $q < |N_j|$, then next agenda setter $a_{j,q+1} \in N_j$ is randomly picked by Nature among members of N_j who have not yet proposed at this stage (so $a_{j,q+1} \neq a_{j,r}$ for $1 \leq r \leq q$), and the game proceeds to step 2 (with q increased by 1). If $q = |N_j|$, the game proceeds to step 6.

6. N_j becomes the ultimate ruling coalition. Each player $i \in N$ receives total payoff

$$U_i = w_i(N_j) - \varepsilon \sum\nolimits_{1 \leq k \leq j} \mathbf{I}_{\{i \in N_k\}}, \tag{6.34}$$

where $\mathbf{I}_{\{\cdot\}}$ is the indicator function taking the value of 0 or 1.

The payoff function (6.34) captures the idea that individual's overall utility is the difference between the baseline $w_i(\cdot)$ and disutility from the number of transitions (rounds of elimination) this individual is involved in. The arbitrarily small cost ε can be interpreted as a cost of eliminating some of the players from the coalition or as an organizational cost that individuals have to pay each time a new coalition is formed. Alternatively, ε may be viewed as a means to refine out equilibria where order of moves matters for the outcome. Note that Γ is a finite game: the total number of moves, including those of Nature, does not exceed $4|N|^3$. Notice also that this game form introduces sequential voting in order to avoid issues of individuals playing weakly-dominated strategies. Our analysis below will establish that the main results hold regardless of the specific order of votes chosen by Nature.

6.2.2. Axiomatic Analysis. Before characterizing the equilibria of the dynamic game Γ, we take a brief detour and introduce four *axioms* motivated by the structure of the game Γ. Although these axioms are motivated by game Γ, they can also be viewed as natural axioms to capture the salient economic forces discussed in the introduction. The analysis in this section identifies an outcome mapping $\Phi : \mathcal{G} \rightrightarrows \mathcal{C}$ that satisfies these axioms and determines the set of (admissible) URCs corresponding to each game Γ. This analysis will be useful for two reasons. First, it will reveal certain attractive features of the game presented in the previous section. Second, we will show in the next section that equilibrium URCs of this game coincides with the outcomes picked by the mapping Φ.

More formally, consider the set of games $\Gamma = (N, \gamma|_N, w(\cdot), \alpha) \in \mathcal{G}$. Holding γ, w and α fixed, consider the correspondence $\phi : \mathcal{C} \rightrightarrows \mathcal{C}$ defined by $\phi(N) = \Phi(N, \gamma|_N, w, \alpha)$ for any $N \in \mathcal{C}$. We adopt the following axioms on ϕ (or alternatively on Φ).

Axiom 1. ***(Inclusion)*** For any $X \in \mathcal{C}$, $\phi(X) \neq \varnothing$ and if $Y \in \phi(X)$, then $Y \subset X$.

Axiom 2. ***(Power)*** For any $X \in \mathcal{C}$, $Y \in \phi(X)$ only if $Y \in \mathcal{W}_X$.

Axiom 3. ***(Self-Enforcement)*** For any $X \in \mathcal{C}$, $Y \in \phi(X)$ only if $Y \in \phi(Y)$.

Axiom 4. ***(Rationality)*** For any $X \in \mathcal{C}$, for any $Y \in \phi(X)$ and for any $Z \subset X$ such that $Z \in \mathcal{W}_X$ and $Z \in \phi(Z)$, we have that $Z \notin \phi(X) \iff \gamma_Y < \gamma_Z$.

Motivated by Axiom 3, we define the notion of a self-enforcing coalition as a coalition that "selects itself". This notion will be used repeatedly in the rest of the paper.

Definition 6.3. Coalition $X \in P(\mathcal{I})$ is *self-enforcing* if $X \in \phi(X)$.

Axiom 1, inclusion, implies that ϕ maps into subcoalitions of the coalition in question (and that it is defined, i.e., $\phi(X) \neq \varnothing$). It therefore captures the feature introduced in Γ that players that have been eliminated (sidelined) cannot rejoin the ruling coalition. Axiom 2, the power axiom, requires a ruling coalition be a winning coalition. Axiom 3, the self-enforcement axiom, captures the key interactions in our model. It requires that any coalition $Y \in \phi(X)$ should be self-enforcing according to Definition 6.3. This property corresponds to the notion that in terms of game Γ, if coalition Y is reached along the equilibrium path, then there should not be any deviations from it. Finally, Axiom 4 requires that if two coalitions $Y, Z \subset X$ are both winning and self-enforcing and all players in $Y \cap Z$ strictly prefer Y to Z, then $Z \notin \phi(X)$ (i.e., Z cannot be the selected coalition). Intuitively, all members of winning coalition Y (both those in $Y \cap Z$ by assumption and those in $Y \setminus Z$ because they prefer to be in the URC) strictly prefer Y to Z; hence, Z should not be chosen in favor of Y. This interpretation allows us to call Axiom 4 the Rationality Axiom. In terms of game Γ, this axiom captures the notion that, when he has the choice, a player will propose a coalition in which his payoff is greater.

At the first glance, Axioms 1–4 may appear relatively mild. Nevertheless, they are strong enough to pin down a unique mapping ϕ. Moreover, under the following assumption, these axioms also imply that this unique mapping ϕ is single valued.

Assumption 6.5. *The power mapping γ is* generic *in the sense that if for any $X, Y \in \mathcal{C}$, $\gamma_X = \gamma_Y$ implies $X = Y$. We also say that coalition N is generic or that numbers $\{\gamma_i\}_{i \in N}$ are generic if mapping $\gamma|_N$ is generic.*

Intuitively, this assumption rules out distributions of powers among individuals such that two different coalitions have exactly the same total power. Notice that mathematically, genericity assumption is without much loss of generality since the set of vectors $\{\gamma_i\}_{i \in \mathcal{I}} \in \mathbb{R}_{++}^{|\mathcal{I}|}$ that are not generic has Lebesgue measure 0 (in fact, it is a union of a finite number of hyperplanes in $\mathbb{R}_{++}^{|\mathcal{I}|}$).

THEOREM 6.16. *Fix a collection of players $\mathcal{I}$, a power mapping γ, a payoff function $w(\cdot)$ such that Assumption 10.1 holds, and $\alpha \in [1/2, 1)$. Then:*

1. There exists a unique mapping ϕ that satisfies Axioms 1–4. Moreover, when γ is generic (i.e. under Assumption 6.5), ϕ is single-valued.

2. This mapping ϕ may be obtained by the following inductive procedure. For any $k \in \mathbb{N}$, let $\mathcal{C}^k = \{X \in \mathcal{C} : |X| = k\}$. Clearly, $\mathcal{C} = \cup_{k \in \mathbb{N}} \mathcal{C}^k$. If $X \in \mathcal{C}^1$, then let $\phi(X) = \{X\}$. If $\phi(Z)$ has been defined for all $Z \in \mathcal{C}^n$ for all $n < k$, then define $\phi(X)$ for $X \in \mathcal{C}^k$ as

$$\phi(X) = \underset{A \in \mathcal{M}(X) \cup \{X\}}{\operatorname{argmin}} \gamma_A, \tag{6.35}$$

where

$$\mathcal{M}(X) = \{Z \in \mathcal{C}_X \setminus \{X\} : Z \in \mathcal{W}_X \text{ and } Z \in \phi(Z)\}. \tag{6.36}$$

Proceeding inductively $\phi(X)$ is defined for all $X \in \mathcal{C}$.

The intuition for the inductive procedure is as follows. For each X, (6.36) defines $\mathcal{M}(X)$ as the set of proper subcoalitions which are both winning and self-enforcing. Equation (6.35) then picks the coalitions in $\mathcal{M}(X)$ that have the least power. When there are no proper winning and self-enforcing subcoalitions, $\mathcal{M}(X)$ is empty and X becomes the URC), which is captured by (6.35). .

Theorem 6.16 establishes not only that ϕ is uniquely defined, but also that when Assumption 6.5 holds, it is single-valued. In this case, with a slight abuse of notation, we write $\phi(X) = Y$ instead of $\phi(X) = \{Y\}$.

COROLLARY 6.1. *Take any collection of players $\mathcal{I}$, power mapping γ, payoff function $w(\cdot)$, and $\alpha \in [1/2, 1)$. Let ϕ be the unique mapping satisfying Axioms 1–4. Then for any $X, Y, Z \in \mathcal{C}$, $Y, Z \in \phi(X)$ implies $\gamma_Y = \gamma_Z$. Coalition N is self-enforcing, that is, $N \in \phi(N)$, if and only if there exists no coalition $X \subset N$, $X \neq N$, that is winning within N and self-enforcing. Moreover, if N is self-enforcing, then $\phi(N) = \{N\}$.*

Corollary 6.1, which immediately follows from (6.35) and (6.36), summarizes the basic results on self-enforcing coalitions. In particular, Corollary 6.1 says that a coalition that includes a winning and self-enforcing subcoalition cannot be self-enforcing. This captures the notion that the stability of smaller coalitions undermines stability of larger ones.

As an illustration to Theorem 6.16, consider again three players A, B and C and suppose that $\alpha = 1/2$. For any $\gamma_A < \gamma_B < \gamma_C < \gamma_A + \gamma_B$, Assumption 6.5 is satisfied and it is easy to see that $\{A\}$, $\{B\}$, $\{C\}$, and $\{A, B, C\}$ are self-enforcing coalitions, whereas $\phi(\{A, B\}) = \{B\}$, $\phi(\{A, C\}) = \phi(\{B, C\}) = \{C\}$. In this case, $\phi(X)$ is a singleton for any X. On the other hand, if $\gamma_A = \gamma_B = \gamma_C$, all coalitions except $\{A, B, C\}$ would be self-enforcing, while $\phi(\{A, B, C\}) = \{\{A, B\}, \{B, C\}, \{A, C\}\}$ in this case.

6.2.3. Equilibrium Characterization. We now characterize the Subgame Perfect Equilibria (SPE) of the game Γ defined above and show that they correspond to the ruling coalitions identified by the axiomatic analysis in the previous section. The next subsection provides the main results. We then provide a sketch of the proofs..

The following two theorems characterize the *Subgame Perfect Equilibrium* (SPE) of game $\Gamma = (N, \gamma|_N, w, \alpha)$ with initial coalition N. As usual, a strategy profile σ in Γ is a SPE if σ induces continuation strategies that are best responses to each other starting in any subgame of Γ, denoted Γ_h, where h denotes the history of the game, consisting of actions in each past periods (stages and steps).

THEOREM 6.17. *Suppose that $\phi(N)$ satisfies Axioms 1-4 (cfr.* (6.35) *in Theorem 6.16). Then, for any $K \in \phi(N)$, there exists a pure strategy profile σ_K that is an SPE and leads to URC K in at most one transition. In this equilibrium player $i \in N$ receives payoff*

$$U_i = w_i(K) - \varepsilon \mathbf{I}_{\{i \in K\}} \mathbf{I}_{\{N \neq K\}}. \tag{6.37}$$

This equilibrium payoff does not depend on the random moves by Nature.

Theorem 6.17 establishes that there exists a pure strategy equilibrium leading to any coalition that is in the set $\phi(N)$ defined in the axiomatic analysis of Theorem 6.16. This is intuitive in view of the analysis in the previous section: when each player anticipates members of a self-enforcing ruling coalition to play a strategy profile such that they will turn down any offers other than K and they will accept K, it is in the interest of all the players in K to play such a strategy for any history. This follows immediately because by the definition of the set $\phi(N)$, because for any deviation to be profitable, the URC that emerges after such deviation must be either not self-enforcing or not winning. But the the first option is ruled out by induction while a deviation to a non-winning URC will be blocked by sufficiently many players. The payoff in (6.37) is also intuitive. Each player receives his baseline payoff $w_i(K)$ resulting from URC K and then incurs the cost ε if he is part of K and if the initial coalition N is not equal to K (because in this latter case, there will be one transition). Notice that Theorem 6.17 is stated without Assumption 6.5 and does not establish uniqueness. The next theorem strengthens these results under Assumption 6.5.

Theorem 6.18. *Suppose Assumption 6.5 holds and suppose $\phi(N) = K$. Then any (pure or mixed strategy) SPE results in K as the URC. The payoff player $i \in N$ receives in this equilibrium is given by* (6.37)*.*

Since Assumption 6.5 holds, the mapping ϕ is single-valued (with $\phi(N) = K$). Theorem 6.18 then shows that even though the SPE may not be unique in this case, any SPE will lead to K as the URC. This is intuitive in view of our discussion above. Since any SPE is obtained by backward induction, multiplicity of equilibria results only when some player is indifferent between multiple actions at a certain nod. However, as we show, this may only happen when a player has no effect on equilibrium play and his choice between different actions has no effect on URC (to put it simple, since ϕ is single-valued in this case, he cannot be indifferent between different actions that lead to different URCs).

6.2.4. Sketch of the Proofs. We now provide an outline of the argument leading to the proofs of the main results presented in the previous subsection and we present two key lemmas that are central for these theorems.

Consider the game Γ and let ϕ be as defined in (6.35). Take any coalition $K \in \phi(N)$. We will outline the construction of the pure strategy profile σ_K which will be a SPE and lead to K as the URC.

Let us first rank all coalitions so as to "break ties" (which are possible, since we have not yet imposed Assumption 6.5). In particular, $n : \mathcal{C} \longleftrightarrow \{1, \ldots, 2^{|\mathcal{I}|} - 1\}$ be a one-to-one mapping such that for any $X, Y \in \mathcal{C}$, $\gamma_X > \gamma_Y \Rightarrow n(X) > n(Y)$, and if for some $X \neq K$ we have $\gamma_X = \gamma_K$, then $n(X) > n(K)$ (how the ties among other coalitions are broken is not important). With this mapping, we have thus ranked (enumerated) all coalitions such that stronger coalitions are given higher numbers, and coalition K receives the smallest number among all coalitions with the same power. Now define the mapping $\chi : \mathcal{C} \to \mathcal{C}$ as

$$\chi(X) = \underset{Y \in \phi(X)}{\operatorname{argmin}}\, n(Y). \tag{6.38}$$

Intuitively, this mapping picks an element of $\phi(X)$ for any X and satisfies $\chi(N) = K$. Also, note that χ is a *projection* in the sense that $\chi(\chi(X)) = \chi(X)$. This follows immediately since Axiom 3 implies $\chi(X) \in \phi(\chi(X))$ and Corollary 6.1 implies that $\phi(\chi(X))$ is a singleton.

The key to constructing a SPE is to consider off-equilibrium path behavior. To do this, consider a subgame in which we have reached a coalition X (i.e., j transitions have occurred and $N_j = X$) and let us try to determine what the URC would be if proposal Y is accepted starting in this subgame. If $Y = X$, then the game will end, and thus X will be the URC. If, on the other hand, $Y \neq X$, then the URC must be some subset of Y. Let us define the strategy profile σ_K such that the URC will be $\chi(Y)$. We denote this (potentially off-equilibrium path) URC following the acceptance of proposal Y by $\psi_X(Y)$, so that

$$\psi_X(Y) = \begin{cases} \chi(Y) & \text{if } Y \neq X; \\ X & \text{otherwise.} \end{cases} \tag{6.39}$$

By Axiom 1 and equations (6.38) and (6.39), we have that

$$X = Y \iff \psi_X(Y) = X. \tag{6.40}$$

We will introduce one final concept before defining profile σ_K. Let $F_X(i)$ denote the "favorite" coalition of player i if the current ruling coalition is X. Naturally, this will be the weakest coalition among coalitions that are winning within X, that are self-enforcing and that include player i. If there are several such coalitions, the definition of $F_X(i)$ picks the one with the smallest n, and if there are none, it picks X itself. Therefore,

$$F_X(i) = \operatorname*{argmin}_{Y \in \{Z: Z \subset X, Z \in \mathcal{W}_X, \chi(Z)=Z, Z \ni i\} \cup \{X\}} n(Y). \tag{6.41}$$

Similarly, we define the "favorite" coalition of players $Y \subset X$ starting with X at the current stage. This is again the weakest coalition among those favored by members of Y, thus

$$F_X(Y) = \begin{cases} \operatorname*{argmin}_{\{Z: \exists i \in Y: F_X(i)=Z\}} n(Z) & \text{if } Y \neq \varnothing; \\ X & \text{otherwise.} \end{cases} \tag{6.42}$$

Equation (6.42) immediately implies that

$$\text{For all } X \in \mathcal{C}: F_X(\varnothing) = X \text{ and } F_X(X) = \chi(X). \tag{6.43}$$

The first part is true by definition. The second part follows, since for all $i \in \chi(X)$, $\chi(X)$ is feasible in the minimization (6.41), and it has the lowest number n among all winning self-enforcing coalitions by (6.38) and (6.35) (otherwise there would exist a self-enforcing coalition Z that is winning within X and satisfies $\gamma_Z < \gamma_{\chi(X)}$, which would imply that ϕ violates Axiom 4). Therefore, it is the favorite coalition of all $i \in \chi(X)$ and thus $F_X(X) = \chi(X)$.

Now we are ready to define profile σ_K. Take any history h and denote the player who is supposed to move after this history $a = a(h)$ if after h, we are at an agenda-setting step, and $v = v(h)$ if we are at a voting step deciding on some proposal P (and in this case, let a be the agenda-setter who made proposal P). Also denote the set of potential agenda setters at this stage of the game by A. Finally, recall that $\tilde{n}$ denotes a vote of "No" and $\tilde{y}$ is a vote of "Yes". Then σ_K is the following simple strategy profile where each agenda setter proposes his favorite coalition in the continuation game (given current coalition X) and each voter votes "No" against proposal P, if the URC following P excludes him or he expects another proposal that he will prefer to come shortly.

(6.44)
$$\sigma_K = \begin{cases} \text{agenda-setter } a \text{ proposes } P = F_X(a); \\ \text{voter } v \text{ votes} \begin{cases} \tilde{n} & \text{if either } v \notin \psi_X(P) \text{ or} \\ & v \in F_X(A), P \neq F_X(A \cup \{a\}), \text{ and } \gamma_{F_X(A)} \leq \gamma_{\psi_X(P)}; \\ \tilde{y} & \text{otherwise.} \end{cases} \end{cases}$$

In particular, notice that $v \in F_X(A)$ and $P \neq F_X(A \cup \{a\})$ imply that voter v is part of a *different* coalition proposal that will be made by some future agenda setter at this stage of the game if the current voting fails, and $\gamma_{F_X(A)} \leq \gamma_{\psi_X(P)}$ implies that this voter will receive weakly higher payoff under this alternative proposal. This expression makes it clear that σ_K is similar to a "truth-telling" strategy; each individual makes proposals according to his preferences (constrained by what is feasible) and votes truthfully.

With the strategy profile σ_K defined, we can state the main lemma, which will essentially constitute the bulk of the proof of Theorem 6.17. For this lemma, also denote the set of voters who already voted "Yes" at history h by V^+, the set of voters who already voted "No" by V^-. Then, $V = P \setminus (V^+ \cup V^- \cup \{v\})$ denotes the set of voters who will vote after player v.

Lemma 6.3. *Consider the subgame Γ_h of game Γ after history h in which there were exactly j_h transitions and let the current coalition be X. Suppose that strategy profile σ_K defined in (6.44) is played in Γ_h. Then:*

(a) If h is at agenda-setting step, the URC is $R = F_X(A \cup \{a\})$; if h is at voting step and $V^+ \cup \{i \in \{v\} \cup V : i \text{ votes } \tilde{y} \text{ in } \sigma_K\} \in \mathcal{W}_X$, then the URC is $R = \psi_X(P)$; and otherwise $R = F_X(A)$.

(b) If h is at the voting step and proposal P will be accepted, player $i \in X$ receives payoff

(6.45)
$$U_i = w_i(R) - \varepsilon\left(j_h + \mathbf{I}_{\{P \neq X\}}\left(\mathbf{I}_{\{i \in P\}} + \mathbf{I}_{\{R \neq P\}}\mathbf{I}_{\{i \in R\}}\right)\right).$$

Otherwise (if proposal P will be rejected or if h is at agenda-setting step), then player $i \in X$ receives payoff

(6.46)
$$U_i = w_i(R) - \varepsilon\left(j_h + \mathbf{I}_{\{R \neq X\}}\mathbf{I}_{\{i \in R\}}\right).$$

The intuition for the results in this lemma is straightforward in view of the construction of the strategy profile σ_K. In particular, part (a) defines what the URC will be. This follows immediately from σ_K. For example, if we are at a agenda-setting step, then the URC will be the favorite coalition of the set of remaining agenda setters, given by $A \cup \{a\}$. This is an immediate implication of the fact that according to the strategy profile σ_K, each player will propose his favorite coalition and voters will vote $\tilde{n}$ ("No") against current proposals if the strategy profile σ_K will induce a more preferred outcome for them in the remainder of this stage. Part (b) simply defines the payoff to each player as the difference between the baseline payoff, $w_i(R)$, as a function of the URC R defined in part (a), and the costs associated with transitions.

Given Lemma 6.3, Theorem 6.17 then follows if strategy profile σ_K is a SPE (because in this case URC will be K and it will be reached with at most one transition). With σ_K defined in (6.44), it is clear that no player can profitably deviate in any subgame.

The next lemma strengthens Lemma 6.3 for the case in which Assumption 6.5 holds by establishing that any SPE will lead to the same URC and payoffs as those in Lemma 6.3.

Lemma 6.4. *Suppose Assumption 6.5 holds and $\phi(N) = \{K\}$. Let σ_K be defined in (6.44). Then for any SPE σ (in pure or mixed strategies) and for any history h, the equilibrium plays induced by σ and by σ_K in the subgame Γ_h will lead to the same URC and to identical payoffs for each player.*

Since $\phi(N) = \{K\}$, Theorem 6.18 follows as an immediate corollary of this lemma (with $h = \varnothing$).

6.2.5. The Structure of Ruling Coalitions. Let us now use the theorems derived above to understand the structure of URCs. Given the equivalence result (Theorems 6.17 and 6.18), we will make use of the axiomatic characterization in Theorem 6.16. Throughout, unless stated otherwise, we fix a game $\Gamma = (N, \gamma, w(\cdot), \alpha)$ with w satisfying Assumption 10.1 and $\alpha \in [1/2, 1)$. In addition, to simplify the analysis in this section, we assume throughout that Assumption 6.5 holds and also we impose the additional assumption:

Assumption 6.6. For no $X, Y \in \mathcal{C}$ such that $X \subset Y$ the equality $\gamma_Y = \alpha\gamma_X$ is satisfied.

Assumption 6.6 guarantees that a small perturbation of a non-winning coalition Y does not make it winning. Similar to Assumption 6.5, this assumption fails only in a set of Lebesgue measure 0 (in fact, it coincides with Assumption 6.5 when $\alpha = 1/2$). All of these results are provided without proof for brevity.

We start with the result that the set of self-enforcing coalitions is open (in the standard topology); this is not only interesting per se but facilitates further proofs. Note that for game $\Gamma = (N, \gamma, w(\cdot), \alpha)$, a power mapping γ (or more explicitly $\gamma|_N$) is given by a $|N|$-dimensional vector $\{\gamma_i\}_{i\in N} \subset \mathbb{R}^{|N|}_{++}$. Denote the subset of vectors $\{\gamma_i\}_{i\in N}$ that satisfy Assumptions 6.5 and 6.6 by $\mathcal{A}(N)$, and the subset of $\mathcal{A}(N)$ for which $\Phi\left(N, \{\gamma_i\}_{i\in N}, w, \alpha\right) = N$ (i.e., the subset of power distributions for which coalition N is stable) by $\mathcal{S}(N)$ and let $\mathcal{N}(N) = \mathcal{A}(N) \setminus \mathcal{S}(N)$.

Lemma 6.5. *1. The set of power allocations that satisfy Assumptions 6.5 and 6.6,* $\mathcal{A}(N)$, *its subsets for which coalition* N *is self-enforcing,* $\mathcal{S}(N)$, *and its subsets for which coalition* N *is not self-enforcing,* $\mathcal{N}(N)$, *are open sets in* $\mathbb{R}^{|N|}_{++}$. *The set* $\mathcal{A}(N)$ *is also dense in* $\mathbb{R}^{|N|}_{++}$.

2. Each connected component of $\mathcal{A}(N)$ *lies entirely within either* $\mathcal{S}(N)$ *or* $\mathcal{N}(N)$.

An immediate corollary of Lemma 6.5 is that if the distribution of powers in two different games are "close," then these two games will have the same URC and also that the inclusion of sufficiently weak players will not change the URC. To state and prove this proposition, endow the set of mappings γ, $\mathcal{R}$, with the sup-metric, so that $(\mathcal{R}, \rho)$ is a metric space with $\rho(\gamma, \gamma') = \sup_{i \in \mathcal{I}} |\gamma_i - \gamma'_i|$. A δ-neighborhood of γ is $\{\gamma' \in \mathcal{R} : \rho(\gamma, \gamma') < \delta\}$.

Proposition 6.1. *Consider* $\Gamma = (N, \gamma, w(\cdot), \alpha)$ *with* $\alpha \in [1/2, 1)$. *Then:*

1. There exists $\delta > 0$ *such that if* $\gamma' : N \to \mathbb{R}_{++}$ *lies within* δ*-neighborhood of* γ, *then* $\Phi(N, \gamma, w, \alpha) = \Phi(N, \gamma', w, \alpha)$.

2. There exists $\delta' > 0$ *such that if* $\alpha' \in [1/2, 1)$ *satisfies* $|\alpha' - \alpha| < \delta'$, *then* $\Phi(N, \gamma, w, \alpha) = \Phi(N, \gamma, w, \alpha')$.

3. Let $N = N_1 \cup N_2$ *with* N_1 *and* N_2 *disjoint. Then, there exists* $\delta > 0$ *such that for all* N_2 *such that* $\gamma_{N_2} < \delta$, $\phi(N_1) = \phi(N_1 \cup N_2)$.

This proposition is intuitive in view of the results in Lemma 6.5. It implies that URCs have some degree of continuity and will not change as a result of small changes in power or in the rules of the game.

Although the structure of ruling coalitions is robust to small changes in the distribution of power within the society, it may be fragile to more sizeable shocks, and in fact the addition or the elimination of a single member of the self-enforcing coalition turns out to be such a sizable shock when $\alpha = 1/2$. This result is established in the next proposition.

Proposition 6.2. *Suppose* $\alpha = 1/2$ *and fix a power mapping* $\gamma : \mathcal{I} \to \mathbb{R}_{++}$. *Then:*

1. If coalitions X *and* Y *such that* $X \cap Y = \varnothing$ *are both self-enforcing, then coalition* $X \cup Y$ *is not self-enforcing.*

2. If X *is a self-enforcing coalition, then* $X \cup \{i\}$ *for* $i \notin X$ *and* $X \setminus \{i\}$ *for* $i \in X$ *are not self-enforcing.*

The most important implication is that, under majority rule $\alpha = 1/2$, the addition or the elimination of a single agent from a self-enforcing coalitions makes this coalition no longer self-enforcing. This result motivates our interpretation in the Introduction of the power struggles in Soviet Russia following random deaths of Politburo members.

Proposition 6.3. *Consider* $\Gamma = (N, \gamma, w(\cdot), \alpha)$.

1. Suppose $\alpha = 1/2$, *then for any* n *and* m *such that* $1 \leq m \leq n$, $m \neq 2$, *there exists a set of players* N, $|N| = n$, *and a generic mapping of powers* γ *such that* $|\phi(N)| = m$. *In particular, for any* $m \neq 2$ *there exists a self-enforcing ruling coalition of size* m. *However, there is no self-enforcing coalition of size* 2.

2. Suppose that $\alpha > 1/2$, *then for any* n *and* m *such that* $1 \leq m \leq n$,*there exists a set of players* N, $|N| = n$, *and a generic mapping of powers* γ *such that* $|\phi(N)| = m$.

These results show that one can say relatively little about the size and composition of URCs without specifying the power distribution within the society further (except that when $\alpha = 1/2$, coalitions of size 2 are not self-enforcing). However, this is largely due to the fact that there can be very unequal distributions of power. For the potentially more interesting case in which the distribution of power within the society is relatively equal, much more can be said about the size of ruling coalitions. In particular, the following proposition shows that, as long as larger coalitions have more power and there is majority rule ($\alpha = 1/2$), only coalitions of size $2^k - 1$ for some integer k (i.e., coalitions of size 3, 7, 15, etc.) can be the URC (Part 1). Part 2 of the proposition provides a sufficient condition for this premise (larger coalitions are more powerful) to hold. The rest of the proposition generalizes these results to societies with values of $\alpha > 1/2$.

Proposition 6.4. *Consider* $\Gamma = (N, \gamma, w(\cdot), \alpha)$ *with* $\alpha \in [1/2, 1)$.

1. Let $\alpha = 1/2$ *and suppose that for any two coalitions* $X, Y \in \mathcal{C}$ *such that* $|X| > |Y|$ *we have* $\gamma_X > \gamma_Y$ *(i.e., larger coalitions have greater power). Then* $\phi(N) = N$ *if and only if* $|N| = k_m$ *where* $k_m = 2^m - 1$, $m \in \mathbb{Z}$. *Moreover, under these conditions, any ruling coalition must have size* $k_m = 2^m - 1$ *for some* $m \in \mathbb{Z}$.

2. For the condition $\forall X, Y \in \mathcal{C} : |X| > |Y| \Rightarrow \gamma_X > \gamma_Y$ *to hold, it is sufficient that there exists some* $\lambda > 0$ *such that*

$$\sum_{j=1}^{|N|} \left| \frac{\gamma_j}{\lambda} - 1 \right| < 1. \tag{6.47}$$

3. Suppose $\alpha \in [1/2, 1)$ *and suppose that* γ *is such that for any two coalitions* $X \subset Y \subset N$ *such that* $|X| > \alpha |Y|$ ($|X| < \alpha |Y|$, *resp.) we have* $\gamma_X > \alpha \gamma_Y$ ($\gamma_X < \alpha \gamma_Y$, *resp.). Then* $\phi(N) = N$ *if and only if* $|N| = k_{m,\alpha}$ *where* $k_{1,\alpha} = 1$ *and* $k_{m,\alpha} = \lfloor k_{m-1,\alpha}/\alpha \rfloor + 1$ *for* $m > 1$, *where* $\lfloor z \rfloor$ *denotes the integer part of* z.

4. There exists $\delta > 0$ such that $\max_{i,j\in N}\{\gamma_i/\gamma_j\} < 1+\delta$ implies that $|X| > \alpha|Y|$ ($|X| < \alpha|Y|$, resp.) whenever $\gamma_X > \alpha\gamma_Y$ ($\gamma_X < \alpha\gamma_Y$, resp.). In particular, coalition $X \in \mathcal{C}$ is self-enforcing if and only if $|X| = k_{m,\alpha}$ for some m (where $k_{m,\alpha}$ is defined in Part 3).

This proposition shows that although it is impossible to make any general claims about the size of coalitions without restricting the distribution of power within the society, a tight characterization of the structure of the URC is possible when individuals are relatively similar in terms of their power.

One might expect that an increase in α—the supermajority requirement—cannot decrease the size of the URC. One might also expect that if an individual increases his power (either exogenously or endogenously), this should also increase his payoff. However, both of these are generally not true. Consider the following simple example: let $w(\cdot)$ be given by (6.31). Then coalition $(3,4,5)$ is self-enforcing when $\alpha = 1/2$, but is not self-enforcing when $4/7 < \alpha < 7/12$, because $(3,4)$ is now a self-enforcing and winning subcoalition. Next, consider game Γ with $\alpha = 1/2$ and five players A, B, C, D, E with powers $\gamma_A = \gamma'_A = 2$, $\gamma_B = \gamma'_B = 10$, $\gamma_C = \gamma'_C = 15$, $\gamma_D = \gamma'_D = 20$, $\gamma_E = 21$, and $\gamma'_E = 40$. Then $\Phi(N,\gamma,w,\alpha) = \{A, D, E\}$, while $\Phi(N,\gamma',w,\alpha) = \{B, C, D\}$, so player E, who is the most powerful player in both cases, belongs to $\Phi(N,\gamma,w,\alpha)$ but not to $\Phi(N,\gamma',w,\alpha)$.

We summarize these results in the following proposition (proof omitted).

PROPOSITION 6.5. *1. An increase in α may reduce the size of the ruling coalition. That is, there exists a society N, a power mapping γ and $\alpha, \alpha' \in [1/2, 1)$, such that $\alpha' > \alpha$ but for all $X \in \Phi(N,\gamma,w,\alpha)$ and $X' \in \Phi(N,\gamma,w,\alpha')$, $|X| > |X'|$ and $\gamma_X > \gamma_{X'}$.*

2. There exist a society N, $\alpha \in [1/2, 1)$, two mappings $\gamma, \gamma' : N \to \mathbb{R}_{++}$ satisfying $\gamma_i = \gamma'_i$ for all $i \neq j$, $\gamma_j < \gamma'_j$ such that $j \in \Phi(N,\gamma,w,\alpha)$, but $j \notin \Phi(N,\gamma',w,\alpha)$. Moreover, this result applies even when j is the most powerful player in both cases, i.e. $\gamma'_i = \gamma_i < \gamma_j < \gamma'_j$ for all $i \neq j$.

Intuitively, higher α turns certain coalitions that were otherwise non-self-enforcing into self-enforcing coalitions, but this implies that larger coalitions are now less likely to be self-enforcing and less likely to emerge as the ruling coalition. This, in turn, makes larger coalitions more stable. The first part of the proposition therefore establishes that greater power or "agreement" requirements in the form of supermajority rules do not necessarily lead to larger ruling coalitions. The second part implies that being more powerful may be a disadvantage, even for the most powerful player. This is for the intuitive reason that other players may wish to be together with less powerful players in order to receive higher payoffs.

This latter result raises the question of when the most powerful player will be part of the ruling coalition. This question is addressed in the next proposition.

PROPOSITION 6.6. *Consider the game* $\Gamma(N, \gamma, w(\cdot), \alpha)$ *with* $\alpha \in [1/2, 1)$, *and suppose that* $\gamma_1, \ldots, \gamma_{|N|}$ *is an increasing sequence. If* $\gamma_{|N|} \in \left(\alpha \sum_{j=2}^{|N|-1} \gamma_j / (1-\alpha), \alpha \sum_{j=1}^{|N|-1} \gamma_j / (1-\alpha)\right)$, *then either coalition* N *is self-enforcing or the most powerful individual,* $|N|$, *is not a part of the URC.*

6.3. References

(1) Acemoglu, Daron, Georgy Egorov and Konstantin Sonin (2007) "Coalition Formation in Nondemocracies" forthcoming *Review of Economic Studies.*

(2) Roberts, Kevin (2005) "Dynamic Voting in Clubs," http://www.nuff.ox.ac.uk/Users/Robertsk/papers.html

CHAPTER 7

Voting and Information

The models studied so far focused on perfect information situations. This in particular implies that voters know both their own preferences and payoff-relevant state variables. Different types of issues arise when we relax these assumptions. In particular, voters might recognize that others who are also taking part in the election might have relevant information for the decision. For example, other voters might also have information on which candidate will do a better job in office. This will lead to a situation similar to a "common value" auction, since the value of each candidate to each voter depends on a state variable on which other voters are also informed. In this case, voting strategies will be more complex, and as we will see, there might exist a "swing voter's curse" in that voters who have relevant information and might actually play the role of a swing voter may abstain from voting (because they want to rely on the information of others). Beyond these static implications, the fact that voters are imperfectly informed might also have dynamic implications. In particular, voter beliefs might evolve as a function of economic outcomes they observe, which are in turn influenced by policies. Beliefs that shape policies are influenced by observed outcomes, which are in turn shaped by policies. This opens the way for potential "dynamic learning failures," whereby incorrect policy choices can become self-sustaining because they generate behavior that "rationally" confirm beliefs that support these policies in the first place.

7.1. Swing Voter's Curse

Feddersen and Pesendorfer (1996) consider the following environment. There our two states of nature, $\theta = \{0, 1\}$, and two policy choices of candidates, $x \in \{0, 1\}$. There are three types of voters, denoted by elements of the type space $T = \{0, 1, i\}$. The first two are committed voters and will always choose $x = 0$ or $x = 1$ either because of distributional are ideal logical reasons. The last one designates "independent" voters, which we normally think as the "swing voters". These independents have preferences given by

$$U_i(x, \theta) = -\mathbb{I}(x \neq \theta),$$

where $\mathbb{I}(x \neq \theta)$ is the indicator function for the position of the candidate from being different than the state of nature. This implies that the voters received negative utility if the "wrong" candidate is elected. A candidate (policy) that obtains an absolute majority is chosen. If both options obtain the same number of votes, then one of them is chosen at random.

Let us suppose, without loss of any generality, that the prior probability that the true state is $\theta = 0$ is $\alpha \leq 1/2$, so that state $\theta = 1$ is more likely ex ante.

To make the model work, there needs to be some uncertainty about the preferences of other voters. One way to introduce this is to suppose that how many other voters there are (meaning how many other voters could potentially turn out to vote) and what fractions of those will be committed types are stochastically generated. Suppose, in particular, that the total number of voters is determined by Nature taking $N + 1$ independent draws from a potentially large pool of voters. At each draw, an actual voter is selected with probability $1 - p_\phi$. This implies that the number of voters is a stochastic variable with the binomial distribution with parameters $(N + 1, 1 - p_\phi)$. Conditional on being selected, an agent is independent with probability $p_i/(1 - p_\phi)$, is committed to $x = 0$ with probability $p_0/(1 - p_\phi)$, and is committed to $x = 1$ with probability $p_1/(1 - p_\phi)$. Therefore, the numbers of voters of different types also follow binomial distributions.

The probability vector (p_ϕ, p_i, p_0, p_1), like preferences and the prior probability α, is common knowledge. Finally, each agent knows her type and also receives a signal $s \in S = \{0, 1, \phi\}$, where the first two entries designate the actual state, i.e., $\theta = 0$ or $\theta = 1$, so that conditional on receiving the signal values the agent will know the underlying state for sure. The last entry means that the agent receives no relevant information and this event has probability q. This formulation implies that some voters will potentially be fully informed, but because all events are stochastic, whether there is indeed such an agent in the population or how many of them there are relative to committed types is not known by any of the voters.

The strategic interactions in this game result from the fact that voters are "rational" and thus rather than voting truthfully, they will choose a decision that will maximize the likelihood of the candidate/policy that they prefer being chosen. In particular, committed voters will vote for their favorite candidate regardless of the signals they receive (this follows immediately from the fact that because the total number of voters is uncertain, voting for anything other than your strict preference would be a strictly dominated strategy). But more interestingly, independent voters might prefer to abstain rather than vote according to their priors. Recall that their priors, $q \leq 1/2$, are informative (one could have also modeled this such that priors were uninformative, but voters received weak signals changing their posteriors).

Taking all of these into account, a pure strategy here is simply

$$\sigma : T \times S \to [\phi, 0, 1],$$

where ϕ denotes abstention. The above argument immediately states that $\sigma(0, \cdot) = 0$ and $\sigma(1, \cdot) = 1$. Moreover, it is also clear that $\sigma(i, z) = z$ for $z \in \{0, 1\}$, meaning that independent informed voters will vote according to their (certain) posterior. This implies that we can simply focus on the decisions by uninformed independent voters, denoted by

$$\tau = (\tau_0, \tau_1, \tau_\phi),$$

which correspond to the probabilities that they will vote for $x = 0$, $x = 1$ and abstain, respectively. Recall that though "uninformed," these voters have posteriors that are not equal to 1/2, thus have relevant information.

The key observation in the analysis of this model, which draws out the parallel to common-value auctions, is that an individual should only care about his or her vote conditional on being pivotal. Since they do not obtain direct utility from their votes and only care about the outcome, their votes when there is a clear majority for one or the other outcome are irrelevant. But this implies that one has to condition on a situation in which one is pivotal, which happens when either an equal number of agents have voted for each choice, or one of the two choices is winning with only one vote.

This intuition is sufficient to establish the following proposition, which captures the idea of the "swing voter's curse". For this proposition, let $U(x, \tau)$ be the expected utility of an uninformed independent agent to choose $x \in \{0, 1, \phi\}$, when all other independents are using (symmetric) mixed strategies given by τ.

PROPOSITION 7.1. *Suppose that $p_\phi > 0$, $q > 0$ and that N is greater than 2 and even. Then if $U(1, \tau) = U(0, \tau)$, then all uninformed independent voters abstain.*

If $U(1, \tau) = U(0, \tau)$, meaning that an uninformed voter is indifferent between voting for either candidate (policy), then he or she must prefer to abstain. By continuity, we could also show that if $|U(1, \tau) - U(0, \tau)| < \varepsilon$ for ε sufficiently small, then the same conclusion will apply. This is despite the fact that uninformed voters actually have relevant information, because the prior α can be arbitrarily small.

Mathematically, we have that

$$\begin{aligned} U(1,\tau) - U(\phi,\tau) &= \frac{1}{2}(1-\alpha)\left[\pi_t(1,\tau) + \pi_1(1,\tau)\right] \\ &\quad -\frac{1}{2}\alpha\left[\pi_t(0,\tau) + \pi_1(0,\tau)\right], \end{aligned}$$

where $\pi_t(\theta,\tau)$ is the probability that in state θ under strategy profile τ (for uninformed independent voters) an equal number of other agents have voted for the two options. Similarly, $\pi_x(\theta,\tau)$ is the probability that in state θ under strategy profile τ, just one less vote has been cast for x than "not x". These are the only scenarios in which the vote of the individual matters, and thus in comparison of expected utility of different actions, all other scenarios can be ignored. The first line is then obtained by noting that with probability $1-\alpha$ the true state is $\theta = 1$, and in this case, the probability that there is a tie is $\pi_t(1,\tau)$ and the probability that $x = 1$ is behind by one vote is $\pi_1(1,\tau)$. In both scenarios are voting in favor of $x = 1$, the voter in question either changes a tie into a win for $x = 1$ or changes a loss into a tie, in both cases gaining utility of $1/2$. The second line has analogous logic, except that in this case the voter is turning a tie into the wrong decision or the right decision into a tie. Also similarly, we have

$$\begin{aligned} U(0,\tau) - U(\phi,\tau) &= \frac{1}{2}\alpha\left[\pi_t(0,\tau) + \pi_0(0,\tau)\right] \\ &\quad -\frac{1}{2}(1-\alpha)\left[\pi_t(1,\tau) + \pi_0(1,\tau)\right]. \end{aligned}$$

Now if $U(1,\tau) = U(0,\tau)$, then we have that these two expressions are equal to each other, and using the binomial expressions for π's, one can simplify $U(1,\tau) - U(\phi,\tau)$ and show that it is strictly negative, meaning that it is a strict best response to abstain.

Intuitively, when a voter expects the same utility from the two options available to him or her, then abstaining and leaving the decision to another voter who is more likely to be informed is better. This is despite the fact that the voter may be leaving the decision to a committed type. The implication is that useful information will be lost in the elections, and this is the essence of the "swing voter's curse". Nevertheless, Feddersen and Pesendorfer also show that in large elections information still aggregates in the sense that the correct choice is made with arbitrarily high probability. In particular:

Proposition 7.2. *Suppose that $p_\phi > 0$, $q > 0$ and $p_i \neq |p_1 - p_0|$, then for every $\varepsilon > 0$, there exists $\bar{N}$ such that for $N > \bar{N}$, the probability that the correct candidate gets elected s greater than $1 - \varepsilon$.*

The idea of this result is that as the size of the electorate becomes large, uninformed independents mix between the "disadvantaged" candidate and abstaining, in such a way that informed independents become pivotal with very high probability.

7.2. Policies and the Evolution of Beliefs

The previous model showed how imperfect information shapes behavior in a single election. More interesting is how imperfect information and the beliefs that agents have about unknown variables and parameters shape policy, which in turn affects beliefs. This issue is discussed in Piketty (1995). The model consists of a sequence of individuals voting over a level of tax rate, which affects effort for success (e.g., upward social mobility). They then beget an offspring, who is potentially affected by the success or failure of their parent. Each individual only cares about current consumption, so there is no experimentation. In addition, Piketty assumes that individuals vote in order to maximize "social welfare" (of their own generation), but they take effort decisions that maximize their own utility.

More specifically, an individual i of generation t has utility

$$U_{it} = y_{it} - \frac{1}{2\alpha} e_{it}^2,$$

where $y_{it} \in \{0, 1\}$ can be thought of as success or failure, and e_{it} is the effort level. The set of individuals is denoted by $\mathcal{I}$ and is taken to be odd for simplicity.

Suppose that success depends on effort and also on

$$\mathbb{P}\left(y_{it} = 1 \mid e_{it} = e \text{ and } y_{it-1} = 0\right) = \pi_0 + \theta e,$$

and

$$\mathbb{P}\left(y_{it} = 1 \mid e_{it} = e \text{ and } y_{it-1} = 1\right) = \pi_1 + \theta e,$$

where $\pi_1 \geq \pi_0$. The gap between these two parameters is the importance of "inheritance" in success, whereas θ is the importance of "hard work". The vector of parameters (θ, π_0, π_1) is unknown. At any given point in time, individuals will have a posterior over this policy vector μ_{it}, shaped by their dynasty's prior experiences as well as other characteristics in the society that they may have observed. The only policy tool is a tax rate on output, which is then redistributed lump sum. Let total output under tax rate τ be $Y(\tau)$.

This implies that given an expectation of a tax rate τ, an individual with a successful or unsuccessful parent denoted by $z = 1$ or $z = 0$ will choose

$$e_z(\tau, \mu) \in \arg\max_e \mathbb{E}_\mu \left[(1 - \pi_z - \theta e)\, \tau Y(\tau) + (\pi_z + \theta e)\left((1 - \tau) + \tau Y(\tau)\right)\right] - \frac{1}{2\alpha} e^2,$$

where the expectation is over the parameters. It can be easily verified that

$$e_z(\tau,\mu) = e(\tau, \mathbb{E}_\mu\theta) = \alpha(1-\tau)\,\mathbb{E}_\mu\theta.$$

Therefore, all that matters for effort is the expectation about the parameter θ.

Now given this expectation, individuals will also choose the tax rate. This choice will be made by voting. Given the quadratic utility function, it can be verified that individuals have single peaked preferences, with bliss point given by

$$\tau(\mu_{it}) \in \arg\max \mathbb{E}_{\mu_{it}} V_t,$$

where $\mathbb{E}_{\mu_{it}} V_t$ is the expectation of an individual with beliefs μ_{it} regarding social welfare at time t. An application of the median voter theorem then gives the equilibrium tax rate is the median of these bliss points.

How will an individual update their beliefs? Straightforward application of Bayes rule gives the evolution of beliefs. For example, for an individual $i \in \mathcal{I}$ with a successful or unsuccessful parent denoted by $z = 1$ or $z = 0$, starting with beliefs μ_{it}, with support $S[\mu_{it}]$, we have that for any $(\theta, \pi_0, \pi_1) \in S[\mu_{it}]$, we have

$$\mu_{it+1}(\theta,\pi_0,\pi_1) = \mu_{it}(\theta,\pi_0,\pi_1)\frac{\pi_z + \theta e\left(\tau_t, \mathbb{E}_{\mu_{it}}\theta\right)}{\displaystyle\int \left[\pi_z' + \theta' e\left(\tau_t, \mathbb{E}_{\mu_{it}}\theta\right)\right] d\mu_{it}}.$$

Note that individuals here are not learning from the realized tax rate, simply from their own experience. This is because individuals are supposed to have "heterogeneous priors". They thus recognize that others have beliefs driven by their initial priors, which are different from theirs and there is no learning from initial priors. [Nevertheless, one could argue that there should be learning from changes in initial priors].

Standard results about Bayesian updating, in particular from the martingale convergence theorem, imply the following:

Proposition 7.3. *The beliefs of individual $i \in \mathcal{I}$, μ_{it}, starting with any initial beliefs μ_{i0} almost surely converges to a stationary belief $\mu_{i\infty}$.*

But if beliefs converge for each dynasty, then the median also converges, and thus equilibrium tax rates also converge.

Proposition 7.4. *Starting with any distribution of beliefs in the society, the equilibrium tax rate τ_t almost surely converges to a stationary tax rate τ_∞.*

The issue, however, is that this limiting tax rate need not be unique, because the limiting stationary beliefs are not necessarily equal to the distribution that puts probability 1 on truth. The intuition for this is the same as "self confirming" equilibria, and can be best seen by considering an extreme set of beliefs in the society that lead to $\tau = 1$ (because effort doesn't matter at all). If $\tau = 1$, then nobody expects any effort and there is no possibility that anybody can learn that effort actually matters.

The characterization of the set of possible limiting beliefs is straightforward. Define $M^*(\tau)$ be the set of beliefs that are "self consistent" at the tax rate τ in the following sense: For any $\tau \in [0,1]$, we have

$$\begin{aligned} M^*(\tau) &= \{\mu : \text{for all } (\theta, \pi_0, \pi_1) \in S[\mu], \\ &\quad \pi_z + \theta e(\tau, \mathbb{E}_\mu \theta) = \pi_z^* + \theta^* e(\tau, \mathbb{E}_\mu \theta) \text{ for } z = 0,1 \text{ and } (\theta^*, \pi_0^*, \pi_1^*) \in S[\mu]\}. \end{aligned}$$

Intuitively, these are the set of beliefs that generate the correct empirical frequencies in terms of upward and downward mobility (success and failure) given the effort level that they imply. Clearly, if the tax rate is in fact τ and $M^*(\tau)$ is not a singleton, a Bayesian cannot distinguish between the elements of $M^*(\tau)$: they all have the same observable implications.

Now the following result is immediate.

PROPOSITION 7.5. *Starting with any initial distribution of beliefs in society $\{\mu_{i0}\}_{i\in\mathcal{I}}$, we have that*

(1) *For all $i \in \mathcal{I}$, $\mu_{i\infty}$ exists and is in $M^*(\tau_\infty)$, and*
(2) *τ_∞ is the median of $\{\tau(\mu_{i\infty})\}_{i\in\mathcal{I}}$.*

This proposition of course does not rule out the possibility that there will be convergence to beliefs corresponding to the true parameter values regardless of initial conditions. But it is straightforward from the above observations establish the next result:

PROPOSITION 7.6. *Suppose $\mathcal{I}$ is arbitrarily large. Then for any $\{\mu_{i\infty}\}_{i\in\mathcal{I}} \in M^*(\tau_\infty)$ such that τ_∞ is the median of $\tau(\mu_{i\infty})$, there exists a set of initial conditions such that there will be convergence to beliefs $\{\mu_{i\infty}\}_{i\in\mathcal{I}}$ and tax rate τ_∞ with probability arbitrarily close to 1.*

This proposition implies that a society may converge and remain in equilibria with very different sets of beliefs and these beliefs will support different amounts of redistribution. Different amounts of redistribution will then lead to different tax rates, which "self confirm" these beliefs because behavior endogenously adjusts to tax rates. Therefore, according to this model, one could have the United States society converge to a distribution of beliefs in which most people believe that θ is high and thus vote for low taxes, and this in turn generates high social mobility, confirming the beliefs that θ is high. Many more Europeans believe that θ is low (and correspondingly $\pi_1 - \pi_0$ is high) and this generates more redistribution and lower social mobility. Neither Americans nor Europeans are being “irrational”.

7.3. References

(1) Austen-Smith, David (1991) “Rational Consumers and Irrational Voters,” *Economics and Politics*, 3:73-92.

(2) Fedderson, Timothy and Wolfgang Pesendorfer (1996) “The Swing Voter’s Curse,” *American Economic Review*, 86: 408-424.

(3) Fey, Mark and J. Kim (2002) “The Swing Voter’s Curse: a Comment,” *American Economic Review*, 92: 1264-1268.

(4) Piketty, Thomas, (1995) “Social Mobility and Redistributive Politics, *Quarterly Journal of Economics*, 110: 551-584.

(5) Strulovici, Bruno (2010) “Voting and Experimentation,” forthcoming *Econometrica.*

CHAPTER 8

Political Agency and Electoral Control

One of the most common conceptions of democratic politics is that based on the principal-agent relationship. Politicians are viewed as agents and the voters are the principals, will incentivize the agents using electoral strategies—voting them out of office when they are not satisfied with their performance. This approach goes back to Barro (1972) and Ferejohn (1986). The political agency model is a very parsimonious approach and is useful for a range of analysis. But as we will discuss below, it also has various limitations. Here we will see the basic model, which makes a range of assumptions (retrospective voting, stationary strategies) and then look at a more game Furyk analysis in the context of a richer model.

8.1. Basic Retrospective Voting Models

The political agency literature typically focuses on the issues by simplifying the voting side. I will follow this practice here, and start with a simple *retrospective* voting strategy on the part of all voters, such that they will replace the politician (with some other party or politician) in case his or her performance is worse than the benchmark.

8.1.1. Static model with full information. To start with, consider the following environment: there is some parameter θ, capturing the cost of supplying public goods. In particular, the government budget constraint is

$$\tau\bar{y} = \theta g + r \tag{8.1}$$

where $\bar{y}$ is average income, τ is the tax rate, g is the level of public good provision, r is the rents captured by the politician or political party in power, and θ is a random variable with distribution $F(\cdot)$, whereby high values of θ correspondent to high costs of providing the public good.

Let the indirect utility function of the representative (or decisive) voter, as a function of the state of the world, the level of public good and the rents captured by the politician, be

$$U(g(\theta), r(\theta), \theta) = \bar{y} - \theta g(\theta) - r(\theta) + H\left(g(\theta)\right),$$

so that the function H captures their valuation of the public good. Assume that H is increasing and strictly concave.

Let $g^*(\theta)$ denote the preferred level of public good when the cost is θ, given by

$$H'(g^*(\theta)) = \theta.$$

The important simplifying feature here comes from the quasi-linear preferences which make $g^*(\theta)$ independent of r, and immediately imply that $g^*(\theta)$ increasing.

We start with an incumbent politician with the following utility function

$$\mathbb{E}(\upsilon_I) = r + p_I R,$$

where R is the total rent from continuing to stay in power, and p_I is the probability that the voters will keep the incumbent in power.

Since we are in the full information, voters observe θ.

Suppose that voters coordinate on the following retrospective voting strategy:

$$p_I = \begin{cases} 1 & \text{if } U(g(\theta), r(\theta), \theta) \geq \varpi(\theta) \\ 0 & \text{otherwise,} \end{cases} \tag{8.2}$$

In other words, the voters will keep the incumbent in power, $p_I = 1$, if he delivers utility greater than $\varpi(\theta)$ to them. [It is in fact possible to check that given the assumptions made here (8.2) is the optimal strategy; see homework].

Given (8.2), the politician solves the following program:

$$\max r(\theta) + p_I R$$

subject to

$$p_I = 1 \text{ only if } \bar{y} - \theta g(\theta) - r(\theta) + H(g(\theta)) \geq \varpi(\theta). \tag{8.3}$$

Let us first look for a solution in which $p_I = 1$. It is then clear that the politician will never leave the constraint slack. Moreover, since $g(\theta)$ does not appear in the objective function, it will be chosen so as to maximize the choice set of the politician, i.e., make (8.3) as slack as possible. Naturally this means setting $g(\theta) = g^*(\theta)$, therefore:

$$r(\theta) = \bar{y} - \theta g^*(\theta) - \varpi(\theta) + H(g^*(\theta)).$$

The next step is to check that the politician indeed prefers $p_I = 1$. With $p_I = 1$, he receives

$$\bar{y} - \theta g^*(\theta) - \varpi(\theta) + H(g^*(\theta)) + R.$$

In contrast, if he chooses $p_I = 0$, then he sets the highest possible level of r, $r = \bar{y}$, and obtains $\bar{y}$. Thus for the politician to behave according to the wishes of the voters, we need

$$\gamma\bar{y} - \theta g^*(\theta) - \varpi(\theta) + H(g^*(\theta)) + R \geq \bar{y} \tag{8.4}$$

Now we have to think about how $\varpi(\theta)$ is determined. Trivially, there is a coordination problem, so various different levels of $\varpi(\theta)$ could emerge. Let us ignore this coordination problem and imagine that voters can coordinate on the best $\varpi(\theta)$. This will be such that (8.4) holds as equality:

$$\varpi(\theta) = [H(g^*(\theta)) - \theta g^*(\theta)] + R,$$

as long as $\bar{y} \geq R$ (if this inequality does not hold, then (8.4) cannot hold as an equality—assuming that there cannot be negative rents). This voting rule (with full information) then yields rents equal to:

$$r^* = \bar{y} - R$$

The important point here is that the incumbent politician is receiving rents despite the fact that there is electoral control over him. This is because of the advantage of incumbency, in other words, his ability to control policy. The issues will become clear when we look at a dynamic model next.

8.1.2. Dynamic model with full information. The only modification now is that the objective function of the politician at time $t = 0$ is changed to

$$\mathbb{E}(\upsilon_I) = \sum_{t=0}^{\infty} \beta^t r_t,$$

with the convention that if $p_t = 0$, so that he is not elected at time t, then $r_{t+k} = 0$ for all $k > 0$.

The budget constraint is not truly modified to:

$$\tau_t \bar{y} = \theta_t g_t + r_t$$

(in other words, there needs to be budget balance in every period).

Let us look for a stationary Markov voting rule, where voters only condition on current performance in deciding whether to elect the politician. Therefore, the voting rule takes the form

$$p_t = \begin{cases} 1 & \text{if } U(g_t(\theta_t), r_t(\theta_t), \theta_t) \geq \varpi(\theta_t) \\ 0 & \text{otherwise,} \end{cases} \tag{8.5}$$

How is $\varpi(\theta_t)$ determined? Now essentially the logic is almost the reverse of the static model. In any given period, the politician can choose $r_t = \bar{y}$, so the problem is to convince him not to do so. Let $R_{t+1} = \sum_{t=1}^{\infty} \beta^{t-1} r_t$ be the continuation value . Then, we have an incentive compatibility constraint for the politician of the form:

$$r_t + \beta R_{t+1} \geq \bar{y}$$

If this condition is violated, the politician will steal everything. Therefore, the objective of the citizens is to make sure that this constraint is satisfied while giving the least amount of rents to the politician. It should be clear that the best way of doing so is to have the politician remain in power with probability one as long as he follows the prescribed policy. Let r be the per period rent, which is constant by the stationarity assumption, and such an equilibrium. Then

$$R_{t+1} = \frac{r}{1-\beta}$$

and we have

$$r = (1-\beta)\,\bar{y}.$$

This then implies that voters set the retrospective voting rule (8.5) with

$$\varpi(\theta) = \beta\bar{y} - \theta g^*(\theta) + H(g^*(\theta)),$$

which will deliver exactly $r(\theta) = (1-\beta)\,\bar{y}$ to the politician in all states.

The dynamic model, therefore, highlights more clearly that the source of the power of the politician is his ability to choose the current policy, and voters have to adjust their voting rule so as to leave him sufficient rents.

8.2. Agency with Asymmetric Information

Now suppose that voters do not observe θ, and let's understand the structure of the static model in this case.

Since voters cannot observe θ, the only static voting policy they can choose is of the form

$$p_I = 1 \quad \text{iff} \quad U(g_t(\theta_t), r_t(\theta_t), \theta_t) \geq \varpi.$$

(here the optimality of voting rules is much more complicated, see the problem set).

Now let us look at the politician's strategy. The same argument as above, implies that if the politician decides to satisfy the voters, he will choose

$$r(\theta) = \bar{y} - \theta g^*(\theta) - \varpi + H(g^*(\theta)).$$

[there is an important point here that the choice of g will reveal the underlying state, but voters do not condition the election rule on this].

In this case, the politician receives

$$\bar{y} - \theta g^*(\theta) - \varpi + H(g^*(\theta)) + R.$$

It is also satisfying the constraint that voters have put as an equality, so

$$U(g_t(\theta_t), r_t(\theta_t), \theta_t) = \varpi.$$

In contrast, as before, if he chooses $p_I = 0$, then he sets the highest possible level of r, $r = \bar{y}$, and obtains $\bar{y}$. Thus, the politician will get reelected if

$$H(g^*(\theta)) - \theta g^*(\theta) - \varpi + R \geq 0$$

Define θ^* such that

$$H(g^*(\theta^*)) - \theta^* g^*(\theta^*) - \varpi + R = 0 \tag{8.6}$$

Then it is clear that voters' utility will be

$$U(g_t(\theta_t), r_t(\theta_t), \theta_t) = \begin{cases} \varpi & \text{for } \theta \leq \theta^* \\ 0 & \theta > \theta^* \end{cases}$$

Therefore, their expected utility is

$$\mathbb{E}(U) = \int_{\underline{\theta}}^{\theta^*} \varpi dF(\theta) + \int_{\theta^*}^{\overline{\theta}} 0 \cdot dF(\theta) = F(\theta^*)\varpi \tag{8.7}$$

Lets us next look at the relationship between ϖ and θ from (8.6). The implicit function theorem gives:

$$\begin{aligned} \frac{d\theta^*}{d\varpi} &= \frac{1}{[H'(g^*(\theta^*)) - \theta^*]\,(g^*)'(\theta^*) - g^*(\theta^*)} \\ &= -\frac{1}{g^*(\theta^*)} < 0, \end{aligned}$$

where the second equality uses the envelope theorem. This highlights the major trade-off that voters face: if they set a high level of ϖ, then their utility is higher when the politician behaves well, but there is a higher probability that he will not do so.

Now consider the maximization of with welfare respect to ϖ:

$$\max_{\varpi} \mathbb{E}(U) = F(\theta^*)\varpi$$

The first-order condition for this is

$$\begin{aligned} F(\theta^*) - f(\theta^*)\,\varpi \frac{d\theta^*}{d\varpi} &= 0 \\ F(\theta^*) - \frac{f(\theta^*)\,\varpi}{g^*(\theta^*)} &= 0 \end{aligned}$$

The interpretation is as follows. A unit increase in ϖ benefits the citizens by an amount $F(\theta^*)$, since this is the probability with which they will receive this increase in welfare. The cost of the increase in ϖ, on the other hand, is the decline in $F(\theta^*)$, the probability with which the citizens will receive this amount, times ϖ. This decline in probability is $-1/g^*(\theta^*)$.

Rearranging, we have

$$\frac{\varpi}{g^*(\theta^*)} = \frac{F(\theta^*)}{f(\theta^*)} \tag{8.8}$$

or alternatively

$$\frac{d\varpi/d\theta^*}{\varpi} = -\frac{f(\theta^*)}{F(\theta^*)}$$

An expression of the form F/f (or f/F) often appears in models with distribution of types, and corresponds to the hazard rate, that is, the relative likelihood that $\theta = \theta^*$ conditional on $\theta \leq \theta^*$. Under the assumption that the hazard rate is monotonically decreasing, which is another standard assumption, (**??**), defines a unique θ^*. [This assumption also corresponds to the assumption that $F(\theta)$ is log concave. To see this, note that the monotone likelihood condition means:

$$\frac{f'(\theta)F(\theta^*) - (f(\theta))^2}{F(\theta)} < 0,$$

in other words, $f'(\theta)$ should not be too positive. Similarly, log concavity means that $\ln F(\theta)$ should be concave, or

$$\frac{f(\theta)}{F(\theta)}$$

should be decreasing. Most familiar distribution functions such as uniform, normal etc. satisfy this assumption.].

The monotone likelihood ratio assumption implies that the right hand side of (8.8) is decreasing in θ^*, while the fact that $g^*(\theta^*)$ is decreasing makes the left-hand side increasing, so θ^* is uniquely defined.

Now, consider a change in the distribution of the costs of public goods from $F(\theta)$ to $\tilde{F}(\theta)$ such that $F(\theta^*) < \tilde{F}(\theta^*)$ and $f(\theta^*) = \tilde{f}(\theta^*)$, so that the probability that $\theta \leq \theta^*$ has increased, but the density of θ around θ^* has not. In this case, the first-condition above implies that ϖ should increase, since there is a higher likelihood of any given level of ϖ to be delivered to the citizens.

What is the effect of asymmetric information on the provision of public goods (or voters ex post welfare). Recall that with symmetric information we had

$$g^F(\theta) = g^*(\theta)$$

In contrast, with incomplete information, we have

$$g^I(\theta) = \begin{cases} 0 & \text{for } \theta > \theta^* \\ g^*(\theta) & \text{for } \theta \leq \theta^* \end{cases}$$

Notice that, when $\theta \leq \theta^*$, the provision of the public good is not distorted because of the form of the voting rule (and quasi-linearity of preferences).

Therefore, $g^I(\theta) \leq g^F(\theta)$. What about rents?

$$r(\theta) = \begin{cases} y & \text{for } \theta > \theta^* \\ r^* + (\tau^*(\theta^*) - \tau^*(\theta))y + H(g^*(\theta)) - H(g^*(\theta^*)) & \text{for } \theta \leq \theta^* \end{cases}$$

Since both $y > r^*$ and $(\tau^*(\theta^*) - \tau^*(\theta))y + H(g^*(\theta)) - H(g^*(\theta^*)) \geq 0$, politicians always receive greater rents with asymmetric information.

This is intuitive: asymmetric information increases the control power of politicians (because voters cannot monitor them as well).

8.3. Career Concerns

A major shortcoming of the previous models is that voters use retrospective voting rules, rather than pursue policies that are in their future interests. One way of overcoming this problem is to look at a different type of agency models, inspired by Holmstrom's career concerns model.

I give a brief overview here.

The main idea is that there is a feature of the politician that voters care about, which also affects outcomes. Then they learn about this feature from past performance. In addition, the politician can also improve performance by other means, so this creates an incentive for the politician to behave better in order to convince the voters that he is of the good type.

Let us illustrate the main issues using a two-period model here. Moreover, let us assume that taxes are fixed at $\bar{\tau}$, so the only problem is to make sure that the tax revenue is spent on public goods. The welfare of the voters is again

$$U_t = \bar{y}(1 - \bar{\tau}) + g_t.$$

Now the technology for public good provision is different, and takes the form

$$g_t = \eta(\bar{\tau}\bar{y} - r_t),$$

where η is the "ability" of the politician, which is fixed in both periods. Let us assume that it is drawn uniformly from the interval

$$\left[1 - \frac{1}{2\xi}, 1 + \frac{1}{2\xi}\right].$$

The important simplifying assumption of the Holmstrom model, which we adopt here, is that there is symmetric information, so the politician is also uncertain about η with the same prior.

The utility of the politician is a simple extension of what we had before:

$$v_I = r_1 + p_I\beta(R + r_2),$$

with $0 < \beta < 1$ again as the discount factor, and now R is interpreted as non-pecuniary grants from being in power.

Moreover, let us assume that there is a maximum on the rents that the politician can extract, $\bar{r}$.

The exact timing of events is as follows:

- Nature determines η.
- The politician chooses r_1.
- Observing g_1 (but not r_1), voters decide whether to keep the politician. If they elect a new politician, he is drawn randomly from the same distribution.
- The politician and power chooses g_2.

Given this structure, the equilibrium is straightforward to determine.

In the second period, there is no control over the politician, so he will set

$$r_2 = \bar{r},$$

and public goods will be

$$g_2 = \eta(\bar{\tau}y - \bar{r})$$

If they appoint a new politician, he will have $\mathbb{E}(\eta) = 1$, so the expected utility of appointing a new politician for the voters is

$$U_2^N = \bar{y}(1 - \bar{\tau}) + (\bar{\tau}y - \bar{r})$$

What about the utility of keeping the incumbent? This would be

$$U_2^I = \bar{y}(1 - \bar{\tau}) + \tilde{\eta}(\bar{\tau}y - \bar{r})$$

where $\tilde{\eta}$ is their posterior about his ability.

Now suppose that voters know that the politician will choose $\tilde{r}_1$ amounts of rents for himself. Then they can estimate

$$\tilde{\eta} = \frac{g_1}{\bar{\tau}y - \tilde{r}_1}$$

and their optimal reelection decision is

$$\tilde{p}_I = \begin{cases} 1 & \text{iff } \tilde{\eta} \geq \mathbb{E}(\eta) = 1 \\ 0 & \text{otherwise.} \end{cases}.$$

The problem is that $\tilde{r}_1$ is an equilibrium choice by the politician. He will try to make this choice in order to ensure that he remains in power if this is beneficial for him. This is why this class of models are sometimes called "signal jamming" models.

To make more progress, let us first look at the probability that he keeps power. This is

$$\begin{aligned} p_I &= \text{Prob}\left[\tilde{p}_I = 1\right] \\ &= \text{Prob}\left[\tilde{\eta} \geq 1\right] \\ &= \text{Prob}\left[\frac{g_1}{\bar{\tau}\bar{y} - \tilde{r}_1} \geq 1\right] \\ &= \text{Prob}\left[\frac{\eta(\bar{\tau}\bar{y} - r_1)}{\bar{\tau}\bar{y} - \tilde{r}_1} \geq 1\right] \\ &= \frac{1}{2} + \xi\left[1 - \frac{\eta(\bar{\tau}\bar{y} - r_1)}{\bar{\tau}\bar{y} - \tilde{r}_1}\right] \end{aligned}$$

where the last equality exploits the uniform assumption. Now the incumbent will choose r_1 to maximize $\upsilon_I = r_1 + p_I\beta(R + r_2)$, which we can write as:

$$\max_{r_1} \upsilon_I = r_1 + \left[\frac{1}{2} + \xi\left(1 - \frac{\eta(\bar{\tau}\bar{y} - r_1)}{\bar{\tau}\bar{y} - \tilde{r}_1}\right)\right]\beta(R + \bar{r})$$

The first-order condition for this gives

$$1 - \frac{\xi(\bar{\tau}\bar{y} - \tilde{r}_1)}{(\bar{\tau}\bar{y} - r_1)^2}\beta(R + \bar{r}) = 0 \tag{8.9}$$

This defines a best-response $r_1(\tilde{r}_1)$ by the incumbent. When voters expect them to play $\tilde{r}_1$, he would play $r_1(\tilde{r}_1)$. Clearly, the equilibrium has to be a fixed point, $r_1(\tilde{r}_1) = \tilde{r}_1$. Substituting this into (8.9), we obtain

$$r_1 = \bar{\tau}y - \xi\beta(R + \bar{r})$$

and the politician keeps power with probability $p_I = \frac{1}{2}$, since in equilibrium nobody's fooled, and with probability 1/2 the politician is worse than average.

This is therefore a more satisfactory model of deriving results in which elections appear as a method of controlling politicians. In fact, we can alternatively express this equilibrium as remarking that voters will reelect the incumbent only if

$$g_1 = \xi\beta(R + \tilde{r})\eta \geq \xi\beta(R + \bar{r})$$

or their utility exceeds a certain threshold as in the retrospective voting models:

$$\bar{y}(1 - \bar{\tau}) + \xi\beta(R + \tilde{r})\eta \geq \varpi \equiv \bar{y}(1 - \bar{\tau}) + \xi\beta(R + \bar{r}),$$

but such rules are not retrospective or pre-determined punishment rules; instead they are derived from the forward-looking optimum behavior of the voters.

8.4. Political Economy of Mechanisms

Let us now relaxed the assumption that citizens have to vote retrospectivelyand that they have to use stationary strategies. Therefore, we will look for subgame perfect equilibria in a political agency model. In particular, since the focuses on how well electoral controls work, let us look for the "best subgame perfect equilibrium" from the viewpoint of the citizens. Since the best subgame perfect equilibrium will be similar to a "mechanism," let us refer to this as *the best sustainable mechanism.* Also, let us enrich the environment so that there is labor supply and capital accumulation as in the standard neoclassical model, so that distortions created by political economy can be seen more transparently.

The model is in infinite horizon and discrete time, and it is populated by a continuum of measure 1 of identical individuals (citizens). Individual preferences at time $t = 0$ are given by

$$\sum_{t=0}^{\infty} \beta^t U(c_t, l_t),$$

where c denotes consumption and l is labor supply. We denote the set of citizens by I and use the subscript i to denote citizens. We impose the standard conditions on U:

Assumption 8.1. ***(utility)*** *$U(c, l)$ is twice continuously differentiable with partial derivatives denoted by U_C and U_L, strictly increasing in c, strictly decreasing in l and jointly concave in c and l. We adopt the normalization $U(0, 0) = 0$. Moreover, $l \in [0, \bar{L}]$.*

The production side of the economy is described by the aggregate production function

$$Y_t = F(K_t, L_t), \tag{8.10}$$

which is defined inclusive of undepreciated capital (i.e., $F(K_t, L_t) \equiv \tilde{F}(K_t, L_t) + (1 - \theta) K_t$ for some other production function $\tilde{F}(K, L)$ and for some depreciation rate $\theta \in (0, 1)$).

Assumption 8.2. ***(production structure)*** *F is strictly increasing and continuously differentiable in K and L with partial derivatives denoted by F_K and F_L, exhibits constant returns to scale, and satisfies $\lim_{L \to 0} F_L(K, L) = \infty$ for all $K \geq 0$ and $\lim_{K \to \infty} F_K(K, L) < 1$ for all $L \in [0, \bar{L}]$.*

The condition that $\lim_{K \to \infty} F_K(K, L) < 1$ together with $L \in [0, \bar{L}]$ implies that there is a maximum steady-state level of output is uniquely defined by $\bar{Y} = F(\bar{Y}, \bar{L}) \in (0, \infty)$. The condition that $\lim_{L \to 0} F_L(K, L) = \infty$ implies that in the absence of distortions there will be positive production.

The allocation of resources is delegated to a politician (ruler). The fundamental political dilemma faced by the society is to ensure that the body to which these powers have been delegated does not use them for its own interests. In the current model, this fundamental dilemma is partly resolved by the control of the politicians via elections.

We assume that there is a large number of potential (and identical) politicians, denoted by the set $\mathcal{I}$. Each politician's utility at time t is given by

$$\sum_{s=0}^{\infty} \delta^s v\left(x_{t+s}\right),$$

where x denotes the politician's consumption (rents) and $v : \mathbb{R}_+ \to \mathbb{R}$ is his instantaneous utility function. Notice also that the politician's discount factor, δ, is potentially different from that of the citizens, β. To simplify the analysis, we assume that potential politicians are distinct from the citizens and never engage in production and that once they are replaced they do not have access to capital markets.

ASSUMPTION 8.3. ***(politician utility)*** *v is twice continuously differentiable, concave, and satisfies $v'(x) > 0$ for all $x \in \mathbb{R}_+$ and $v(0) = 0$. Moreover $\delta \in (0, 1)$.*

The politician in power decides the allocation of resources (or equivalently decides a general set of taxes and transfers). The only restriction on the allocation of resources, in addition to $c_t \geq 0$ and $l_t \in \left[0, \bar{L}\right]$, comes from the *participation constraint* of the citizens, which requires that $U(c_t, l_t) \geq 0$ for each t. We denote the three constraints $c_t \geq 0$, $l_t \in \left[0, \bar{L}\right]$ and $U(c_t, l_t) \geq 0$ by

$$(c_t, l_t) \in \Lambda \text{ for all } t. \tag{8.11}$$

Since $U(c, l)$ is concave and continuous, Λ is closed and convex (and also nonempty). We use IntΛ to denote the interior of the set Λ, so that $(c_t, l_t) \in$IntΛ implies that $c_t > 0$, $l_t \in \left(0, \bar{L}\right)$ and $U(c_t, l_t) > 0$.

We consider the following game. At each time t, the economy starts with a politician $\iota_t \in \mathcal{I}$ in power and a stock of capital inherited from the previous period, K_t. Then:

1. Citizens make labor supply decisions, denoted by $[l_{i,t}]_{i \in I}$, where $l_{i,t} \geq 0$. Output $F(K_t, L_t)$ is produced, where $L_t = \int_{i \in I} l_{i,t} di$.

2. The politician chooses the amount of rents $x_t \in \mathbb{R}_+$, a consumption function $\mathbf{c}_t : \mathbb{R}_+ \to \mathbb{R}_+$, which assigns a level of consumption for each level of (current) labor supply, and next period's capital stock $K_{t+1} \in \mathbb{R}_+$, subject to the constraint

$$K_{t+1} \leq F(K_t, L_t) - C_t - x_t,$$

where $C_t = \int_{i \in I} \mathbf{c}_t(l_{i,t}) \, di$ is aggregate consumption. We denote a triple $(x_t, \mathbf{c}_t, K_{t+1})$ that is feasible for the politician by $(x_t, \mathbf{c}_t, K_{t+1}) \in \Phi_t$.

3. Elections are held and citizens jointly decide whether to keep the politician or replace him with a new one, $\rho_t \in \{0, 1\}$, where $\rho_t = 1$ denotes replacement.

The important feature here is that even though individuals make their economic decisions independently, they make their political decisions—elections to replace the politician—jointly. This is natural since there is no conflict of interest among the citizens over the replacement decision. Joint political decisions can be achieved by a variety of procedures, including various voting schemes (e.g., Persson and Tabellini, 2000). Here we simply assume that the decision $\rho_t \in \{0, 1\}$ is taken by a randomly chosen citizen.

We assume that at each date there is a public random variable z_t and all agents can condition their strategies on the history of this variable. This will enable us to convexify the value function of the citizens. Let

$$h^t \equiv \left(K_0, \iota_0, z_0, [l_{i,0}]_{i \in I}, x_0, \mathbf{c}_0, \rho_0, K_1, ...K_t, \iota_t, z_t, [l_{i,t}]_{i \in I}, x_t, \mathbf{c}_t, \rho_t, K_{t+1}\right)$$

denote the history of the game up to date t, and H^t be the set of all such histories. In the text, to simplify notation we suppress the conditioning on the history of z^t. A *subgame perfect equilibrium* (SPE) is given by labor supply decisions $\left[l^*_{i,t}\right]_{i \in I}$ at time t given history h^{t-1}, policy decisions $x^*_t, \mathbf{c}^*_t, K^*_{t+1}$ by the politician in power given h^{t-1} and $[l_{i,t}]_{i \in I}$, and electoral decisions by the citizens, ρ^*_t at time t, given history h^{t-1} and $[l_{i,t}]_{i \in I}$, $x^*_t, \mathbf{c}^*_t, K^*_{t+1}$ that are best responses to each other for all histories. In addition, we will show below that the SPE we focus on are "renegotiation-proof". Although the issue of how renegotiation should be handled in dynamic games is not settled and there are many alternative notions in the literature (e.g., Fudenberg and Tirole, 1994), for our purposes the simplest notion of renegotiation-proofness is sufficient. In particular, we say that a SPE is *renegotiation-proof* if after any history h^t there does not exist another SPE that can make all active players weakly better off (and some strictly better off). In the present context, this implies that there should not exist an alternative SPE that can make the citizens and the politician in power better off than in the candidate SPE. Note is that in the definition of renegotiation-proofness, if the the qualification "all active players" were removed, the concept would lose its bite and renegotiation-proofness would be much easier to guarantee. This is because an alternative SPE might make the citizens and the current politician better off, but reduce the utility of some future politician, who becomes less likely to come to power. The definition of renegotiation-proofness here is more demanding and more interesting in the context of political games.

We focus on *best SPE*, defined as a SPE that maximizes the utility of the citizens. Consider the following constrained optimization problem:

$$\textbf{MAX}: \max_{\{C_t, L_t, K_{t+1}, x_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t U(C_t, L_t) \tag{8.12}$$

subject to the participation constraint (8.11), the resource constraint,

$$C_t + K_{t+1} + x_t \leq F(K_t, L_t) \text{ for all } t, \tag{8.13}$$

the sustainability constraint for the politician in power,

$$w_t \equiv \sum_{s=0}^{\infty} \delta^s v(x_{t+s}) \geq v(F(K_t, L_t)) \text{ for all } t, \tag{8.14}$$

and given the initial capital stock $K_0 > 0$. We have written this program using capital letters, since the consumption and labor supply levels refer both to individual and aggregate quantities. Notice also that in (8.14) we have defined w_t as the expected discounted utility of the politician at time t. This notation will be used in Theorem 8.1 below.

The sustainability constraint, (8.14), requires the equilibrium utility of the politician to be such that he does not wish to choose the maximum level of rents this period, $x_t = F(K_t, L_t)$, which would give him utility $v(F(K_t, L_t))$. We refer to a sequence $\{C_t, L_t, K_{t+1}, x_t\}_{t=0}^{\infty}$ that is a solution to this problem as a *best sustainable mechanism* (since it implicitly defines a resource allocation mechanism). The constraint, (8.14), is sufficient to ensure that the politician does not wish to deviate from the mechanism.

Proposition 8.1. *The allocation of resources in the best SPE (best sustainable mechanism) is identical to the solution of the maximization problem in* (MAX) *and involves no replacement of the initial politician along the equilibrium path. Moreover, this allocation can be supported as a renegotiation-proof SPE.*

Proof. First, in view of the concavity of U, no (possibly stochastic) feasible allocation can provide higher ex ante utility to citizens than $\{\tilde{C}_t, \tilde{L}_t, \tilde{K}_{t+1}, \tilde{x}_t\}_{t=0}^{\infty}$ that is a solution to (MAX). If it did, it would either violate the participation constraint, (8.11), the resource constraint, (8.13), or the sustainability constraint, (8.14), after some history h^t, and would thus not be feasible. Therefore, to prove the proposition it suffices to show that there exists a renegotiation-proof SPE achieving the solution to (MAX).

Let $\{\tilde{C}_t, \tilde{L}_t, \tilde{K}_{t+1}, \tilde{x}_t\}_{t=0}^{\infty}$ be a solution to (MAX). We next show that it can be supported as a SPE with no politician replacement along the equilibrium path. Introduce the following notation: $h^t = \hat{h}^t$ if $(K_{s+1}(h^s), x_s(h^s)) = (\tilde{K}_{s+1}, \tilde{x}_s)$ and $\mathbf{c}_s(l_{i,s} \mid h^s) = \tilde{C}_s$ for $l_{i,s} = \tilde{L}_s$ and $\mathbf{c}_s(l_{i,s} \mid h^s) = 0$ for $l_{i,s} \neq \tilde{L}_s$ for all $s \leq t$. Consider the strategy profile ρ for the citizens such that $\rho(h^t) = 0$ if $h^t = \hat{h}^t$ and $\rho(h^t) = 1$ if $h^t \neq \hat{h}^t$; that is, citizens replace the politician unless the politician has always chosen a strategy inducing the allocation $\{\tilde{C}_s, \tilde{L}_s, \tilde{K}_{s+1}, \tilde{x}_s\}_{s=0}^{t}$ up at to time t. It is a best response for the politician to continue to choose $\{\tilde{C}_s, \tilde{L}_s, \tilde{K}_{s+1}, \tilde{x}_s\}_{s=t}^{\infty}$

after history $h^{t-1} = \hat{h}^{t-1}$ only if

$$\mathbb{E}\left[\sum_{s=0}^{\infty} \delta^s v\left(\tilde{x}_{t+s}\left(h^{t+s}\right)\right) \mid h^t\right] \geq \max_{\left(x_t',\mathbf{c}_t',K_{t+1}'\right)\in\Phi_t} \mathbb{E}\left[v\left(x_t'\right) + \delta v_t^c\left(K_{t+1}', \mathbf{c}_t', x_t'\right) \mid h^t\right]$$

where $v_t^c\left(x_t', \mathbf{c}_t', K_{t+1}'\right)$ is the politician's continuation value following a deviation to a feasible $\left(x_t', \mathbf{c}_t', K_{t+1}'\right)$. Under the candidate equilibrium strategy, $v^c = 0$ following any deviation, thus the best deviation for the politician is $x_t' = F(\tilde{K}_t, \tilde{L}_t)$, which gives (8.14). Consequently, (8.14) is sufficient for the politician not to deviate from $\{\tilde{C}_t, \tilde{L}_t, \tilde{K}_{t+1}, \tilde{x}_t\}_{t=0}^{\infty}$.

Next, suppose, to obtain a contradiction, that a solution to (MAX), $\{\tilde{C}_t, \tilde{L}_t, \tilde{K}_{t+1}, \tilde{x}_t\}_{t=0}^{\infty}$, can be supported as a SPE with replacement of the initial politician. Consider an alternative allocation $\{\tilde{C}_t', \tilde{L}_t', \tilde{K}_{t+1}', \tilde{x}_t'\}_{t=0}^{\infty}$ such that the initial politician is kept in power along the equilibrium path and receives exactly the same consumption sequence as the new politicians would have received after replacement. Since $\{\tilde{C}_t, \tilde{L}_t, \tilde{K}_{t+1}, \tilde{x}_t\}_{t=0}^{\infty}$ satisfies (8.14) for the new politicians at all t, $\{\tilde{C}_t', \tilde{L}_t', \tilde{K}_{t+1}', \tilde{x}_t'\}_{t=0}^{\infty}$ satisfies (8.14) for all t for the initial politician. Moreover, since $\{\tilde{C}_t, \tilde{L}_t, \tilde{K}_{t+1}, \tilde{x}_t\}_{t=0}^{\infty}$ must involve at least some positive consumption for the new politicians, $\{\tilde{C}_t', \tilde{L}_t', \tilde{K}_{t+1}', \tilde{x}_t'\}_{t=0}^{\infty}$ yields a higher $t = 0$ utility to the initial politician. Thus, x_0 can be reduced and C_0 can be increased without violating (8.14), so $\{\tilde{C}_t, \tilde{L}_t, \tilde{K}_{t+1}, \tilde{x}_t\}_{t=0}^{\infty}$ cannot be a solution to (MAX). This yields a contradiction and proves that there is no replacement of the initial politician along the equilibrium path.

We next show that citizens' strategy (in particular, $\rho\left(h^t\right) = 1$ if $h^t \neq \hat{h}^t$ and $\rho\left(h^t\right) = 0$ if $h^t = \hat{h}^t$) is subgame perfect and the equilibrium characterized here is renegotiation-proof. Let us denote the equilibrium value of the initial politician starting with capital stock K by $w_0(K)$ and the maximum feasible value that can be promised to a politician when the capital stock is K by $\bar{w}(K)$. Consider the following continuation equilibrium. If $\rho(h^t) = 1$ and $h^t \neq \hat{h}^t$, then the continuation equilibrium is a solution to (MAX), with initial value for the next politician $w' = w_0\left(K\left(h^t\right)\right)$, where $K\left(h^t\right)$ is the capital stock after history h^t (that is, after the deviation if there is any). If $\rho(h^t) = 1$ and $h^t = \hat{h}^t$, then the continuation equilibrium is a solution to (MAX), with initial value for the next politician given by $w' = \bar{w}\left(K\left(h^t\right)\right) \geq w_0\left(K\left(h^t\right)\right)$. Consequently, $\rho(h^t) = 0$ following $h^t = \hat{h}^t$ and $\rho\left(h^t\right) = 1$ following $h^t \neq \hat{h}^t$ are best responses for the citizens and are subgame perfect. Moreover, they involve the continuation play of a best SPE, thus the citizens and the politician in power cannot both be made better off. This establishes that the best sustainable mechanism outlined above, which achieves the solution to (MAX), can be supported as a renegotiation-proof SPE. □

This proposition enables us to focus on the constrained maximization problem given in (MAX). Moreover, it implies that in the best SPE, the initial politician will be kept in power forever (and that this best SPE is renegotiation-proof). The initial politician is kept in power forever because all politicians are identical and more effective incentives can be provided to a politician when he has a longer planning horizon (i.e., when he expects to remain in power for longer). Naturally, he is only kept in power along the equilibrium path—if he deviates from the implicitly-agreed mechanism, he will be replaced.

For future reference, let us define an *undistorted* allocation as a sequence $\{C_t, L_t, K_{t+1}, x_t\}_{t=0}^{\infty}$ that maximizes (8.12) without the sustainability constraint (8.14) (for a given sequence of $\{x_t\}_{t=0}^{\infty}$). An undistorted allocation where $(C_t, L_t) \in \text{Int}\Lambda$ satisfies

$$F_L(K_t, L_t)\, U_C(C_t, L_t) = -U_L(C_t, L_t), \tag{8.15}$$

$$U_C(C_t, L_t) = \beta F_K(K_{t+1}, L_{t+1})\, U_C(C_{t+1}, L_{t+1}). \tag{8.16}$$

We say that an allocation $\{C_t, L_t, K_{t+1}, x_t\}_{t=0}^{\infty}$ features *downward labor distortions* at time t if the left-hand side of (8.15) is strictly greater than the right-hand side. Similarly, there are *downward intertemporal distortions* when the left-hand side of (8.16) is strictly less than the right-hand side. Downward distortions imply that there is less labor supply and less capital accumulation than in an undistorted allocation. We will interpret these distortions as corresponding to "aggregate tax distortions," since allocations that involve downward labor and intertemporal distortions can be decentralized by using linear labor and capital taxes.

8.4.1. The Best Sustainable Mechanism without Capital. Let us start with the economy without capital, so that instead of Assumption 8.2, we have $Y_t = L_t$. An allocation can now be represented by $\{C_t, L_t, x_t\}$ (thus dropping K_t). An undistorted allocation with $(C_t, L_t) \in \text{Int}\Lambda$ now satisfies $U_C(C_t, L_t) = -U_L(C_t, L_t)$.

We next introduce a sustainability assumption, which ensures that when the maximum amount of utility is given to the politician in every period, this is sufficient to satisfy the sustainability constraint (8.14). More formally:

ASSUMPTION 8.4. ***(sustainability)*** Let $(\tilde{C}, \tilde{L}) \in \arg\max_{(C,L)\in\Lambda} \{L - C\}$. Then $v(\tilde{L} - \tilde{C})/(1-\delta) > v(\tilde{L})$.

The main result of this section is given in the following theorem.

THEOREM 8.1. *Suppose that $Y_t = L_t$, that Assumptions 8.1, 8.3 and 8.4 hold and that $U_C(0,0) > U_L(0,0)$. Then in the best SPE (best sustainable mechanism), we have:*

1. *there are downward labor distortions at $t = 0$.*

2. when $\beta \leq \delta$, the values promised to the politician $\{w_t\}_{t=0}^{\infty}$ form a nondecreasing sequence and converge to some w^*. Moreover, $\{C_t, L_t, x_t\}_{t=0}^{\infty}$ converges to some (C^*, L^*, x^*), which satisfies the no-distortion condition $U_C(C^*, L^*) = -U_L(C^*, L^*)$.

3. when $\beta > \delta$, then there are downward labor distortions even asymptotically.

The allocation described above can be supported as a renegotiation-proof SPE.

Part 1 of the theorem illustrates the additional distortion arising from the sustainability constraints. As output increases, the sustainability constraint, (8.14), requires more rents to be given to the politician in power and this increases the effective cost of production for the citizens. The best SPE creates distortions so as to reduce the level of output and thus the rents that have to be paid to the politician. Starting from an undistorted allocation, this is always beneficial. Loosely speaking, a marginal distortion, reducing labor supply and output by a small amount, creates a "second-order" loss for the citizens, but a "first-order" reduction in the amount of rents that have to be paid to the politician and thus a "first-order" increase in their consumption and utility.

Part 2 states that as long as $\beta \leq \delta$, the economy asymptotically converges to an equilibrium (C^*, L^*, x^*) where there are no aggregate distortions; even though there will be rents provided to the politician, these will be financed without introducing distortions. This result is important as it implies that in the long run there will be "efficient" provision of rents to politicians, with the necessary tax revenues raised without distortions (e.g., with lump-sum taxes in a decentralized allocation). This part of the theorem also shows that the (promised) rewards to the politician, given by the sequence $\{w_t\}_{t=0}^{\infty}$, are nondecreasing. Intuitively, current incentives to the politician are provided by both consumption in the current period, x_t, and by consumption in the future represented by the promised value, w_{t+1}. Future consumption by the politician not only relaxes the sustainability constraint in the future but does so in all prior periods as well. Thus, all else equal, optimal incentives for the politician should be backloaded. As discussed in the Introduction, this intuition for backloading in this political environment is the same as the intuition for backloading in the principal-agent literature.

Part 3 of the theorem states that if the politicians are less patient than the citizens, distortions will never disappear. Since in many realistic political economy models politicians are—or act—more short-sighted than the citizens, this part of the theorem implies that in a number of important cases, political economy considerations will lead to additional distortions that will not disappear even asymptotically. Finally, Theorem 8.1 also shows that the best SPE can be supported as a renegotiation-proof equilibrium.

To provide an intuition for the proof of the theorem, let us represent the maximization problem in (MAX) recursively (for the special case without capital and ignoring the feasibility constraint on w^+):

$$V(w) = \max_{(C,L)\in\Lambda,x,w^+} \left\{U(C,L) + \beta V(w^+)\right\} \tag{8.17}$$

subject to

$$C + x \leq L, \tag{8.18}$$

$$w = v(x) + \delta w^+, \tag{8.19}$$

$$v(x) + \delta w^+ \geq v(L). \tag{8.20}$$

Here $V(w)$ is the value of the citizens when they have promised value w to the politician and w^+ denotes next period's promised value. Constraint (8.18) imposes the resource constraint (8.13). Constraint (8.19) imposes promise keeping, incorporating the fact that the politician will not be replaced. It requires that the promised value w be equal to the sum of the current utility, $v(x)$, and the continuation utility, δw^+. Finally, constraint (8.20) is the recursive version of the sustainability constraint, (8.14). Let γ and $\psi \geq 0$ be the Lagrange multipliers on the constraints (8.19) and (8.20) respectively. We can now show that $V(w)$ is concave and differentiable. Furthermore, for the intuitive argument here, suppose that $(C,L) \in$IntΛ. The first-order condition with respect to w^+ and the envelope theorem then imply

$$\frac{\beta}{\delta}V'(w^+) = -\gamma - \psi = V'(w) - \psi. \tag{8.21}$$

Combining the first-order conditions for C and L gives

$$U_C(C,L) + U_L(C,L) = \psi v'(L). \tag{8.22}$$

Equation (8.22) makes it clear that aggregate distortions are related to the Lagrange multiplier on the sustainability constraint, ψ. Moreover, we must have $\psi > 0$ at $t = 0$, otherwise the politician would receive $w_0 = 0$ initially, which together with (8.20) would imply $C_t = L_t = 0$ for all t. However, $C_t = L_t = 0$ for all t cannot be a solution when $\psi = 0$ at $t = 0$. Equation (8.22) then yields $U_C(C,L) + U_L(C,L) > 0$ at $t = 0$.

To obtain the intuition for the second part of Theorem 8.1, consider the case where $\beta = \delta$ (for the argument for $\beta < \delta$ is somewhat different and committed from the notes). Then equation (8.21) implies

$$V'(w^+) = V'(w) - \psi \leq V'(w). \tag{8.23}$$

Concavity of the value function $V(\cdot)$ then implies that $w^+ \geq w$, with $w^+ > w$ if $\psi > 0$, and $w^+ = w$ if $\psi = 0$. Therefore, the values promised to the politician form a nondecreasing sequence and converge to some w^* and (8.23) implies that ψ must converge to 0. This also implies that $\{C_t, L_t, x_t\}_{t=0}^{\infty}$ converges to some (C^*, L^*, x^*), which satisfies (8.20) as stated in part 2 of Theorem 8.1.

This argument breaks down in part 3 of the theorem when $\delta < \beta$ because the politician does not value future rewards sufficiently and the sequence $\{w_t\}_{t=0}^{\infty}$ is not necessarily nondecreasing. In fact, (8.21) implies that if $\{w_t\}_{t=0}^{\infty}$ converges to some $\hat{w}$, then $\beta V'(\hat{w})/\delta = V'(\hat{w}) - \psi$. Since V' is negative, ψ must be strictly positive in this case and there will necessarily be asymptotic distortions.

8.4.2. The Best Sustainable Mechanism with Capital. We now extend Theorem 8.1 to an environment with capital, where the production function is given by Assumption 8.2. We first strengthen the sustainability assumption, Assumption 8.4. Let us define $\bar{C}$ and $\bar{K}$ such that

$$\bar{C} = \min\left\{C : \left(C, \bar{L}\right) \in \Lambda\right\} \text{ and } \bar{K} = \arg\max_{K \geq 0}\left\{F\left(K, \bar{L}\right) - K - \bar{C}\right\}. \tag{8.24}$$

Clearly, $\bar{C}$ is uniquely defined (since $C \geq 0$ and Λ is closed). In view of this and Assumption 2, $\bar{K}$ is also uniquely defined.

Assumption 4′ (sustainability with capital)
(1) $\delta v\left(F\left(\bar{K}, \bar{L}\right) - \bar{C} - \bar{K}\right)/(1-\delta) > v\left(F\left(\bar{K}, \bar{L}\right)\right)$; and (2) $\bar{C} + \bar{K} \leq F\left(0, \bar{L}\right)$.

The first part of Assumption 4′ states that there exists a feasible allocation delivering sufficient utility to the politician so that the sustainability constraint (8.14) can be satisfied as a strict inequality. This implies that the maximum utility to the politician can be provided without distortions. A high discount factor δ is sufficient to ensure that this part of the assumption is satisfied. The second part of the assumption is a technical condition, which guarantees that the equilibrium allocation does not get stuck at some arbitrary capital level, and naturally requires that $F\left(0, \bar{L}\right) > 0$. Both parts of this assumption are used only in part 2 of the next theorem to characterize the equilibrium when the utility provided to a politician reaches the boundary of the set of feasible values.

Theorem 8.2. *Suppose that Assumptions 8.1-8.3 and 4′ hold. Then in the best SPE:*

1. there are downward labor distortions at some $t < \infty$ and downward intertemporal distortions at $t-1$ (provided that $t \geq 1$);

2. when $\beta \leq \delta$, the best sustainable mechanism $\{C_t, K_{t+1}, L_t, x_t\}_{t=0}^{\infty}$ converges to some (C^*, K^*, L^*, x^*). At this allocation, the labor and intertemporal distortions disappear asymptotically, i.e., (8.15) and (8.16) hold as $t \to \infty$;

3. when $\beta > \delta$, there are downward labor and intertemporal distortions, even asymptotically.

The allocation described above can be supported as a renegotiation-proof SPE..

This theorem generalizes the results of Theorem 8.1 to an environment that is identical to the standard neoclassical growth model. The results are slightly weaker than in Theorem 8.1. In particular, there may not necessarily be distortions at the initial date, though such distortions will necessarily exist at some date. Perhaps more importantly, expected rewards to the politician are no longer always increasing. In fact, it is straightforward to construct examples in which the initial capital stock is sufficiently high so that these rewards are decreasing in the best SPE. Also noteworthy is that when $\beta > \delta$, the best SPE not only generates labor distortions but also intertemporal distortions. These can be thought of as "aggregate capital taxes," since they create a wedge between the marginal product of capital and the ratio of marginal utilities of consumption. Therefore, this model generates a political economy rationale for long-run capital taxation.

8.4.3. Stationary Equilibria. Let us not return to stationary equilibria as the typical applications of the Barro-Ferejohn approach assume. Let us focus on best *stationary* SPE and ignore capital. With stationary strategies and no capital, x_t has to be constant (conditional on the politician remaining in power). Although a similar result can be stated for the economy with capital, in this case the politician's consumption x would be a function of K, which complicates the analysis. The economy without capital allows us to emphasize the importance of non-stationary SPEs in a clearer fashion.

Proposition 8.2. *Consider the environment without capital in Theorem 8.1 and suppose that Assumptions 8.1, 8.3 and 8.4 hold and that $U_C(0,0) > U_L(0,0)$. Then, in the best stationary SPE distortions never disappear.*

Proof. Along a stationary equilibrium path, $x_t = x$ and $L_t = L$ so that

$$\frac{v(x)}{1-\delta} \geq v(L) \tag{8.25}$$

replaces the sustainability constraint (8.14). Constraint (8.25) must bind in all periods with $\psi > 0$, since otherwise the solution to the stationary equivalent of (MAX) would involve $x = 0$ and no distortions. The assumption that $U_C(0,0) > U_L(0,0)$ then implies that in this case $L > 0$, thus $x = 0$ would violate (8.25). Condition (8.22), which still applies in this case, then shows that there is a positive distortion on labor in all periods. □

This proposition illustrates the role of nonstationary SPE in our analysis. Stationary equilibria do not allow the optimal provision of dynamic incentives to politicians and imply that political economy distortions never disappear, even when $\beta \leq \delta$.

8.5. References

(1) Acemoglu, Daron, Michael Golosov and Aleh Tsyvinski (2008) "Political Economy of Mechanisms" *Econometrica*, 76: 619-641.

(2) Barro, Robert (1973) "The Control of Politicians: An Economic Model," *Public Choice*, 14, 19-42.

(3) Chari, V.V. and Patrick Kehoe (1990) "Sustainable Plans," *Journal of Political Economy*, 94, 783-802.

(4) Ferejohn, John (1986) "Incumbent Performance and Electoral Control," *Public Choice*, 50, 5-25.

(5) Harris, Milton and Bengt Holsmtrom (1982) "A Theory of Wage Dynamics", *Review of Economic Studies*, 49(3), 315-333.

(6) Lazear, Edward (1981) "Agency, Earnings Profiles, Productivity, and Hours Restrictions," *American Economic Review*, 71, 606-20.

(7) Persson, Torsten and Guido Tabellini (2000) *Political Economics: Explaining Economic Policy*, MIT Press.

(8) Ray, Debraj (2002) "Time Structure of Self-Enforcing Agreements," *Econometrica*, 70, 547-82.

(9) Thomas, Jonathan and Tim Worrall (1990) "Income Fluctuations and Asymmetric Information: An Example of Repeated Principle-Agent Problem," *Journal of Economic Theory* 51, 367-90.

CHAPTER 9

Failures of Electoral Control

The models so far emphasize aggregation of preferences via voting and how quotations can be controlled by electoral processes. In many cases, especially in many less-developed economies, electoral controls seem to be weak or nonexistent. In this chapter, we briefly discuss how money may matter in politics. In the next chapter, we discuss how in the presence of weak institutions, electoral control may totally disappear and politicians may become the "principals" rather than the "agents".

9.1. Lobbying

Consider next a very different model of policy determination, a lobbying model. In a lobbying model, different groups make campaign contributions or pay money to politicians in order to induce them to adopt a policy that they prefer. With lobbying, political power comes not only from voting, but also from a variety of other sources, including whether various groups are organized, how much resources they have available, and their marginal willingness to pay for changes in different policies. Nevertheless, the most important result for us will be that equilibrium policies under lobbying will also look like the solution to a weighted utilitarian social welfare maximization problem.

To see this, I will quickly review the lobbying model due to Grossman and Helpman (1996). Imagine again that there are G groups of agents, with the same economic preferences. The utility of an agent in group g, when the policy that is implemented is given by the vector $p \in \mathcal{P} \subset \mathbb{R}^K$, is equal to

$$U^g(p) - \gamma^g(p)$$

where $U^g(p)$ is the usual indirect utility function, and $\gamma^g(p)$ is the per-person lobbying contribution from group g. We will allow these contributions to be a function of the policy implemented by the politician, and to emphasize this, it is written with p as an explicit argument.

Following Grossman and Helpman, let us assume that there is a politician in power, and he has a utility function of the form

$$V(p) \equiv \sum_{g=1}^{G} \lambda^g \gamma^g(p) + a \sum_{g=1}^{G} \lambda^g U^g(p), \tag{9.1}$$

where as before λ^g is the share of group g in the population. The first term in (9.1) is the monetary receipts of the politician, and the second term is utilitarian aggregate welfare. Therefore, the parameter a determines how much the politician cares about aggregate welfare. When $a = 0$, he only cares about money, and when $a \to \infty$, he acts as a utilitarian social planner. One reason why politicians might care about aggregate welfare is because of electoral politics (for example, they may receive rents or utility from being in power as in the last subsection and their vote share might depend on the welfare of each group).

Now consider the problem of an individual j in group g. By contributing some money, he might be able to sway the politician to adopt a policy more favorable to his group. But he is one of many members in his group, and there is a natural free-rider problem. He might let others make the contribution, and simply enjoy the benefits. This will typically be an outcome if groups are unorganized (for example, there is no effective organization coordinating their lobbying activity and excluding non-contributing members from some of the benefits). On the other hand, organized groups might be able to collect contributions from their members in order to maximize group welfare.

Suppose that out of the G groups of agents, $G' < G$ are organized as lobbies, and can collect money among their members in order to further the interests of the group. The remaining $G - G'$ are unorganized, and will make no contributions. Without loss of any generality, let us rank the groups such that groups $g = 1, ..., G'$ to be the organized ones.

The lobbying game takes the following form: every organized lobby g simultaneously offers a schedule $\gamma^g(p) \geq 0$ which denotes the payments they would make to the politician when policy $p \in \mathcal{P}$ is adopted. After observing the schedules, the politician chooses p. Notice the important assumption here that contributions to politicians (campaign contributions or bribes) can be conditioned on the actual policy that's implemented by the politicians. This assumption may be a good approximation to reality in some situations, but in others, lobbies might simply have to make up-front contributions and hope that these help the parties that are expected to implement policies favorable to them get elected.

This is a potentially complex game, since various different agents (here lobbies) are choosing functions (rather than real numbers or vectors). Nevertheless, the equilibrium of this lobbying game takes a relatively simple form.

Theorem 9.1. ***(Lobbying Equilibrium)*** *In the lobbying game described above, contribution functions for groups $g = 1, 2...J$, $\{\hat{\gamma}^g(\cdot)\}_{g=1,2..J}$ and policy p^* constitute a subgame perfect Nash equilibrium if:*

(1) *$\hat{\gamma}^g(\cdot)$ is feasible in the sense that $0 \leq \hat{\gamma}^g(p) \leq U^g(p)$.*

(2) *The politician chooses the policy that maximizes its welfare, that is,*

$$p^* \in \arg\max_p \left(\sum_{g=1}^{G'} \lambda^g \hat{\gamma}^g(p) + a \sum_{g=1}^{G} \lambda^g U^g(p) \right).$$

(3) *There are no profitable deviations for any lobby, $g = 1, 2, .., G'$, that is,*

(9.2)

$$p^* \in \arg\max_p \left(\lambda^g \left(U^g(p) - \hat{\gamma}^g(p) \right) + \sum_{g'=1}^{G'} \lambda^{g'} \hat{\gamma}^{g'}(p) + a \sum_{g'=1}^{G} \lambda^{g'} U^{g'}(p) \right) \text{ for all } g = 1, 2, .., G'.$$

(4) *There exists a policy p^g for every lobby $g = 1, 2, .., G'$ such that*

$$p^g \in \arg\max_p \left(\sum_{g'=1}^{G'} \lambda^{g'} \hat{\gamma}^{g'}(p) + a \sum_{g'=1}^{G} \lambda^{g'} U^{g'}(p) \right)$$

and satisfies $\hat{\gamma}^g(p^g) = 0$. That is, the contribution function of each lobby is such that there exists a policy that makes no contributions to the politician, and gives her the same utility.

Proof. **(Sketch)** Conditions 1, 2 and 3 are easy to understand. No group would ever offer a contribution schedule that does not satisfy Condition 1. Condition 2 has to hold, since the politician chooses the policy. If Condition 3 did not hold, then the lobby could change its contribution schedule slightly and improve its welfare. In particular suppose that this condition does not hold for lobby $g = 1$, and instead of p^*, some $\hat{p}$ maximizes (9.2). Denote the difference in the values of (9.2) evaluated at these two vectors by $\Delta > 0$. Consider the following contribution schedule for lobby $g = 1$:

$$\tilde{\gamma}^1(p) = \lambda_1^{-1} \left[\sum_{g=1}^{G'} \lambda^g \hat{\gamma}^g(p^*) + a \sum_{g=1}^{G} \lambda^g U^g(p^*) - \sum_{g=2}^{G'} \lambda^g \hat{\gamma}^g(p) - a \sum_{g=1}^{G} \lambda^g U^g(p) + \varepsilon c^1(p) \right]$$

where $c^1(p)$ is an arbitrary function that reaches its maximum at $p = \hat{p}$. Following this contribution offer by lobby 1, the politician would choose $p = \hat{p}$ for any $\varepsilon > 0$. To see this note that by part (1), the politician would choose policy $\tilde{p}$ that maximizes

$$\lambda^1 \tilde{\gamma}^1(p) + \sum_{g=2}^{G'} \lambda^g \hat{\gamma}^g(p) + a \sum_{g=1}^{G} \lambda^g U^g(p) = \sum_{g=1}^{G'} \lambda^g \hat{\gamma}^g(p^*) + a \sum_{g=1}^{G} \lambda^g U^g(p^*) + \varepsilon c^1(p).$$

Since for any $\varepsilon > 0$ this expression is maximized by $\hat{p}$, the politician would choose $\hat{p}$. The change in the welfare of lobby 1 as a result of changing its strategy is $\Delta - \varepsilon c^1(\hat{p})$. Since $\Delta > 0$, for small enough ε, the lobby gains from this change, showing that the original allocation could not have been an equilibrium.

Finally, condition 4 ensures that the lobby is not making a payment to the politician above the minimum that is required. If this condition were not true, the lobby could reduce its contribution function by a constant, still induce the same behavior, and obtain a higher payoff. □

Next suppose that these contribution functions are differentiable. Then, it has to be the case that for every policy choice, p^k, within the vector p^*, we must have from the first-order condition of the politician that

$$\sum_{g=1}^{G'} \lambda^g \frac{\partial \hat{\gamma}^g(p^*)}{\partial p^k} + a \sum_{g=1}^{G} \lambda^g \frac{\partial U^g(p^*)}{\partial p^k} = 0 \text{ for all } k = 1, 2, .., K$$

and from the first-order condition of each lobby that

$$\lambda^g \left(\frac{\partial \hat{\gamma}^g(p^*)}{\partial p^k} - \frac{\partial U^g(p^*)}{\partial p^k} \right) + \sum_{g'=1}^{G} \lambda^g \frac{\partial \hat{\gamma}^{g'}(p^*)}{\partial p^k} + a \sum_{g'=1}^{G} \lambda^g \frac{\partial U^{g'}(p^*)}{\partial p^k} = 0 \text{ for all } k = 1, 2, .., K \text{ and } g = 1, 2, .., G'.$$

Combining these two first-order conditions, we obtain

$$\frac{\partial \hat{\gamma}^g(p^*)}{\partial p^k} = \frac{\partial U^g(p^*)}{\partial p^k} \tag{9.3}$$

for all $k = 1, 2, .., K$ and $g = 1, 2, .., G'$. Intuitively, at the margin each lobby is willing to pay for a change in policy exactly as much as this policy will bring them in terms of marginal return.

But then this implies that the equilibrium can be characterized as

$$p^* \in \arg\max_p \left(\sum_{j=1}^{G'} \lambda^g U^g(p) + a \sum_{j=1}^{G} \lambda^g U^g(p) \right).$$

Consequently, there is an interesting parallel between the lobbying equilibrium and the pure strategy equilibria of probablilistic voting models analyzed before. Like the latter, the lobbying equilibrium can also be represented as a solution to the maximization of a weighted social welfare function, with individuals in unorganized groups getting a weight of a and those in organized group receiving a weight of $1 + a$. Intuitively, $1/a$ measures how much money matters in politics, and the more money matters, the more weight groups that can lobby receive. As $a \to \infty$, we converge to the utilitarian social welfare function.

At this point, we can ask why contribution functions have to be differentiable. There is really nothing in the analysis that implies that these functions have to be differentiable;

in fact, equilibria with non-differentiable functions are easy to construct. Nevertheless, it is generally thought that equilibria with non-differentiable functions are more "fragile" and thus less relevant. See the discussion in Grossman and Helpman (1994) and Bernheim and Whinston (1986).

9.2. Application of Lobbying to Distributional Conflict

Next consider an application of this tax policy. Imagine there are two groups, the rich and poor. A fraction λ of the agents is rich with income h^r, and the remaining agents are poor with income $h^p < h^r$. Average income in the economy is

$$h = \lambda h^r + (1 - \lambda) h^p$$

There is a linear tax rate τ imposed on all incomes, and the proceeds are distributed lump sum to all agents. Taxation creates a dead weight loss of

$$c(\tau) h$$

where $c(\tau)$ is strictly increasing and convex, and assume that $c'(0) > 0$ and $c'(0) < \varepsilon$. The overall amount of lump sum subsidy is therefore

$$T = [\tau - c(\tau)] h$$

First recall that with majority voting, the preferred tax rate for the poor (who are more numerous) will result. In particular, we will have

$$\tau^m = \arg\max_{\tau} (1 - \tau) h^p + [\tau - c(\tau)] h$$

or

$$h - h^p = c'(\tau^m) h$$

So as long as $h - h^p > \varepsilon$, we will have $\tau^m > 0$, and there will be redistributive taxation.

Next consider the lobbying model, and assume that only the rich are organized. Then, the equilibrium tax rate will be given by

$$\tau^l = \arg\max_{\tau} \begin{pmatrix} (1 + a) \lambda [(1 - \tau) h^r + [\tau - c(\tau)] h] \\ +a (1 - \lambda) [(1 - \tau) h^p + [\tau - c(\tau)] h] \end{pmatrix}$$

The first-order condition to this problem (taking into account the possibility of a corner solution) gives

$$\begin{aligned} (1 + a) \lambda \left[h - h^r - c'\left(\tau^l\right) h\right] + a (1 - \lambda) \left[h - h^p - c'\left(\tau^l\right) h\right] &\leq 0 \\ &\text{or} \\ \lambda \left[h - h^r - c'\left(\tau^l\right) h\right] - a c'\left(\tau^l\right) h &\leq 0, \end{aligned} \tag{9.4}$$

and clearly since $h - h^r < 0$, we will have $\tau^l = 0$. Therefore, with lobbying there will be no redistributive taxation.

Interestingly, this conclusion also extends to the case in which the poor are also organized. In this case, we will have

$$\tau^{l'} = \arg\max_{\tau} \begin{pmatrix} \lambda\left[(1-\tau)h^r + [\tau - c(\tau)]h\right] \\ +(1-\lambda)\left[(1-\tau)h^p + [\tau - c(\tau)]h\right] \end{pmatrix}$$

The first-order condition to this problem gives

$$\lambda\left[h - h^r - c'\left(\tau^l\right)h\right] + (1-\lambda)\left[h - h^p - c'\left(\tau^l\right)h\right] \leq 0 \text{ and } \tau^l \geq 0$$

with complementary slackness. Since $\lambda[h - h^r] + (1-\lambda)[h - h^p] = 0$ by definition, this implies

$$-c'\left(\tau^l\right)h \geq 0,$$

which is only possible if $\tau^l = 0$.

This result basically reflects the fact that with costly taxation, the utilitarian social welfare maximizing policy is zero taxes.

In contrast, imagine a situation in which redistribution is socially beneficial. This might be because taxes are redistributed to the poor agents who are then able to invest in human capital which they were unable to do before because of credit constraints. Let us capture this in a very simple way by assuming that $c'(\tau) < 0$ for $\tau \leq \bar{\tau}$ and $c'(\bar{\tau}) = 0$. This implies that the utilitarian social welfare maximizing policy is to set $\tau = \bar{\tau}$.

In this case there would be equilibrium redistribution at the rate $\tau = \bar{\tau}$ when both the poor and the rich have organized to form lobbies. To see this note that with both the poor and the rich organized, the same condition as in (**??**) applies, so we need $c'\left(\tau^l\right)h = 0$, or in other words $\tau^l = \bar{\tau}$.

But with only the rich organized, the relevant condition is given by (9.4), or

$$\lambda\left[h - h^r - c'\left(\tau^l\right)h\right] - ac'\left(\tau^l\right)h = 0 \text{ or } \leq 0,$$

and if $|c'(\tau)|$ is not very large, there will not be redistribution.

This illustrates the fact that with the rich organized, public policy will cater more to their preferences, so policies that redistribute away from the rich to the poor will not be adopted even if they are socially beneficial.

9.3. Campaign Contributions

Now let us briefly discuss an alternative model of lobbying, where instead of bribes to politicians, lobbying expenditures take the form of campaign contributions and affect equilibrium election outcomes.

Consider a probabilistic voting model with campaign contributions. In particular, let contributions to party P be where

$$C_P = \sum\nolimits_g O^g \lambda^g C_P^g$$

O^g is an indicator variable for whether group g is organized or not, C_P^g is contribution per member, and α^g denotes the size of group g.

The effect of contributions is introduced as affecting the balance of different politicians. In particular, suppose as before that individuals in a group will vote for

$$U^i(p_A) - U^i(p_B) - \delta \geq \sigma^i,$$

where δ is an aggregate random variable affecting all voters. Assume that

$$\delta = \tilde{\delta} + \eta \times (C_B - C_A),$$

so campaign spending influences this balance parameter. The parameter η measures the effectiveness of campaign spending.

Now, the swing voter in group J becomes:

$$\sigma^g = U^g(q_A) - U^g(q_B) + \eta(C_A - C_B) - \tilde{\delta}.$$

In addition, assume that all groups are symmetric, and have σ^g distributed uniformly over

$$\left[-\frac{1}{2\phi}, \frac{1}{2\phi}\right].$$

The parameter $\tilde{\delta}$, on the other hand, has a uniform distribution on

$$\left[-\frac{1}{2\psi}, \frac{1}{2\psi}\right].$$

This implies that the probability of party A winning the election is

$$p_A = \frac{1}{2} + \psi \left[U(q_A) - U(q_B) + \eta(C_A - C_B)\right]$$

where $U(q_P) = \sum_g \lambda^g U^g(q_P)$ is a measure of average preferences. A utilitarian social planner would have simply maximized this. Moreover, given the symmetry of all the groups, we know from our above analysis that probabilistic voting would have also maximized this.

We continue to assume that the only objective of the parties is to come to power.

The question is how lobbying changes this. To understand this, let us look at the objective function of lobbies. Assume that the lobby for group J has the objective function:

$$p_A U^g(q_A) + (1 - p_A) U^g(q_B) - \frac{1}{2}\left(\left(C_A^g\right)^2 + (C_B^g)^2\right),$$

which means that they don't care about which party comes to power, only about the implemented policy. And there are convex costs of contributing to each party.

The exact timing of events is as follows:

- The two parties simultaneously choose their platforms, q_A and q_B;
- Lobbies, observing the platforms, decide how much to give to each party.
- Voters observe their own σ's and vote.

Important assumption here is that voters are essentially myopic, in the sense that they can be swayed by campaign contributions.

This implies the following first-order condition for campaign contributions (for all groups that are organized)

$$\eta\psi\lambda^g \left[U^g(q_A) - U^g(q_B)\right] - C_A^g \leq 0,$$

and

$$-\eta\psi\lambda^g \left[U^g(q_A) - U^g(q_B)\right] - C_B^g \leq 0,$$

which exploits the fact that $\partial p_A / \partial C_A^g = \eta\psi\lambda^g$ and taking into account that we may be at the corner solution. This yields

$$\begin{aligned} C_A^g &= \max\left\{0, \psi\eta\alpha^g(U^g(q_A) - U^g(q_B))\right\} \\ C_B^g &= -\min\left\{0, \psi\eta\alpha^g(U^g(q_A) - U^g(q_B))\right\}. \end{aligned} \tag{9.5}$$

In other words, despite the convexity of the contribution schedules, each lobby only contributes to one party, in particular to the party that has a platform that gives its members greater utility [can you see why the same result would have been even more immediate if the cost were $\frac{1}{2}\left(C_A^g + C_B^g\right)^2$? What would go wrong in this case?]

Now consider the first stage of the game where each party chooses their platform. Since parties only care about coming to power, party A will maximize:

$$\psi\left[U(q_A) - U(q_B) + \eta \times \sum_g \left(\begin{array}{c} \max\left[0, \psi\eta\alpha^g(U^g(q_A) - U^g(q_B))\right] + \\ \min\left[0, \psi\eta\alpha^g(U^g(q_A) - U^g(q_B))\right] \end{array}\right)\right]$$

while party B will try to minimize this object.

It is clear that this is a concave problem, so the parties will again adopt symmetric platforms. This has a very important implication: in equilibrium the lobbies will make no contribution from (9.5), but still influence policy with the threat of campaigning against the party that deviates from a particular equilibrium platform!

The first-order conditions will be:

$$\sum_g \alpha^g \left[\psi + O^g \alpha^g (\psi\eta)^2\right] \nabla U^g(q_A) = 0.$$

In other words, parties will again be maximizing a weighted utility function.

$$\sum_g \alpha^g \left[1 + O^g \alpha^g \psi\eta^2\right] U^g(q_A).$$

When no group is organized, i.e., $O^g = 0$ for all g, this is equivalent to the maximization of utilitarian social welfare (the assumption that $\phi^g = \phi$ this of course important for this). Otherwise, organized groups will get more weight, and interestingly larger groups will get more weight, because they can generate greater campaign contributions. The additional weight that organized groups receive will be a function of η, the effectiveness of lobbies.

CHAPTER 10

Politics in Weakly-Institutionalized Environments

10.1. Introduction

We will now discuss whether and why we may need somewhat different models to think of the dynamics of political power in "weakly institutionalized polities" such as those in Africa, where institutions do not provide ways of checking the power of rulers or expressing the views of citizens. The motivation for introducing this terminology is to emphasize that there are places in the world where political actions and incentives do not seem to be tightly conditioned by the formal rules of the political game and where politics seems qualitatively distinct from Western Europe or North America. Looking at the constitution is not going to give us much clue to the political order of such places. Some of these issues came up in our discussion of institutional persistence where we looked at how de facto and de jure power may offset each other implying that particular formal political institutions did not have unique implications for the equilibrium distribution of political power. In political science these issues come up in African politics where the main paradigm is what is called "Neo-Patrimonialism" or "Personal Rule." This is basically an extreme personalization of politics where it is not the formal rules of the game that matters, but personalities, cliques, families, social networks which may function completely outside the formal rules. Bratton and van der Walle (1997, p. 62) put it as follows:

> "the right to rule in neopatrimonial regimes is ascribed to a person rather than to an office, despite the official existence of a written constitution. One individual ... often a president for life, dominates the state apparatus and stands above its laws. Relationships of loyalty and dependence pervade a formal political and administrative system, and officials occupy bureaucratic positions less to perform public service ... than to acquire personal wealth and status. Although state functionaries receive an official salary, they also enjoy access to various forms of illicit rents, prebends, and petty corruption, which constitute ... an entitlement of office. The chief executive and his inner circle undermine the effectiveness of the nominally modern state administration by

> using it for systematic patronage and clientelist practices in order to maintain political order."

Jackson and Rosberg (1982, pp.17-19) note that personal rule is

> "a system of relations linking rulers ... with patrons, clients, supporters, and rivals, who constitute the 'system.' If personal rulers are restrained, it is by the limits of their personal authority and power and by the authority and power of patrons, associates, clients, supporters, and—of course—rivals. The systems is 'structured' ... not by institutions, but by the politicians themselves."

These insights are interesting but the essence of personal rule has not been formalized and how this relates more generally to the absence of institutionalization is unclear. Though we are very far from an understanding of "weakly institutionalized polities" we believe that this terminology is useful. This is an amazingly exciting area for future research and in these lectures we just present some ideas and two models about different aspects of things that may be relevant. We cannot hope to propose a theory of politics in such societies. An interesting research program is the systematic study of policymaking in weakly-institutionalized societies, and ultimately, a study of the process via which strongly-institutionalized societies emerge.

To frame and begin the discussion, let us focus on "kleptocratic" regimes, where a ruler and his close (often ethnic) associates managed to capture a large fraction of the GDP for their own benefit and consumption.

Many countries in Africa and the Caribbean suffer under such "kleptocratic" regimes (or personal rule). Examples of kleptocratic regimes include the Democratic Republic of the Congo (Zaire) under Mobutu Sese Seko, the Dominican Republic under Rafael Trujillo, Haiti under the Duvaliers, Nicaragua under the Somozas, Uganda under Idi Amin, Liberia under Charles Taylor, and the Philippines under Ferdinand Marcos. In all these cases, kleptocratic regimes appear to have been disastrous for economic performance and caused the impoverishment of the citizens.

Perhaps the most puzzling feature of kleptocracies, illustrated by the examples from the Congo, the Dominican Republic, Nicaragua or Haiti, is their longevity, despite the disastrous policies pursued by the rulers. This longevity is made even more paradoxical by the fact that such regimes apparently lacked a political base (a core constituency) that supported them. Despite the absence of formal institutional mechanisms for deposing unpopular rulers, constraints on the behavior of rulers exist even in weakly-institutionalized societies (e.g., the threat of revolution, or competition from other strongmen). Why do, then, the heavily-taxed producers or the poverty-stricken citizens not replace the kleptocrat? Why do they rarely form an effective opposition constraining the kleptocrat? How can a regime that apparently benefits nobody outside of the narrowest of cliques survive? The basic answer, which distinguishes weakly-institutionalized polities from those that are strongly-institutionalized, is the absence of strong institutions allows rulers to adopt political strategies which are highly effective at defusing any opposition to their regime.

The seminal book by Robert Bates, *Markets and States in Tropical Africa*, provides many clues towards an answer. Bates described the web of inefficient transfers and policies in effect in many parts of Africa, but most notably in Ghana and Zambia, and suggested the following logic: many of these inefficient policies are in place to transfer resources from the population to the ruling groups, while at the same time ensuring their political survival. In particular, the nexus of inefficient policies appeared to be useful for creating an environment where any group that became politically mobilized against the rulers could be punished, while those that remained loyal were rewarded. With this logic, the Ghanaian government heavily taxed cocoa producers, while at the same time subsidizing their inputs of seeds and fertilizers. The subsidies could be allocated selectively as a potential reward for not attempting to change the status quo. Similarly, the exchange rate was kept overvalued because then the government could allocate or withhold valuable rations of foreign exchange in order to guarantee support.

One formalization of these ideas is the use of a "divide-and-rule" strategy by the ruler. Divide-and-rule is a method used by kleptocrats to maintain power in weakly-institutionalized polities while simultaneously pursuing policies costly to society. The logic of the divide-and-rule strategy is to enable a ruler to bribe politically pivotal groups off the equilibrium path, ensuring that he can remain in power against challenges. To remove a ruler from power requires the cooperation of distinct social groups which is made difficult by the collective action problem. By providing selective incentives and punishments, the divide-and-rule strategy exploits the fragility of social cooperation: when faced with the threat of being ousted, the kleptocratic ruler intensifies the collective action problem and destroys the coalition against him by bribing the pivotal groups.

10.2. A Model of Divide-and-Rule

Basic idea: consider a dynamic game between the ruler and two producer groups. The kleptocratic ruler taxes production and uses the ensuing tax revenue, the rents from natural resources and potential foreign aid from outside donors for his own consumption. The two producer groups, if they can cooperate, can remove the ruler from power and establish democracy (a regime more favorable to their interests).

This cooperation is modeled as follows: one of the two groups (the "proposer") makes a proposal to remove the ruler from power, and if the other group (the "proposed") agrees, the ruler is removed and democracy is established. The ruler-friendly political institutions, however, imply that before the proposed group responds to the proposal, the ruler can make a counteroffer. This counteroffer enables him to use a divide-and-rule strategy: following a challenge, the ruler uses all his resources and the tax revenues to bribe the proposed group (and compensate them for future higher taxes if they turn down the proposal and keep the ruler in power). If he can do so successfully, he can fight off the challenge, and anticipating this, no group will challenge the ruler. Therefore, the divide-and-rule strategy remains off the equilibrium path, and its anticipation implies that the ruler can follow highly distortionary (kleptocratic) policies without being challenged. Not only is the kleptocrat able to stay in power, but the threat of divide-and-rule implies that there will be no challenges to remove him from power along the equilibrium path.

10.2.1. The Environment. Consider a small open economy (alternatively, an economy with linear technology) producing three goods: a natural resource, Z, and two goods, q_1 and q_2. We normalize the prices of all goods to 1, which is without loss of any generality, since we will allow differences in the technology of production of the two goods. To start with, we focus on the case where the production of the natural resource good Z_t is constant in all periods,

$$Z_t = Z.$$

Natural resources create rents in this economy, which, in turn, affect political equilibria. We assume that the natural resource rents accrue to the government, and can then be distributed to the producers or consumed by the ruler.

There are two (large) groups of agents, n_1 that produce q_1 and n_2 that produce q_2. We normalize $n_1 = n_2 = 1$. Both groups have utility at time t given by:

$$\sum_{s=t}^{\infty} \beta^s u_{is}\left(y_{is}, l_{is}\right) = \sum_{s=t}^{\infty} \beta^s \left(y_{is} - \frac{\eta}{1+\eta} l_{is}^{\frac{1+\eta}{\eta}} \right) \tag{10.1}$$

where $\beta < 1$ is the discount factor, y_{it} denotes their after-tax income, and l_{it} is labor supply at time t. This specification implies that labor is supplied with elasticity $\eta > 0$.

For each producer of group i, the production technology is:

$$q_{it} = \omega_i l_{it}, \tag{10.2}$$

where ω_i is the productivity of group $i = 1, 2$. Without loss of generality, we assume that group 1 is more productive, i.e., $\omega_1 \geq \omega_2$. To parametrize the degree of inequality between the two groups, we denote:

$$\omega_1 = \overline{\omega}(1+x) \text{ and } \omega_2 = \overline{\omega}(1-x), \tag{10.3}$$

where, by construction, $\overline{\omega}$ is the average productivity of the economy, and $x \in [0,1]$. A greater x corresponds to greater inequality between the two groups.

The only redistributive tools in the economy are a linear income tax that is potentially specific to each group, and group-specific lump-sum transfers. The option to use group-specific taxes and transfers are important for the results, and plausible in the context of African societies, where there are clear geographic and ethnic distinctions between producer groups. The post-tax income of the two groups are

$$y_{it} = (1-\tau_{it})\,\omega_i l_{it} + T_{it}, \tag{10.4}$$

where $\tau_{it} \in [0,1]$ is the income tax imposed on group i at time t and $T_{it} \in [0,\infty)$ is a (non-negative) lump-sum transfer to group i.

In each period, each producer maximizes his utility function (10.1) taking the tax rate τ_{it} as given. This implies a labor supply function:

$$l_{it}(\tau_{it}) = [(1-\tau_{it})\,\omega_i]^{\eta}. \tag{10.5}$$

This equation relates labor supply, and therefore output, to taxes, and illustrates the distortionary effects of taxation: greater taxes reduce labor supply and output.

Using (10.5), the instantaneous indirect utility of a representative agent in group i is found to be:

$$U_i(\tau_{it}, T_{it}) = \frac{1}{1+\eta}[\omega_i(1-\tau_{it})]^{1+\eta} + T_{it}, \tag{10.6}$$

and tax revenues are:

$$R(\tau_{1t}, \tau_{2t}) = \tau_{1t} q_{1t} + \tau_{2t} q_{2t} = \tau_{1t}\ (1-\tau_{1t})^{\eta}\ {\omega_1}^{1+\eta} + \tau_{2t}\ (1-\tau_{2t})^{\eta}\ {\omega_2}^{1+\eta}. \tag{10.7}$$

The government budget constraint is

$$T_{1t} + T_{2t} + C_{Kt} \leq R(\tau_{1t}, \tau_{2t}) + Z + F_t, \tag{10.8}$$

where $C_{Kt} \in [0,\infty)$ is the consumption of the (kleptocratic) ruler, $R(\tau_{1t}, \tau_{2t})$ is tax revenue given by (10.7), and $F_t \in [0,\infty)$ is foreign aid, if any. The ruler is assumed to have the utility function at time t:

$$\sum_{s=t}^{\infty} \beta_K^s C_{Ks},$$

where $\beta_K < 1$ is the discount factor of the ruler, which could differ from those of the citizens.

The political system is either "dictatorship" (controlled by the ruler), K, or democracy, D. Our focus is whether dictatorship can survive and to what extent it will be "kleptocratic" (i.e., to what extent the ruler will be able to tax producers for his own consumption, while ensuring the survival of the dictatorship). To focus on this question, we model democracy in the simplest possible way, and assume that in democracy the two producer groups are in power jointly, thus they set zero taxes, and share the natural resource rents and foreign aid equally (and therefore, set $C_{Kt} = 0$).

In what follows, we assume that only the ruler receives foreign aid (i.e., there is no foreign aid in democracy). More formally, denoting the political state by S_{t-1}, we have $F_t = 0$ if $S_{t-1} = D$ and $F_t = F$ if $S_{t-1} = K$. This assumption is not as extreme as it appears, since any part of foreign aid that is perpetual can be included in the natural resource rents, Z, and therefore, F_t can be interpreted as the additional portion of foreign aid that the ruler receives because under his rule there is more poverty or famine, or because the ruler pursues a foreign policy in line with the donors' interests. Moreover, if a democracy will receive more foreign aid than the ruler, we can allow this by letting $F < 0$. Finally, we can interpret F as the fungible part of foreign aid. In this case, even if foreign aid will continue after democracy it will be allocated to some specific purposes such as poverty reduction or education for disadvantaged groups. In this case in democracy the two producer groups will not have access to aid, and this situation is equivalent to the one here with $F_t = 0$ when $S_{t-1} = D$.

Given this assumption, the instantaneous utilities of the two groups in democracy are:

$$U_i^D = \frac{\omega_i^{1+\eta}}{1+\eta} + \frac{Z}{2}. \tag{10.9}$$

In contrast, in kleptocracy, the ruler will maximize his consumption, subject to the constraint that he keeps power (alternatively, he can be removed from power, and in this case, democracy will result, and $C_{Kt} = 0$ for all future periods). Before describing the constraints facing the ruler in detail, let us write the "unconstrained" solution. This is given by maximizing $R(\tau_{1t}, \tau_{2t})$, which is achieved at the tax rates:

$$\tau_{1t}^* = \tau_{2t}^* = \tau^* \equiv \frac{1}{1+\eta}, \tag{10.10}$$

and paying 0 transfers, i.e., $T_{it}^* = 0$, thus setting $C_{Kt} = R(\tau^*, \tau^*) + Z + F$. The instantaneous utilities of the two groups under these tax rates are given by:

$$U_i^* = U_i\left(\tau_{it} = \tau^*, T_{it} = 0\right) = \frac{\omega_i^{1+\eta}}{1+\eta}\left(\frac{\eta}{1+\eta}\right)^{1+\eta}. \tag{10.11}$$

10.2.2. The Political Game and Definition of Equilibrium. The timing of events in the political game are as follows. In each period, t, society inherits a political state, either $S_{t-1} = D$ or $S_{t-1} = K$. $S_{t-1} = D$ is an absorbing state, so if the economy ever becomes a democracy, it remains so forever. If $S_{t-1} = D$, the two producer groups play the simple game described above, denoted by $\Gamma_t(D)$ for convenience, where they set the taxes and share the natural resources rents equally. If society is a dictatorship, i.e., $S_{t-1} = K$, then the following game, $\Gamma_t(K)$, is played:

- The ruler announces tax rates (τ_{1t}, τ_{2t}) and transfers (T_{1t}, T_{2t}).
- Each group i decides whether to make a proposal to remove the ruler from power. $p_{it} = 1$ denotes that group i has made a proposal and $p_{it} = 0$ denotes otherwise (if both groups simultaneously choose to make a proposal, one of them is chosen randomly to have $p_{it} = 1$ and the other one has $p_{it} = 0$). If one of the two groups makes a proposal to replace the ruler, we denote this by $P_t = 1$, with $P_t = 0$ otherwise.
- If $P_t = 0$, then $(\tau_{1t}, \tau_{2t}, T_{1t}, T_{2t})$ is implemented and the political system remains at $S_t = K$.
- If $P_t = 1$, i.e., if $p_{jt} = 1$ for one of the groups, then: the ruler makes a new offer of taxes and transfers, $(\tau_{1t}^r, \tau_{2t}^r, T_{1t}^r, T_{2t}^r)$ such that this policy vector satisfies the government budget constraint, (10.8). Group $i \neq j$ then responds to the proposal of "proposer" group j and the ruler's new policy vector. If the "proposed", group i, chooses $d_{it} = 1$, the ruler is replaced and there is a switch to democracy, i.e., $S_t = D$. If $d_{it} = 0$, the political system remains at $S_t = K$, and $(\tau_{1t}^r, \tau_{2t}^r, T_{1t}^r, T_{2t}^r)$ is implemented.
- Given the policy vector, either $(\tau_{1t}, \tau_{2t}, T_{1t}, T_{2t})$ or $(\tau_{1t}^r, \tau_{2t}^r, T_{1t}^r, T_{2t}^r)$, individuals in both groups choose labor supply.
- Output is produced, tax revenues are collected and consumption takes place.
- If $P_t = 1$ and the proposed group plays $d_{it} = 1$, then in the next period the stage game switches to $\Gamma_{t+1}(D)$, and otherwise it is $\Gamma_{t+1}(K)$.

There are a number of noteworthy features: first, we assume that all individuals within a producer group act in cohesion in the political game. This is a natural assumption here, since there are no costs of political action, and all agents within a group have the same preferences, so there is no free-rider problem. Second, there is a specific (political) structure built in the timing of the political game: the ruler can only be replaced if the two groups agree to replace him. This assumption captures the fact that in weakly-institutionalized societies, those controlling the state may have considerable power, and cannot be easily removed from office by one of the social groups alone. An alternative political game, where the party in power needs to receive support from all social groups or compete against potential rivals, would correspond to "political institutions" placing checks on politicians. However, such strong political institutions are absent in a number of less developed countries. In these weakly-institutionalized polities, the implied power of the ruler, combined with the fact that after the proposal to remove him from power he can offer a different policy vector, gives him the opportunity to use a "divide-and-rule" strategy, which will be the focus of our analysis.

To simplify the analysis, we will focus on the (pure strategy) Markov Perfect Equilibria (MPE) of the above game (though this restriction is not important for the results). An MPE is a mapping from the current state of the game (and from the actions taken previously in the same stage game) to strategies. Here, the only state variable is S_{t-1}, which denotes whether the political state is either democracy or dictatorship.

10.2.3. Analysis. The MPE will be characterized by backward induction. When $S_{t-1} = D$, there are no interesting actions, and the ruler receives zero utility, while the two groups receive lifetime utilities of:

$$V_i^D = \frac{U_i^D}{1-\beta}, \tag{10.12}$$

with U_i^D given by (10.9). Note also that V_i^D given by (10.12) is what the proposed group will receive if it chooses $d_{it} = 1$ and removes the ruler from power.

On the other hand, if, in response to the reaction of the ruler (τ_{it}^r, T_{it}^r), the proposed group chooses $d_{it} = 0$, its members will receive

$$V_i^C\left(\tau_{it}^r, T_{it}^r \mid \tau_i^e, T_i^e\right) = U_i\left(\tau_{it}^r, T_{it}^r\right) + \frac{\beta U_i(\tau_i^e, T_i^e)}{1-\beta}, \tag{10.13}$$

where U_i is given by (10.6) and (τ_i^e, T_i^e) is the MPE tax transfer combination that applies to this group. The reasoning for this expression is that in this period, the proposed group receives (τ_{it}^r, T_{it}^r), and the kleptocrat remains in power, so in the future, the play goes back to the equilibrium policy of (τ_i^e, T_i^e).

In addition, the response of the ruler must satisfy the government budget constraint:

$$T_{1t}^r + T_{2t}^r \leq R(\tau_{1t}^r, \tau_{2t}^r) + Z + F. \tag{10.14}$$

The "divide-and-rule" strategy will be successful and the ruler will keep power only if

$$V_i^C\left(\tau_{it}^r, T_{it}^r \mid \tau_i^e, T_i^e\right) \geq V_i^D. \tag{10.15}$$

It is useful to distinguish two cases:

- The ruler will be able to maintain power, with the equilibrium strategy of $\tau_{it}^e = \tau^* \equiv 1/(1+\eta)$ as given by (10.10) and $T_{it}^e = 0$ for $i = 1, 2$ and for all t. We denote the set of parameters such that this happens by Σ^*, i.e., if $\sigma' = (\eta, \beta, Z, F, \bar{\omega}, x) \in \Sigma^*$, then $(\tau_{1t}^e, \tau_{2t}^e, T_{1t}^e, T_{2t}^e) = (\tau^*, \tau^*, 0, 0)$.

- The ruler will not be able to maintain power if he sets $(\tau^*, \tau^*, 0, 0)$, thus $\sigma' \notin \Sigma^*$. As we will see below, in this case, the ruler can reduce taxes and increase transfers so as to maintain power. We will also see that in this case $(\tau_1^e, \tau_2^e) < (\tau^*, \tau^*)$, that is, the ruler will necessarily be forced to reduce taxes, and policy will be less distortionary.

To characterize Σ^*, let us start with the subgame in which group j has proposed to replace the ruler, and denote the policies initially chosen by the ruler by $(\tau_{1t}^e, \tau_{2t}^e, T_{1t}^e, T_{2t}^e)$. If the ruler responds with (τ_{it}^r, T_{it}^r) for $i \neq j$ such that $V_i^C(\tau_{it}^r, T_{it}^r \mid \tau_i^e, T_i^e) < V_i^D$, then he will be replaced. This shows that the ruler must ensure (10.15).

To analyze how, and when, the ruler can do so, let us first define

$$V_i^C[\tau_i^e, T_i^e] = \max_{\tau_{1t}^r, \tau_{2t}^r, T_{1t}^r, T_{2t}^r} V_i^C(\tau_{it}^r, T_{it}^r \mid \tau_i^e, T_i^e) \tag{10.16}$$

subject to (10.14). If $V_i^C[\tau_i^e, T_i^e] < V_i^D$ for $i = 1$ or 2, then group $j \neq i$, anticipating that its proposal will be accepted, will propose to replace the ruler, and the ruler will be deposed. Therefore, the ruler must guarantee that $V_i^C[\tau_i^e, T_i^e] \geq V_i^D$ for $i = 1$ and 2 to remain in power.

Consequently, we first need to find $V_i^C[\tau_i^e, T_i^e]$, the maximum utility that the ruler can give to the proposed group off the equilibrium path. To do this, we need to maximize (10.16) subject (10.14). Straightforward differentiation establishes that

$$\tau_i^r = 0 \text{ and } \tau_j^r = \frac{1}{1+\eta}.$$

Therefore, in fighting off a challenge from group j, the ruler will set the revenue-maximizing tax rate on this group, and set zero taxes on the proposed group i. In addition, the ruler will clearly give the minimum possible amount to the proposer group, thus $T_j^r = 0$. Then the government budget constraint, (10.8), implies:

$$T_i^r = R(\tau_i^r, \tau_j^r) + Z + F.$$

Using these expressions, we can derive the maximum off-the-equilibrium-path payoff of the proposed group, as a function of the MPE policy vector $(\tau_{1t}^e, \tau_{2t}^e, T_{1t}^e, T_{2t}^e)$. This is

$$V_i^C[\tau_i^e, T_i^e] = \frac{\omega_i^{1+\eta}}{1+\eta} + \frac{\omega_j^{1+\eta}}{1+\eta}\left(\frac{\eta}{1+\eta}\right)^\eta + Z + F + \frac{\beta U_i(\tau_i^e, T_i^e)}{1-\beta}. \tag{10.17}$$

This expression is the maximum utility that the ruler can give to group i, following a proposal by group j, as a function of the equilibrium tax and transfer rates on group i.

Given this analysis, the problem of finding the MPE is equivalent to finding a solution to the following maximization problem of the ruler:

$$\max_{\tau_{1t}^e, \tau_{2t}^e, T_{1t}^e, T_{2t}^e} \frac{1}{1-\beta_K}[R(\tau_1^e, \tau_2^e) + Z + F] \tag{10.18}$$

subject the constraint set:

$$V_i^C\left[\tau_i^e, T_i^e\right] \geq V_i^D \text{ for } i = 1, 2. \tag{10.19}$$

We can now characterize the solution to this constrained maximization problem. First, notice that combining (10.11), (10.12) and (10.17), the constraint set, (10.19), can be rewritten as:

$$\begin{aligned} &\frac{\omega_i^{1+\eta}}{1+\eta} + \frac{\omega_j^{1+\eta}}{1+\eta}\left(\frac{\eta}{1+\eta}\right)^{\eta} + Z + F + \frac{\beta}{1-\beta}\left(\frac{1}{1+\eta}\left[\omega_i\left(1-\tau_i\right)\right]^{1+\eta} + T_i\right) \\ &\geq \frac{1}{1-\beta}\left(\frac{\omega_i^{1+\eta}}{1+\eta} + \frac{Z}{2}\right), \end{aligned} \tag{10.20}$$

for $i = 1, 2$ and $j \neq i$. Then exploiting the fact that $\omega_1 = \overline{\omega}(1+x)$ and $\omega_2 = \overline{\omega}(1-x)$, we can write the constraint set as

$$\Psi(\tau_1, T_1, x) \geq Z\left(\beta - \frac{1}{2}\right) - (1-\beta)\, F, \text{ and} \tag{10.21}$$

$$\Psi(\tau_2, T_2, -x) \geq Z\left(\beta - \frac{1}{2}\right) - (1-\beta)\, F, \tag{10.22}$$

where

$$\begin{aligned} &\Psi(\tau, T, x) \\ \equiv\; &\frac{(1-\beta)\,\overline{\omega}^{1+\eta}\,(1-x)^{1+\eta}}{1+\eta}\left(\frac{\eta}{1+\eta}\right)^{\eta} - \frac{\beta\,\overline{\omega}^{1+\eta}\,(1+x)^{1+\eta}}{1+\eta}\left(1-(1-\tau)^{1+\eta}\right) + \beta T. \end{aligned}$$

Moreover, in the case where there is no inequality between the two groups, i.e., when $x = 0$, the constraint set is simply:

$$\Psi(\tau_i, T_i) \equiv \Psi(\tau_i, T_i, x = 0) \geq Z\left(\beta - \frac{1}{2}\right) - (1-\beta)\, F. \tag{10.23}$$

It is already possible to see why the divide-and-rule strategy can arise in equilibrium. The constraint set, characterized by (10.21) and (10.22), will be satisfied when, off-the-equilibrium path, the ruler can shift enough resources to the proposed group. In other words, a very inefficient set of policies can be supported when each group knows that if it proposes to replace the ruler, the ruler will bribe the other group successfully and remain in power. Recognizing this off-the-equilibrium path threat, no group will challenge the ruler, who will then be able to pursue kleptocratic policies along the equilibrium path.

10.2.4. Equilibrium Without Inequality. Let us start with the case in which $x = 0$ and there is no inequality between the two groups. First, note that whenever he can, the ruler would like to set the tax rates that maximize (10.18), i.e., $(\tau_{1t}^e, \tau_{2t}^e, T_{1t}^e, T_{2t}^e) = (\tau^*, \tau^*, 0, 0)$. Therefore, the first step is to characterize the set of parameters Σ^* such that these best tax rates (from the point of view of the kleptocratic ruler) can be supported as equilibria. Using

(10.23) and substituting $(\tau_{1t}^e, \tau_{2t}^e, T_{1t}^e, T_{2t}^e) = (\tau^*, \tau^*, 0, 0)$, we immediately obtain the set Σ^* as the set of parameters such that $\Psi\left(\frac{1}{1+\eta}, 0\right) \geq Z\left(\beta - \frac{1}{2}\right) - (1-\beta) F$, or more explicitly:

(10.24)
$$\Sigma^* = \left\{\sigma = (\eta, \beta, Z, F, \overline{\omega}) : \frac{\overline{\omega}^{1+\eta}}{1+\eta}\left(\frac{\eta}{1+\eta}\right)^{\eta} - \frac{\beta\, \overline{\omega}^{1+\eta}}{1+\eta} \geq Z\left(\beta - \frac{1}{2}\right) - (1-\beta) F\right\}.$$

If $\sigma = (\eta, \beta, Z, F, \overline{\omega}) \in \Sigma^*$, then the MPE involves $(\tau_{1t}^e, \tau_{2t}^e, T_{1t}^e, T_{2t}^e) = (\tau^*, \tau^*, 0, 0)$. What happens if $\sigma \notin \Sigma^*$. Then, $(\tau_{1t}^e, \tau_{2t}^e, T_{1t}^e, T_{2t}^e) = (\tau^e, \tau^e, T^e, T^e)$ will be chosen such that $\Psi(\tau^e, T^e) = Z\left(\beta - \frac{1}{2}\right) - (1-\beta) F$ (given the symmetry between the two groups, the ruler will choose the same taxes and transfers for both groups). Moreover, inspection of the expression for $\Psi(\tau, T, x)$ establishes that as long as $\Psi(\hat{\tau}, T = 0) = Z\left(\beta - \frac{1}{2}\right) - (1-\beta) F$ for some $\hat{\tau} \in [0, \tau^*]$, the ruler will reduce taxes to $\hat{\tau}$ and sets 0 lump-sum transfers (this is intuitive, since taxes are distortionary). The important point to note is that the ruler can always satisfy (10.24), and therefore remain in power. This highlights the importance of the underlying political institutions in this context: by allowing the ruler to use divide-and-rule, the current set of political institutions make sure that he always remains in power. Nevertheless, the extent to which he can transfer rents to himself and distort the allocation of resources will depend on parameter values as we will see below.

This discussion establishes the following proposition:

Proposition 10.1. *Let Σ^* be given by (10.24). Then we have:*

- When $\sigma \in \Sigma^*$, then the unique MPE is an *unconstrained kleptocratic regime* where $(\tau_{1t}^e, \tau_{2t}^e, T_{1t}^e, T_{2t}^e) = (\tau^*, \tau^*, 0, 0)$ for all t and $i = 1, 2$.
- When $\sigma \notin \Sigma^*$, then the unique MPE is a *constrained kleptocratic regime* where the equilibrium policies are $(\tau_{1t}^e, \tau_{2t}^e, T_{1t}^e, T_{2t}^e) = (\hat{\tau}, \hat{\tau}, 0, 0)$ if

(10.25)
$$\Psi(\hat{\tau}, T = 0) = Z\left(\beta - \frac{1}{2}\right) - (1-\beta) F,$$

for some $\hat{\tau} \in [0, \tau^*]$, and $(\tau_{1t}^e, \tau_{2t}^e, T_{1t}^e, T_{2t}^e) = (0, 0, \hat{T}, \hat{T})$ where $\Psi(\tau = 0, \hat{T}) = Z\left(\beta - \frac{1}{2}\right) - (1-\beta) F$ otherwise.

In both cases, a challenge from group j, i.e., $p_{jt} = 1$, is met by $(\tau_{jt}^r, T_{jt}^r) = (\tau^*, 0)$ and (τ_{it}^r, T_{it}^r) for $i \neq j$ such that $V_i^C(\tau_{it}^r, T_{it}^r \mid \tau_i^e, T_i^e) = V_i^D$ where $V_i^C(\tau_{it}^r, T_{it}^r \mid \tau_i^e, T_i^e)$ is given by (10.13) and V_i^D is given by (10.12).

The discussion above establishes this proposition. The only part that may need more comment is the uniqueness of equilibrium. Recall that if group j makes a proposal to remove the ruler from power, the ruler will respond with $\tau_{jt}^r = \tau^*$, and when $\sigma \in \Sigma^*$, we also have $\tau_{jt}^e = \tau^*$. It may therefore appear that we can construct equilibria where there are challenges along the equilibrium path when $\sigma \in \Sigma^*$, and thus the equilibrium described in part 1 of Proposition 1 is not unique. This is not the case, however. Any combination of strategies where $p_{jt} = 1$ cannot be an equilibrium. If it were, a deviation to $\left(\tau^e_{jt}, T^e_{jt}\right) = (\tau^*, \varepsilon)$ for $\varepsilon > 0$ would be a best response for the ruler, and the strategy of $p_{jt} = 1$ would then cost group j an amount $\varepsilon > 0$. Since a smaller ε is always preferred by the ruler, the only combination of best response strategies is when $\varepsilon \to 0$, which is the one described in the proposition.

This proposition therefore formalizes how the ruler remains in power and is able to transfer resources to himself thanks to the divide-and-rule strategy. He achieves this as follows: when threatened by the "proposer" group, he can always gain the allegiance of the other, "proposed" group, by shifting resources to them. Because the proposed group is pivotal, the ruler can remain in power if he can successfully buy off the proposed group. If this is the case, anticipating this outcome, neither group will attempt to remove the ruler from power, and he will be able to establish a kleptocratic regime transferring resources to himself at the expense of the productive groups in society.

The proposition also highlights the notion of "constrained kleptocratic regime," where the ruler is able to pursue kleptocratic policies transferring resources to himself, but in this endeavor he is constrained by the threat that the two groups will coordinate and remove him from power. To avoid this possibility, the ruler reduces the equilibrium taxes (or sometimes sets 0 taxes and makes positive transfers) to the two groups.

Notice that when $\sigma \in \Sigma^*$, the equilibrium does not feature the notion of "punishment", which was discussed in the introduction. According to this notion, kleptocrats are in power because they can threaten to punish challengers and reward loyal groups. In this case, both on and off the equilibrium path, the group that challenges the ruler is taxed at the rate τ^* and receives 0 transfers. In contrast, when $\sigma \notin \Sigma^*$, $\tau^e_j < \tau^*$, and if group j challenges the ruler, not only will group $i \neq j$ be bribed to cooperate with the ruler, but also group j will be punished with the tax increased to τ^*.

Next we turn to a discussion of a number of natural comparative statics in this model. Most of those are immediate from the inspection of (10.24) and (10.25):

- Greater F makes $\sigma \in \Sigma^*$ more likely, and when $\sigma \notin \Sigma^*$, greater F increases taxes.

  This comparative static is intuitive: greater F, i.e., greater foreign aid, relaxes the budget constraint of the ruler and provides him with more resources to buy off the pivotal group off the equilibrium path. Therefore, greater F makes the kleptocratic regime easier to sustain. This comparative static result suggests that the foreign aid given to many African regimes by the United States and the United Nations during the Cold War period may have had the unforeseen consequence of consolidating kleptocratic regimes. As we discussed in the introduction, this comparative static result may help us understand why in the postwar period, foreign aid appears to have had no positive effect on economic growth on average, and in fact, it may have had a negative effect on the economic outcomes in certain non-democratic countries.

- Greater β makes $\sigma \in \Sigma^*$ less likely, and when $\sigma \notin \Sigma^*$, greater β reduces taxes.

  Greater β means that both groups are more patient. Since the benefit of replacing the ruler—greater returns in democracy—accrues in the future, greater patience makes it less likely that the ruler will be able to maintain his kleptocratic regime. This comparative static suggests that kleptocratic regimes are more likely to emerge in societies where citizens or their political representatives value the future less.

- If $\beta < 1/2$, then greater Z makes $\sigma \in \Sigma^*$ more likely, and when $\sigma \notin \Sigma^*$, it increases taxes. If $\beta > 1/2$, the opposite comparative statics apply.

  Inspection of (10.24) shows that greater natural resource rents create two opposing forces. First, like foreign aid, greater Z relaxes the budget constraint of the ruler, and enables him to sustain his kleptocratic regime by buying off pivotal groups when challenged. Second, greater Z increases the value of democracy. When $\beta < 1/2$, the two groups are sufficiently short-sighted that the first effect dominates. When $\beta > 1/2$, the second effect dominates. The reason why the relevant threshold is 1/2 is that in democracy natural resource rents will be divided between the two groups, whereas off the equilibrium path, the ruler can pay all the rents to the proposed group. Casual empiricism suggests that the case with $\beta < 1/2$ appears more relevant here, and suggests that natural-resource-rich countries may be more prone to kleptocratic regimes. In fact, the comparative static that greater rents from natural resources make kleptocracy more likely (i.e., the case with $\beta < 1/2$) may help us explain why kleptocratic and neopatrimonial regimes have often emerged in resource-rich countries, such as the Democratic Republic of the Congo (Zaire), Sierra Leone, Liberia, and Nigeria.

- Finally, if $\beta > (\eta/(1+\eta))^\eta$, then greater ϖ makes $\sigma \in \Sigma^*$ less likely, and when $\sigma \notin \Sigma^*$, it reduces taxes. If $\beta < (\eta/(1+\eta))^\eta$, the opposite comparative statics apply.

  This result can be obtained by differentiating (10.24). The intuition for the result may be better understood by considering condition (10.20), which separately shows the effect of the productivity of the proposed and the proposer groups. Higher productivity of the proposed group, ω_i, makes the condition less likely to hold (and kleptocracy more likely to survive), because the proposed group has more to gain from democracy. On the other hand, a greater level of productivity of the proposer group, ω_j, makes kleptocracy more likely to survive, because it implies greater tax revenues that the ruler can use to bribe the proposed group. Consequently, higher average productivity $\overline{\omega}$ creates two opposing forces. When the discount factor β is sufficiently large, i.e., $\beta > (\eta/(1+\eta))^{\eta}$, the own effect dominates and greater productivity makes kleptocracy less likely. This comparative static suggests that, as long as β is not very low, societies that are otherwise less productive are also more likely to suffer from kleptocratic regimes and distortionary policies of rulers.

10.2.5. Equilibrium With Inequality. Let us now return to the case where the two groups do not necessarily have the same productivity, i.e., $x \in (0,1]$. First, recall that despite the differences in productivity between the two groups, the analysis above established that the most preferred (unconstrained) policy for the ruler is still $(\tau_{1t}^e, \tau_{2t}^e, T_{1t}^e, T_{2t}^e) = (\tau^*, \tau^*, 0, 0)$. The question is when this policy vector will be possible for the ruler.

To answer this question, recall that $\sigma' = (\overline{\omega}, \eta, \beta, Z, F, x)$ is the vector of parameters, and let Σ_1 the the set of parameters such that $\Psi(\tau^*, T = 0, x) \geq Z\left(\beta - \frac{1}{2}\right) - (1-\beta)F$ and Σ_2 the set of parameters such that $\Psi(\tau^*, T = 0, -x) \geq Z\left(\beta - \frac{1}{2}\right) - (1-\beta)F$. In other words, $\sigma' \in \Sigma_1$ implies that if group 2 makes a proposal to remove the ruler from power, the ruler can make a counteroffer that co-opts group 1 even when along the MPE group 1's members are taxed at the rate τ^* and receive no transfers. Σ_2 is the corresponding set for group 2.

Formally, these two sets are defined by

$$(10.26)\quad \Sigma_1 = \left\{\sigma' : \frac{(1-\beta)\,\overline{\omega}^{1+\eta}\,(1-x)^{1+\eta}}{1+\eta}\left(\frac{\eta}{1+\eta}\right)^{\eta} - \frac{\beta\,\overline{\omega}^{1+\eta}\,(1+x)^{1+\eta}}{1+\eta}\left[1-\left(\frac{\eta}{1+\eta}\right)^{1+\eta}\right] \geq Z\left(\beta-\frac{1}{2}\right)-(1-\beta)F\right\}$$

and

$$(10.27)\quad \Sigma_2 = \left\{\sigma' : \frac{(1-\beta)\,\overline{\omega}^{1+\eta}\,(1+x)^{1+\eta}}{1+\eta}\left(\frac{\eta}{1+\eta}\right)^{\eta} - \frac{\beta\,\overline{\omega}^{1+\eta}\,(1-x)^{1+\eta}}{1+\eta}\left[1-\left(\frac{\eta}{1+\eta}\right)^{1+\eta}\right] \geq Z\left(\beta-\frac{1}{2}\right)-(1-\beta)F\right\}.$$

Since $\Psi(\tau^*, T = 0, x)$ is a decreasing function of x, we have that $\Sigma_1 \subset \Sigma_2$. In other words, when the producers in one of the groups become more productive they also become more willing to oust the ruler. Consequently, the tighter constraint faced by the ruler is to satisfy the more productive group off the equilibrium path. This result reflects the fact that the more productive group has more to gain from democracy, where its members will not be taxed (or more generally, where they will be taxed more lightly). The logic of the political game above therefore implies that, everything else equal, the constraints that the ruler has to worry about is group 2 making an offer and group 1, the more productive group, accepting this proposal.

This observation and a similar analysis to the one above lead to the following proposition:

- Let Σ_1 and Σ_2 be given by (10.26) and (10.27). Then we have:

When $\sigma' \in \Sigma_1$, then the unique MPE is an *unconstrained kleptocratic regime* where $(\tau^e_{1t}, \tau^e_{2t}, T^e_{1t}, T^e_{2t}) = (\tau^*, \tau^*, 0, 0)$ for all t and $i = 1, 2$.

When $\sigma' \notin \Sigma_1$ but $\sigma' \in \Sigma_2$, then the unique MPE is a *partially constrained kleptocratic regime* where the equilibrium policy combination is $(\tau^e_{1t}, \tau^e_{2t}, T^e_{1t}, T^e_{2t}) = (\hat{\tau}_1, \tau^*, \hat{T}_1, 0)$ with

$$\Psi(\hat{\tau}_1, \hat{T}_1, x) = Z\left(\beta - \frac{1}{2}\right) - (1-\beta)\, F. \tag{10.28}$$

When $\sigma' \notin \Sigma_2$, then the unique MPE is a *fully constrained kleptocratic regime* where $(\tau^e_{1t}, \tau^e_{2t}, T^e_{1t}, T^e_{2t}) = (\hat{\tau}_1, \hat{\tau}_2, \hat{T}_1, \hat{T}_2)$ with

$$\Psi(\hat{\tau}_1, \hat{T}_1, x) = \Psi(\hat{\tau}_2, \hat{T}_2, -x) = Z\left(\beta - \frac{1}{2}\right) - (1-\beta)\, F. \tag{10.29}$$

This proposition extends Proposition 1 to a situation with potential heterogeneity in the productivities of the two groups. It also introduces the notion of partially- and fully-constrained kleptocratic regimes: as before, an unconstrained kleptocratic regime pursues the policy most preferred by the ruler. When it is partially constrained, the ruler has to reduce the tax rate on the more productive group, but can tax the less productive group as heavily as he wishes. When the regime is fully constrained, the tax rates on both groups are constrained.

The comparative static results discussed previously continue to apply in this extended model. The new result here is with respect to x, the degree of inequality between the two producer groups. A greater x—greater inequality between the producer groups—makes the unconstrained kleptocratic regime less likely (i.e., it makes it less likely that $\sigma' \in \Sigma_1$). Intuitively, the more binding constraint from the point of view of the ruler is to satisfy the more productive group: when this group becomes even more productive, democracy becomes more attractive for the producers in this group, and therefore, it becomes more difficult for the ruler to buy them off when challenged. This comparative static captures the notion that when there is a strong producer group in society, the ruler has less room to maneuver, and therefore the unconstrained kleptocratic regime is less likely to emerge. Loosely speaking, we can say that a highly productive producer group creates a “balance of power", and this balance between one of the major producer groups and the ruler prevents the most egregious kleptocratic policies. This comparative static result might help us understand why kleptocracies are rare in African countries with powerful producer groups, such as the cattleowners in Botswana or the sugar planters in Mauritius.

10.3. A Model of Politics of Fear

Basic idea: weakly-institutionalized polities in ethnically-divided societies lead to three peculiar features:

- Differential taxes on different ethnic groups
- Support from own ethnic group sufficient to remain in power with high probability
- Replacing rulers leads to uncertainty.

To capture these features, consider an infinitely repeated economy populated by a continuum of citizens of mass 1.

Citizens belong to one of two ethnic groups, A and B. The size of group A is n^A. A group is defined by two distinct sets of characteristics. First, there are some ascriptive characteristics like language or skin color (maybe geographical distribution) that are identifiable and not easily changeable. Second, different groups obtain wealth from different portfolios of economic activities. These portfolios generate w^A and w^B per period, respectively.

Economic activity is assumed to be the only characteristic that a group can change. In particular, a group may decide to switch its efforts to the activity portfolio of the other group, but in the process it loses a fraction of its wealth. Hence if group A switches to B's activity, it obtains $w^A(1-\phi^A)$ instead of w^A. ϕ^i thus captures the extent to which a group's wealth is specific to a particular activity. Let $\omega_t^i = 1$ if group i switches activities in period t. Otherwise $\omega_t^i = 0$.

There is a state that performs two functions: it taxes portfolios of economic activities and uses the proceeds to provide group specific goods.

These group specific goods might be public goods that are so dependent on taste that only one of the groups enjoys them. Alternatively, they may constitute pure patronage such as the allocation of public resources to the region of a group in the form of infrastructure or the granting of lucrative bureaucratic posts (or posts in the army, police, etc.) to members of the favored group. The state is able to discriminate across recipients for public expenditure thanks to the ascriptive characteristics of groups.

On the other hand, taxes are activity specific because in particularly poor developing countries as in Africa, the absence of a competent bureaucracy forces the governments to raise their revenue from indirect taxation. Thus note that the fundamental difference between expenditure and taxation is that patronage can be perfectly targeted to specific groups.

Both groups have identical preferences represented by $\mathbb{E}\sum_{t=0}^{\infty}\delta^t C_t^j$, where C_t^j is the consumption of group j at time t, and δ is the discount factor.

At any point in time, one ethnic group has control of the government. Even though a group nominally has the state captured, real power is exercised by a narrow elite inside the group, and I will call it the Leader. Denote by L^i the leader if she is from group i. In the remainder of the paper, I call the group to which the leader belongs "supporter" group, and the other is the "excluded" group for reasons that will become apparent. Each group has an unlimited supply of identical leaders from which to choose.

Denote τ^{ij} the tax level that a leader of group i levies on (the activities of) group j. Similarly, let Z^{ij} be the amount that leader of group i spends on patronage for group j. Obviously $i, j \in \{A, B\}$. The amount Z^{ij} provides utility $R(Z^{ij})$ to group j with $R' > 0$, $R'(0) > 1$, $R'' < 0$ and $R(0) = 0$. Group $-j$ receives no utility from Z^{ij}.

This economy has two fundamental states, $S_t \in \{A, B\}$, denoting whether power is captured by group A or group B in period t.

The instantaneous utility of a citizen of group j in state S is thus:

$$C(S, \omega^j) = (1 - \omega^j)(w^j - \tau^{Sj}) + \omega^j((1 - \phi^j)w^j - \tau^{S-j}) + R(Z^{Sj})$$

where time subscripts have been omitted for notational simplicity.

Even though the leader belongs to group S_t she has self-serving interests. In particular, she wants to maximize the funds that she can divert for her own uses. A leader of group A obtains instantaneous utility (the expression for B is just symmetric) as long as she is in power:

$$U^A = n^A(\tau^{AA} - Z^{AA}) + (1 - n^A)(\tau^{AB} - Z^{AB})$$

and discounts future payoffs by δ. When a leader is not in power, she obtains 0 utility per period.

The weakness of institutions and the importance of ascriptive links is captured in the model by the following assumptions. First, assume that as long as the incumbent leader retains the support of her kin group, she maintains her position. The unique credible source of support and thus the unique credible promise of future patronage is given by the ruler's ethnic linkage with her own group. For simplicity I assume that the "excluded" group has no chance of recovering power if the incumbent leader keeps the support from her group.

Second, if the supporters of an incumbent leader decide to subvert the authority of their leader and want to oust her from power, they succeed automatically. Hence the relevant constraint on the rapacious interests of the leader is the need to keep the support of her group. This is the sense in which the position of the leader is weak: she needs the active support of a sizable share of the population to maintain power.

Third, when a leader is ousted from power, the state does not perform its functions for that period. Moreover, the group that is not in power will try to use this opportunity to grab power and seat a leader from its ranks. This captures the reality of Personal Rule regimes in which successions are always uncertain matters, resolved in non-institutionalized ways. Thus, I assume that the status of the group in power will change with probability $1-\gamma^S$. γ^S captures the degree to which the grip on power of group S is solid in the presence of upheaval from its own members. On the other hand it could capture the entrenchment of the leader or the relative competence of the next replacement from the group

The timing of each stage game, given state S_t, is the following:

(1) Leader L^S announces the policy vector $P_t = \{\tau_t^{SA}, \tau_t^{SB}, Z_t^{SA}, Z_t^{SB}\}$

(2) The citizens of group S_t decide to "subvert" $s_t = 1$ or not $s_t = 0$

(3) If $s_t = 0$, the citizens decide ω_t^A and ω_t^B and afterwards the policy vector is implemented, payoffs are realized and the next period starts with $S_{t+1} = S_t$

(4) If $s_t = 1$, the leader is ousted immediately and the "revolt" vector $P_r = \{0, 0, 0, 0\}$ is implemented. With probability $1-\gamma^{S_t}$, S_t loses power and the next period starts with $S_{t+1} = -S_t$. Otherwise, the next period starts with $S_{t+1} = S_t$

There are a number of features of the model that are worth stressing.

First, note that collective action within a group is not an issue in this model. The focus of the argument is on the forces that allow rents to be appropriated by a weak leader, and not competed away by different elites inside the same group. Adding heterogeneity and a collective action problem would only help the current leader to steal even more, because she would find it easier to disrupt coordination.

Second, and in the same spirit, I do not allow the leader access to any repression instrument: if she loses the support of her group, she is replaced at no explicit cost.

Third, the economy is modeled in a very reduced form: taxes entail no efficiency cost. As a consequence, all potential for inefficiency is in the allocation of Z. A more complex model would allow us to analyze other aspects of the number of mislead policies that these leaders imposed on their economies, but again, the focus of the model is on political survival and capture of the electorate.

Finally, note that no difference is made between democracy and dictatorship in the model. The evidence from Africa shows that democracies have not behaved differently than dictatorships at the time of supporting kleptocracies and corruption. In my analysis, the reason is that both types of regimes have been able to play ethnic divisions and patronage networks in exactly the same ways. The analysis will reveal that institutional reform needs to go further than getting people to vote. It has to include effective constraints on the capacity of the leaders to treat ethnic groups differently, and it should include mechanisms directed to smooth intragroup competition.

10.3.1. Definition of Equilibrium. The equilibrium concept is (pure strategy) Markov Perfect Equilibrium. In this type of equilibria, strategies can only be contingent on the payoff-relevant state of the world and the prior actions taken within the same period. In the present case, this will provide an equilibrium that is stationary as long as the group in power, which characterizes completely the state space, does not change.

The state space of this economy includes only two elements, $\Theta = \{A, B\}$, denoting whether power is captured by group A or group B at the beginning of period t. Denote the state at each period by S_t, where obviously $S_t \in \Theta$, $\forall t = 0, 1, 2....$ Assume that each group has a set of potential leaders from which replacements will be drawn randomly. Call these two sets of leaders Δ^A and Δ^B. At any point in time, the leader in power is denoted by L^A or L^B depending on the group she was drawn from. Denote by $\tilde{}L^A$ the potential leaders that belong to Δ^A but are not in power currently. $\tilde{}L^B$ is defined symmetrically. The strategy of the current leader L^A is denoted by P^A and it is a four-tuple $\{\tau^{AA}, \tau^{AB}, Z^{AA}, Z^{AB}\} \in \mathbb{R}^4_+$ when $S_t = A$. When either $S_t = B$ or $S_t = A$ but a leader belongs to $\tilde{}L^A$, her set of strategies is empty. The symmetric definition holds for the strategies of leaders L^B. Note that I do not consider the identity of the leader part of the state space.

The strategy of group A is denoted $\sigma^A(S, P^S)$ and depends on both the state of political capture and the policy vector proposed by the leader. It determines two actions, $\{s^A, \omega^A\}$ that have been defined above as the decision to subvert and the decision to switch economic activities. If $S_t = A$, $s^A \in \{0, 1\}$, that is, if the leader is from group A, this group can decide to give her support or to subvert her authority. On the other hand, if $S_t = B$, $s^A = \emptyset$. $\omega^A \in \{0, 1\}$ independently of the state. The symmetric definition holds for the strategy space of citizens of group B.

State transitions work as follows. $S_{t+1} = S_t$ whenever $s_t^S = 0$. If $s_t^S = 1$, that is, if there is subversion, $S_{t+1} = -S_t$ with probability $1 - \gamma^{S_t}$. Denote this transition function $T(\sigma^S, S)$. Hence, power only changes hands with positive probability when the supporter group subverts. Otherwise the state remains the same. Since in equilibrium I will show that there is never subversion, the unique equilibrium is stationary.

A (pure strategy) Markov Perfect Equilibrium for this game is a combination of strategies denoted by $\{\tilde{P}^A, \tilde{P}^B, \tilde{\sigma}^A, \tilde{\sigma}^B\}$ such that all four strategies are best responses to the other three for all possible states. In particular, consider the following set of Bellman equations:

$$V^A(S) = \max_{\sigma^A}\{C^A(S, \tilde{P}^S, \sigma^A(S, P^S), \tilde{\sigma}^B) + \delta \sum_{S \in \Theta} V^A(S')T(\sigma^S, S)\} \tag{10.30}$$

$$V^B(S) = \max_{\sigma^B}\{C^B(S, \tilde{P}^S, \sigma^B(S, P^S), \tilde{\sigma}^A) + \delta \sum_{S \in \Theta} V^B(S')T(\sigma^S, S)\} \tag{10.31}$$

$$W^A_{L^A}(A) = \max_{P^A}\{U^A(P^A, \tilde{\sigma}^A, \tilde{\sigma}^B) + \delta \sum_{S \in \Theta} W^A_{\Delta}(S')T(\tilde{\sigma}^A(A, P^A), A)\} \tag{10.32}$$

$$W^B_{L^B}(B) = \max_{P^B}\{U^B(P^B, \tilde{\sigma}^B, \tilde{\sigma}^A) + \delta \sum_{S \in \Theta} W^B_{\Delta}(S')T(\tilde{\sigma}^B(B, P^B), B)\} \tag{10.33}$$

where C^j denotes the consumption of citizen j as a function of the state S and the strategies of the leader in power and both sets of citizens. $V^j(S)$ denotes the value function for citizen j in state S. $W^i_{L^S}(S)$ denotes the value function for leader from group i in state S, when she is the current leader L^S. To complete the definition, note that $W^A_{\Delta}(B)$, $W^A_{\sim L^A}(A)$, $W^B_{\Delta}(A)$ and $W^B_{\sim L^B}(B)$ are completely independent of any decision that the particular leader could take. They only depend on the probability that, in equilibrium, a particular leader will be in power in the future. Since it will be shown that there is no revolt in equilibrium, whether the opposite group is in power or whether another leader from her own group is in power, the continuation value of an out-of-power leader is 0. As a consequence, these are not interesting strategic objects in this game.

A Markov Perfect Equilibrium is thus a combination of strategies $\{\tilde{P}^A, \tilde{P}^B, \tilde{\sigma}^A, \tilde{\sigma}^B\}$ such that $\tilde{\sigma}^A$ solves (10.30), $\tilde{\sigma}^B$ solves (10.31), $\tilde{P}^A$ solves (10.32) and $\tilde{P}^B$ solves (10.33).

10.3.2. Analysis. Assume without loss of generality that $S_t = A$. The equilibrium is characterized by backwards induction within each stage game. Examine first the decision to switch the sector of production. Take first B producers. Note that the decision to switch does not affect continuation utilities, hence only the static difference in payoffs is relevant. After observing the policy vector P_t, they will switch sector only if the loss in wealth is smaller than the difference in taxation. Formally,

$$\omega_t^B = 1 \text{ iff } w^B - \tau^{AB} < (1 - \phi^B)w^B - \tau^{AA}$$

Since it is in the interest of the ruler not to allow this switch, which is wasteful, this ability to switch provides an upper bound on the differential taxation that the ruler can levy on group B. The effective constraint on the ruler will thus be

$$\tau^{AB} \leq \phi^B w^B + \tau^{AA} \tag{10.34}$$

The equivalent restriction for group A is then

$$\tau^{AA} \leq \phi^A w^A + \tau^{AB} \tag{10.35}$$

Obviously, both restrictions cannot be binding at the same time.

Let us examine now the decision to subvert by A supporters. Note that the leader is the first player to act in the stage game. As a consequence, since strategies can only be conditional on the state of the economy, a leader L^A always proposes the same policy vector P^A. Upon observing P^A, if there is no subversion ($s_t = 0$), A supporters obtain:

$$w^A - \tau^{AA} + R(Z^{AA}) + \delta V^A(A)$$

Alternatively, if they subvert, $s_t = 1$, they expect:

$$w^A + \delta\gamma^A V^A(A) + \delta(1-\gamma^A)V^A(B)$$

Hence the non-subversion condition reduces to:

$$\tau^{AA} - R(Z^{AA}) \leq \delta(1-\gamma^A)(V^A(A) - V^A(B)) \tag{10.36}$$

Note that the ruler will always satisfy this constraint by subgame perfection. Not satisfying it gives her no benefit because in the period she is thrown out she already receives 0 utility, plus she will obtain 0 forevermore, while being in power implies receiving positive rents each period. Hence in any MPE there will never be any ousting of a ruler. Since no change of ruler implies no change of state, any MPE of this game will be stationary, with the ruler proposing always the same P^A which will be accepted every time. After some rearranging and knowing that the payoffs will be stationary, the constraint that the ruler has to observe to avoid being thrown out can be written as follows:

$$\tau^{AA} - R(Z^{AA}) \leq \frac{\delta}{1-\delta}(1-\gamma^A)[\tilde{\tau}^{BA} - R(\tilde{Z}^{BA}) - \tilde{\tau}^{AA} + R(\tilde{Z}^{AA})]$$

Where the superscript $\sim$ denotes equilibrium values. Note that the right hand side of the inequality contains all expected future terms. For notational simplicity we will denote $\Phi^A = \frac{\delta}{1-\delta}(1-\gamma^A)[\tilde{\tau}^{BA} - R(\tilde{Z}^{BA}) - \tilde{\tau}^{AA} + R(\tilde{Z}^{AA})]$. This term summarizes the way in which future expected equilibrium play affects present decisions. With these ingredients, now we are able to posit the problem of ruler L^A:

$$\max_{\{\tau^{AA},\tau^{AB},Z^{AA},Z^{AB}\}} \quad n^A(\tau_t^{AA} - Z_t^{AA}) + (1-n^A)(\tau_t^{AB} - Z_t^{AB}) + \delta W_{L^A}^A(A) \tag{10.37}$$

subject to

$$\begin{aligned}
\tau^{AB} &\leq \phi^B w^B + \tau^{AA} & [\lambda] \\
\tau^{AA} &\leq \phi^A w^A + \tau^{AB} & [\nu] \\
\tau^{AA} - R(Z^{AA}) &\leq \Phi^A & [\mu] \\
0 &\leq Z^{AB} & [\rho]
\end{aligned}$$

The ruler thus maximizes her returns per period, conditional on avoiding any wasteful switching and subversion. The first order conditions of this program yield:

$$n^A + \lambda - \nu - \mu = 0 \tag{10.38}$$

$$1 - n^A - \lambda + \nu = 0 \tag{10.39}$$

$$-n^A + \mu R'(Z^{AA}) = 0 \tag{10.40}$$

$$-(1-n^A) + \rho = 0 \tag{10.41}$$

The first order conditions are simple and easy to interpret. From (10.41) it is obvious that $Z^{AB} = 0$. The reason is that providing patronage good to the excluded group is costly and yields no benefit, since the supporter group is enough to maintain power. From (10.38) and (10.39) and the fact that λ and ν cannot both be strictly positive at the same time we learn that $\nu = 0$, $\lambda = 1 - n^A$ and $\mu = 1$. $\nu = 0$ implies that the second restriction is not saturated. The analysis thus reveals that the ruler endogenously chooses to discriminate against the "excluded" group. The reason is that she only needs the support of her own group to remain in power. Quite intuitively, then, the leader will tax the excluded group as much as she can, that is, to the point in which the first constraint is binding.

Note that every dollar that the ruler is able to tax her own supporters is worth more than one for her, because it allows her to tax an extra dollar on the excluded group. Note from (10.40) that μ multiplies the return from the last unit of patronage given to group A. The cost of this last unit is only n^A, but its return is increased taxation from the whole population (because $\mu = 1$). This disparity is the reason for inefficient patronage provision. The mechanism works as follows: the non-subversion constraint is binding and hence an increase in $R(Z^{AA})$ allows the ruler to increase taxation on her supporters and, since the no-switching constraint is also binding, taxation on the excluded increases in parallel. Hence, patronage good for A is overprovided in equilibrium: the ruler considers the benefits from increasing taxation from the whole population, while a social planner would only consider the group that receives utility from it. This distortion is worse the narrower the basis of support of the ruler (the smaller n^A).

Formally, the stage program yields the following solution:

$$Z^{AB} = 0 \tag{10.42}$$
$$R'(Z^{AA}) = n^A \tag{10.43}$$
$$\tau^{AA} = \Phi^A + R(Z^{AA}) \tag{10.44}$$
$$\tau^{AB} = \phi^B w^B + \Phi^A + R(Z^{AA}) \tag{10.45}$$

The solution for patronage public goods (10.42) and (10.43) is thus independent of expectations of future play, but this is not the case for the amount of resources that the leader can extract from both groups. In fact, the solution above presents a mapping between future equilibrium play and current taxation. Remember that, in equilibrium, another symmetric problem is solved by any L^B leader in power. The solution to the program for L^B is:

$$Z^{BA} = 0$$
$$R'(Z^{BB}) = 1 - n^A$$
$$\tau^{BB} = \Phi^B + R(Z^{BB}) \tag{10.46}$$
$$\tau^{BA} = \phi^A w^A + \Phi^B + R(Z^{BB}) \tag{10.47}$$

Denote the mapping from expectations to current play $\Gamma(\Phi^A, \Phi^B) = (\tau^{AA}, \tau^{AB}, \tau^{BA}, \tau^{BB})$, given by (10.44), (10.45), (10.46) and (10.47). Moreover, the definition of Φ^A (and the symmetric definition of Φ^B) provides a mapping from actual play to consistent expectations $\Psi(\tau^{AA}, \tau^{AB}, \tau^{BA}, \tau^{BB}) = (\Phi^{CA}, \Phi^{CB})$. The equilibrium posits the requirement that these expectations be consistent with future play. In this context this reduces to finding a fixed point of the mapping that relates expectations into themselves: $\Psi(\Gamma(\Phi^A, \Phi^B)) = (\Phi^{CA}, \Phi^{CB})$. Explicitly, this mapping is the following:

$$\Phi^{CA} = \frac{\delta}{1-\delta}(1-\gamma^A)[\phi^A w^A + \Phi^B + R(Z^{BB}) - \Phi^A - R(Z^{AA}) + R(Z^{AA})]$$
$$\Phi^{CB} = \frac{\delta}{1-\delta}(1-\gamma^B)[\phi^B w^B + \Phi^A + R(Z^{AA}) - \Phi^B - R(Z^{BB}) + R(Z^{BB})]$$

For simplicity denote $\zeta^i = \frac{\delta}{1-\delta}(1-\gamma^i)$. Solving this system for the fixed point $(\Phi^A, \Phi^B) = (\Phi^{CA}, \Phi^{CB})$ yields:

$$\begin{aligned}\Phi^A &= \frac{\zeta^A(1+\zeta^B)(\phi^A w^A + R(Z^{BB})) + \zeta^A\zeta^B(\phi^B w^B + R(Z^{AA}))}{1+\zeta^A+\zeta^B} \\ \Phi^B &= \frac{\zeta^B(1+\zeta^A)(\phi^B w^B + R(Z^{AA})) + \zeta^A\zeta^B(\phi^A w^A + R(Z^{BB}))}{1+\zeta^A+\zeta^B}\end{aligned}$$

Since there is a single fixed point, uniqueness of MPE is shown. This discussion establishes the following proposition.

The model presents a unique MPE. If, without loss of generality, initial conditions determine that group A is in control, the equilibrium is such that:

PROPOSITION 10.2. *L^A proposes the following policy vector:*

$$Z^{AA} \equiv Z^A_* \text{ such that } R'(Z^A_*) = n^A$$

$$Z^{AB} = 0$$

$$\tau^{AA} = \frac{\zeta^A(1+\zeta^B)\phi^A w^A + \zeta^A\zeta^B\phi^B w^B}{1+\zeta^A+\zeta^B} + \frac{(1+\zeta^A)(1+\zeta^B)R(Z^A_*) + \zeta^A(1+\zeta^B)R(Z^B_*)}{1+\zeta^A+\zeta^B} \tag{10.48}$$

$$\tau^{AB} = \frac{\zeta^A(1+\zeta^B)\phi^A w^A + (1+\zeta^A)(1+\zeta^B)\phi^B w^B}{1+\zeta^A+\zeta^B} + \frac{(1+\zeta^A)(1+\zeta^B)R(Z^A_*) + \zeta^A(1+\zeta^B)R(Z^B_*)}{1+\zeta^A+\zeta^B} \tag{10.49}$$

The citizens of group A accept this policy vector: $s^A = 0$.

No activity switch occurs: $\omega^A = \omega^B = 0$.

If group B starts in control, the equilibrium is symmetric.

The unique MPE of the model provides a potential explanation for many features of the post-colonial political economy of Africa.

First, the model endogenously generates inefficient policies. Note that in the simple framework proposed here, the unique potential source of inefficiency is the excessive allocation of patronage to a particular group. Since the opportunity cost of public funds is 1, the fact that the marginal return to patronage for the supporter group is $n^A < 1$ shows that political needs cause inefficiencies. This feature of the equilibrium helps explain the patterns of inefficient taxation and inefficient transfers coexisting in the same group highlighted in the seminal work by Bates (1981) for agricultural policies in tropical Africa. The ruler needs to buy support from her own group while, at the same time, wants to extract a lot of resources from the economy. The best way of doing so, given the absence of lump-sum taxation, is by taxing both groups and then returning some patronage to the supporters even if it is highly wasteful. This is a general pattern of statism in Africa.

Second, the model predicts a very strong bias in the allocation of public funds. The excluded group receives no public benefits while the supporter group receives public resources beyond the optimal point. The use of public money in the form of bureaucratic posts, infrastructure or even access to schools as a form of patronage, as well as the ethnic bias in the allocation of these goods has been widely documented in Africa. Gikuyus and later Kalenjin in Kenya, northern groups in both Nigeria and Uganda or Tutsis in Burundi are just salient examples that reproduce across the continent. The bias in favor of the ruling group is conspicuous and is actually one of the basic sources of resentment between ethnic groups. Not only access to these positions is biased, but is accompanied by an absence of meritocratic pressure that makes them ripe for all kinds of corruption, official and unofficial.

Third, the bias is not only present in the allocation of patronage: taxation is also differential across groups. In particular, in addition to taxes levied on the supporter group, the model shows that the excluded group is expropriated from the non-transferable share of its wealth. Bates (1981) and Bates (1989) provide evidence of this pattern: in Ghana and Uganda, among other examples, the coalition that supported the leader extracted resources from the coffee and cocoa planters. These are crops that involve a lot of specific long term investment. On the contrary, in Kenya the Gikuyu controlled the coffee growing parts of the country, and hence the discrimination against these crops was much less evident.

The combination of higher taxation and absence of patronage makes the excluded group obviously worse off than the supporter group. As a consequence, whenever there is a change in the group controlling power, the patterns of taxation change and purges follow in order to make space for the new elites. For instance, the ascension to power by Moi in Kenya was followed by a substitution of Gikuyus by Kalenjin in all echelons of the state. In Ghana, cocoa has been heavily taxed by all governments, civil and military, except the one headed by Kofi Busia, a native from the Ashanti region which contains a large share of smallholders that grow cocoa. In Cameroon, the substitution of Ahidjo in 1982 unleashed another deep ethnic purge of the bureaucracy. Similar dynamics are found in Nigeria. Ironically, these purges tend to take place under the excuse of anti-corruption initiatives. These switches prove the use of public resources as patronage, as well as the conscious status of "ruling" groups versus "excluded" groups.

This pattern of discrimination both in raising revenue and in public expenditures supports the vision that a particular ethnic group has the government captured. The model suggests that the actual benefits of such capture are not spread throughout the group. The particular elite that holds power extracts so much resources that part of the money comes from the pockets of non-elite members of the group. In equilibrium it is very easy to see that (10.48) implies that $R(Z_*^A) - \tau^{AA} < 0$.

This result is also consistent with casual empiricism. In Kenya, a potential political cleavage, and a reason why Kenyatta used the ethnic card to maintain power was the situation of landless Gikuyus, most of them ex Mau Mau fighters. These downtrodden masses did not obtain anything from the regime, even though it was clear to all observers and political participants that Kenyatta was at the helm of a "Gikuyu" regime. Emphasizing the fact that the majority of Gikuyu were not actually receiving their share of the spoils was obviously threatening to the regime. That is why the leadership had to act hastily whenever any political entrepreneur tried to shed light on these facts. This included the assassination of a popular politician in 1975. Wa Wamwere describes this unbalance in the reception of spoils in a colorful way:

> "The cream of government service goes to the ruling ethnic elites, the crumbs to the lesser ethnic elites, and dust to members of the so-called ruling ethnic community" and "Among the Gikuyu of Kenya, the approving masses are called grill lickers, *njuna ndara*".

The fact that non-elites are not receiving much from the government is by no means unique to Kenya.

Fourth, the results of the model rationalize the existence of kleptocratic elites supported by masses of impoverished ethnic followers. Even though in absolute terms the masses are made worse off by the existence of rent-creating policies, in relative terms it is much better to belong to the group in power than to the excluded group, and hence they are willing to defend the status quo vis à vis a leader from another group. The members of a narrow elite around the leader are thus the ones extracting the lion's share of the rents that these inefficient policies create. Evidence of Kleptocratic tendencies abound in Africa, but Mobutu's Zaire is probably the most cited example. Sani Abacha in Nigeria or Daniel arap Moi in Kenya have been able to amass personal fortunes counted in the billions of dollars. The "clan de la madame" in Habyarimana's Rwanda is another example of concentrated wealth extracted from the state. Even a relatively well-considered leader, such as Houphouet-Boigny in Cote d'Ivoire had his share of personal aggrandizement projects, such as a marble covered cathedral in his home town. Consistent with this concentration of wealth at the highest levels of leadership, Africa is the continent with highest capital flight.

Finally note that the perverse effect of divisions in this model has nothing to do with diversity in the utility functions, or animosity. The simple inconsequential fact that, in a given society, some people have a blue skin while others have it green is enough for bad governance if institutions are weak. The leader uses these ascriptive characteristics to build a coalition of support just by deciding that a condition to receive patronage is to belong to the "right" group.

The theoretical reason that supports kleptocratic regimes in this model is summarized in expression (10.36). It makes clear that as long as the supporter group observes a difference between being in the supporter status and being excluded under the leadership of the opponent group, there is a surplus that the current leader can expropriate from her own supporters. In addition, the more a leader can extract from her supporters, the more she can extract from the excluded group, thanks to (10.34) being binding in equilibrium. As a consequence, there is an amplification effect of any characteristic of the economy that allows one ruler to steal. Assume that the institutional or economic technology of this society changes so that L^B is now able to steal more from her group if she is ever in power. An A citizen understands that, in equilibrium, this will mean that should he ever fall into an excluded status, his plight will be worse. This reduces $V^A(B)$ in equilibrium. But obviously, this looses the non-subversion constraint for L^A and as a consequence, L^A is able to increase τ^{AA} to the point where her supporters are again indifferent between giving her support or subverting and taking a lottery that now is much less favorable, since both $V^A(A)$, and $V^A(B)$ are reduced. This amplification mechanism is the reason why in the expressions for equilibrium taxation in Proposition 1 the economic and institutional characteristics of both groups appear: γ^A, γ^B as well as ϕ^A, ϕ^B play a role in the equilibrium expressions for both τ^{AA} and τ^{AB}.

On the other hand, the substantive forces that allow the leader to create a wedge between supporters and excluded are clarified in expressions (10.48) and (10.49). In particular, (10.48) shows that the leader is able to exploit the absence of institutional constraints in two dimensions: the non-transferabilities in the economy -$\phi^A$and ϕ^B- allow her to use differential taxation, and her discretion in the allocation of patronage allows her to build clientelist networks. These two effects are separated in the two summands of (10.48) and (10.49).

In the model, the net amount of funds that the leader L^A is able to extract equals $X^A = \Phi^A + R(Z_*^A) + (1-n^A)\phi^B w^B - n^A Z_*^A$. An analysis of comparative statics yields:

$$\begin{aligned}\frac{\partial X^A}{\partial \phi^A} &= \frac{\zeta^A(1+\zeta^B)}{1+\zeta^A+\zeta^B} w^A > 0 \\ \frac{\partial X^A}{\partial \phi^B} &= (\frac{\zeta^A\zeta^B}{1+\zeta^A+\zeta^B} + 1 - n^A) w^B > 0\end{aligned}$$

Rent extraction from both groups, independently of the allegiance of the leader, is increasing in the share of non-transferable resources in the economy. This result implies that starting from a situation with low $\phi^A$and ϕ^B, an increase in the share of non-transferable wealth anywhere in the economy increases equilibrium misbehavior by the ruler. This result provides a rationale for the well documented “natural resource curse":

Comparative statics with respect to the ethnic demographic balance are ambiguous. On the one hand, all the direct effects predict a reduction in stealing: increasing n^A reduces the benefits from distorting the patronage good for two reasons: first, rents are reduced at each level of provision because it becomes more expensive to provide it. Moreover, the optimal level of distortion is reduced because the returns are reduced (less people in the excluded group to pay for it).

In addition, increasing n^A reduces the fraction of population excluded, and hence reduces the extra revenue that comes from the extraction of their non-transferable resources.

However, there is a third, indirect effect, that makes the overall effect ambiguous: increasing n^A means that, should the group ever lose power, a potential L^B would be able to steal more because she would additionally distort the allocation of Z_*^B since her basis of support would now be smaller. Using the same logic of amplification, this allows an L^A leader extra room for stealing. Explicitly, the partial derivative has the following expression:

$$\frac{\partial X^A}{\partial n^A} = \frac{\zeta^A\zeta^B}{1+\zeta^A+\zeta^B}\frac{R'(Z_*^A)}{R''(Z_*^A)} - Z_*^A - \phi^B w^B - \frac{\zeta^A(1+\zeta^B)}{1+\zeta^A+\zeta^B}\frac{R'(Z_*^B)}{R''(Z_*^B)}$$

In this expression, the first two summands represent the rents lost from the ability to distort Z_*^A and the third is the direct loss that is a consequence of the smaller size of the excluded group. The last summand represents the indirect effect, and it is positive. For general functional forms of $R(.)$ this expression cannot be signed, but note that if $R(.)$ is a power function, $\frac{R'(Z)}{R''(Z)}$ is increasing in Z. Hence, if the third indirect effect ever dominates, it will do so at high levels of n^A. That is, when the A group includes a wide majority of the population, the prospect of falling under an L^B is most terrifying because she will have a very narrow basis of support, and hence she will use extreme distortions of patronage to steal.

An analysis of the relationship between institutional certainty and the amount of stealing yields:

$$\frac{\partial X^A}{\partial \gamma^A} = -\frac{\delta(1+\zeta^B)[(1+\zeta^B)(\phi^A w^A + R(Z_*^B)) + \zeta^B(\phi^B w^B + R(Z_*^A))]}{(1+\zeta^A+\zeta^B)^2} < 0$$
$$\frac{\partial X^A}{\partial \gamma^B} = -\frac{\delta\zeta^A[\zeta^A(\phi^A w^A + R(Z_*^B)) + (1+\zeta^A)(\phi^B w^B + R(Z_*^A))]}{(1+\zeta^A+\zeta^B)^2} < 0$$

From these comparative statics it is clear that the level of rent extraction is diminishing in both γ^A and γ^B. Hence it diminishes with institutional certainty. Note that the leader can extract more the lighter is the grip on power of her followers, that is, the higher is the probability that replacing her will be accompanied by a loss of power. While γ^i depends on characteristics of the polity beyond the control of the ruler, such as the demographic ethnic balance, it certainly depends as well on institutional factors. In particular, it has been shown that uncertainty in the succession process is a characteristic of systems of personal rule. The logic of the model shows that the leader has no incentive to strengthen the institutional framework if this means increasing her accountability. This sheds light on several facts.

First, this is consistent with the behavior of the leadership in most African countries: from the moment of independence, even the first prophetic leaders such as Nkrumah, clamped down on opposition, banned political parties, used the police and the military in a partisan way, did not respect judicial independence or any kind of separation of powers and imposed censorship on the press. These are not actions of rulers interested in institutional consolidation.

Second, it is very important for the leaders not to allow the presence of a strong and obvious second-in-command. This would permit her followers to replace her and quickly coordinate on giving support to this alternative focal point and thus reduce the risk of being taken over.

10.4. Incumbency Veto Power and Persistence of Bad Governments

Electoral failures can also result because of some degree of incumbency veto power, for example, because some member of the government needs to give his or her consent to a change in government ("somebody to switch off the lights"). This is based on Acemoglu, Egorov and Sonin (2010). The following example illustrates this point and shows that even a minimal amount of veto power by government incumbents who enjoy remaining in office (because of rents or corruption) can lead to the emergence of very "bad" governments in the sense that the government could consist of the least competent individuals in society and fail to provide adequate public goods.

EXAMPLE 10.1. *Suppose that the society consists of* $n \geq 6$ *of individuals, and that any* $k = 3$ *individuals could form a government. A change in government requires both the support of the majority of the population and the consent of* $l = 1$ *member of the government, so that there is a "minimal" degree of incumbency veto power. Suppose that individual* j *has a level of competence* γ_j*, and order the individuals, without loss of any generality, in descending order according to their competence, so* $\gamma_1 > \gamma_2 > ... > \gamma_n$*. The competence of a government is the sum of the competences of its three members. Each individual obtains utility from the competence level of the government and also a large rent from being in office, so that each prefers to be in office regardless of the competence level of the government. Suppose also that individuals have a sufficiently high discount factor, so that the future matters relative to the present.*

It is straightforward to determine the stable governments that will persist and remain in power once formed. Evidently, $\{1, 2, 3\}$ *is a stable government, since it has the highest level of competence, so neither a majority of outsiders nor members of the government would like to initiate a change (some outsiders may want to initiate a change: for example,* 4*,* 5*, and* 6 *would prefer government* $\{4, 5, 6\}$*, but they do not have the power to enforce such a change). In contrast, governments of the form* $\{1, i, j\}$*,* $\{i, 2, j\}$*, and* $\{i, j, 3\}$ *are unstable (for* $i, j > 3$*), which means that starting with these governments, there will necessarily be a change. In particular, in each of these cases,* $\{1, 2, 3\}$ *will receive support from both one current member of government and from the rest of the population, who would be willing to see a more competent government.*

Consider next the case where $n = 6$ *and suppose that the society starts with the government* $\{4, 5, 6\}$*. This is also a stable government, even though it is the lowest competence government and thus the worst possible option for the society as a whole. This is because any change in government must result in a new government of one of the following three forms:* $\{1, i, j\}$*,* $\{i, 2, j\}$*, or* $\{i, j, 3\}$*. But we know that all of these types of governments are unstable. Therefore, any of the more competent governments will ultimately take the society to* $\{1, 2, 3\}$*, which does not include any of the members of the initial government. Since individuals are relatively patient, none of the initial members of the government would support (consent to) a change that will ultimately exclude them. As a consequence, the initial worst government persists forever. Returning to our discussion of the unwillingness of certain governments to include skilled technocrats, this example shows why such a technocrat, for example individual* 1*, will not be included in the government* $\{4, 5, 6\}$*, even though he would potentially increase the quality and competence of the government substantially.*

One can further verify that $\{4,5,6\}$ *is also a stable government when* $l = 3$*, since in this case any change requires the support of all three members of government and none of them would consent to a change. In contrast, under* $l = 2$*,* $\{4,5,6\}$ *is not a stable government, and thus the quality of the government is higher under intermediate incumbency veto power,* $l = 2$*, than under* $l = 1$ *or* $l = 3$*.*

Now consider the same environment as above, but with potential changes in the competences of the agents. For example, individual 4 may see an increase in his competence, so that he becomes the third most competent agent (i.e., $\gamma_4' \in (\gamma_3, \gamma_2)$*). Suppose that shocks are sufficiently infrequent so that stability of governments in periods without shocks is given by the same reasoning as for the nonstochastic case. Consider the situation starting with the government* $\{4,5,6\}$ *and* $l = 1$*. Then, this government remains in power until the shock occurs. Nevertheless, the equilibrium government will eventually converge to* $\{1,2,3\}$*; at some point a shock will change the relative competences of agents 3 and 4, and the government* $\{4,5,6\}$ *would become unstable; individual 4 would support the emergence of the government* $\{1,2,4\}$*, which now has the highest competence. In contrast, when* $l = 3$*, the ruling government remains in power even after the shock. This simple example thus illustrates how, even though a regime with fewer veto players does not ensure better outcomes in nonstochastic environments, it may provide greater flexibility and hence better long-run outcomes in the presence of shocks.*

Let us next model this situation in the context of a dynamic game.

10.4.1. Model. Consider a dynamic game in discrete time indexed by $t = 0, 1, 2, \ldots$. The population is represented by the set $\mathcal{I}$ and consists of $n < \infty$ individuals. We refer to non-empty subsets of $\mathcal{I}$ as *coalitions* and denote the set of coalitions by $\mathcal{C}$. We also designate a subset of coalitions $\mathcal{G} \subset \mathcal{C}$ as the set of *feasible governments*. For example, the set of feasible governments could consist of all groups of individuals of size k_0 (for some integer k_0) or all groups of individuals of size greater than k_1 and less than some other integer k_2. To simplify the discussion, we define $\bar{k} = \max_{G \in \mathcal{G}} |G|$, so $\bar{k}$ is the upper bound for the size of any feasible government: i.e., for any $G \in \mathcal{G}$, $|G| \leq \bar{k}$. It is natural to presume that $\bar{k} < n/2$.

In each period, the society is ruled by one of the feasible governments $G^t \in \mathcal{G}$. The initial government G^0 is given as part of the description of the game and G^t for $t > 0$ is determined in equilibrium as a result of the political process described below. The government in power at any date affects three aspects of the society:

(1) It influences collective utilities (for example, by providing public goods or influencing how competently the government functions).

(2) It determines individual utilities (members of the government may receive additional utility because of rents of being in office or corruption).

(3) It indirectly influences the future evolution of governments by shaping the distribution of political power in the society (for example, by creating incumbency advantage in democracies or providing greater decision-making power or veto rights to members of the government under alternative political institutions).

We now describe each of these in turn. The influence of the government on collective utilities is modeled via its *competence*. In particular, at each date t, there exists a function

$$\Gamma^t : \mathcal{G} \to \mathbb{R}$$

designating the competence of each feasible government $G \in \mathcal{G}$ (at that date). We refer to $\Gamma^t_G \in \mathbb{R}$ as government G's competence, with the convention that higher values correspond to greater competence. We can also assume that each individual has a certain level of competence or ability, and the competence of a government is a function of the abilities of its members. For now, this additional assumption is not necessary. Note also that the function Γ^t depends on time. This generality is introduced to allow for changes in the environment (in particular, changes in the relative competences of different individuals and governments).

Individual utilities are determined by the competence of the government that is in power at that date and by whether the individual in question is herself in the government. More specifically, each individual $i \in \mathcal{I}$ at time τ has discounted (expected) utility given by

$$U_i^\tau = \mathbb{E} \sum\nolimits_{t=\tau}^{\infty} \beta^{(t-\tau)} u_i^t, \tag{10.50}$$

where $\beta \in (0,1)$ is the discount factor and u_i^t is individual's stage payoff, given by

$$u_i^t = w_i \left(G^t, \Gamma^t_{G^t}\right) = w_i \left(G^t\right), \tag{10.51}$$

where in the second equality we suppress dependence on $\Gamma^t_{G^t}$ to simplify notation; we will do this throughout unless special emphasis is necessary. Throughout, let us impose the following assumptions on w_i.

Assumption 10.1. *The function w_i satisfies the following properties:*

(1) *for each $i \in \mathcal{I}$ and any $G, H \in \mathcal{G}$ such that $\Gamma^t_G > \Gamma^t_H$: if $i \in G$ or $i \notin H$, then $w_i(G) > w_i(H)$.*
(2) *for any $G, H \in \mathcal{G}$ and any $i \in G \setminus H$, $w_i(G) > w_i(H)$.*

Part 1 of this assumption is a relatively mild restriction on payoffs. It implies that all else equal, more competent governments give higher stage payoff. In particular, if an individual belongs to both governments G and H, and G is more competent than H, then she prefers G. The same conclusion also holds when the individual is not a member of either of these two governments or when she is only a member of G (and not of H). Therefore, this part of the assumption implies that the only situation in which an individual may prefer a less competent government to a more competent one is when she is a member of the former, but not of the latter. This simply captures the presence of rents from holding office or additional income from being in government due to higher salaries or corruption. The interesting interactions in our setup result from the "conflict of interest": individuals prefer to be in the government even when this does not benefit the rest of the society. Part 2 of the assumption strengthens the first part and imposes that this conflict of interest is always present; that is, individuals receive higher payoffs from governments that include them than from those that exclude them (regardless of the competence levels of the two governments). We impose both parts of this assumption throughout. It is important to note that Assumption 10.1 implies that all voters, who are not part of the government, care about a one-dimensional government competence; this feature simplifies the analysis considerably. Nevertheless, the tractability of our framework makes it possible to enrich this environment by allowing other sources of disagreement or conflict of interest among voters, and we return to this issue in the Conclusion.

Example 10.2. *As an example, suppose that the competence of government G, Γ_G, is the amount of public good produced in the economy under feasible government G, and*

$$w_i(G) = v_i(\Gamma_G) + b_i \mathbf{I}_{\{i \in G\}}, \tag{10.52}$$

where $v_i : \mathbb{R} \to \mathbb{R}$ is a strictly increasing function (for each $i \in \mathcal{I}$) corresponding to the utility from public good for individual i, b_i is a measure of the rents that individual i obtains from being in office, and $\mathbf{I}_X$ is the indicator of event X. If $b_i \geq 0$ for each $i \in \mathcal{I}$, then (10.52) satisfies part 1 of Assumption 10.1. In addition, if b_i is sufficiently large for each i, then each individual prefers to be a member of the government, even if this government has a very low level of competence, thus part 2 of Assumption 10.1 is also satisfied.

Finally, the government in power influences the determination of future governments whenever consent of some current government members is necessary for change. We represent the set of individuals (regular citizens and government members) who can, collectively, induce a change in government by specifying the set of *winning coalitions*, $\mathcal{W}_G$, which is a function of current government G (for each $G \in \mathcal{G}$). This is an economical way of summarizing the relevant information, since the set of winning coalitions is precisely the subsets of the society that are able to force (or to block) a change in government. We only impose a minimal amount of structure on the set of winning coalitions.

Assumption 10.2. *For any feasible government $G \in \mathcal{G}$, $\mathcal{W}_G$ is given by*

$$\mathcal{W}_G = \{X \in \mathcal{C} : |X| \geq m_G \text{ and } |X \cap G| \geq l_G\},$$

where l_G and m_G are integers satisfying $0 \leq l_G \leq |G| \leq \bar{k} < m_G \leq n - \bar{k}$ (recall that $\bar{k}$ is the maximal size of the government and n is the size of the society).

The restrictions imposed in Assumption 10.2 are intuitive. In particular, they state that a new government can be instituted if it receives a sufficient number of votes from the entire society (m_G total votes) *and* if it receives support from some subset of the members of the current government (l_G of the current government members need to support such a change). This definition allows l_G to be any number between 0 and $|G|$. One special feature of Assumption 10.2 is that it does not relate the number of veto players in the current government, l_G, to the total number of individuals in the society who wish to change the government, m_G. This aspect of Assumption 10.2 can be relaxed without affecting our general characterization; we return to a discussion of this issue in the Conclusion.

Given this notation, the case where there are no incumbency veto power, $l_G = 0$, can be thought of as *perfect democracy*, where current members of the government have no special power. The case where $l_G = |G|$ can be thought of as *extreme dictatorship*, where unanimity among government members is necessary for any change. Between these extremes are imperfect democracies (or less strict forms of dictatorships), which may arise either because there is some form of (strong or weak) incumbency veto power in democracy or because current government (junta) members are able to block the introduction of a new government. In what follows, one might wish to interpret l_G as an inverse measure of the degree of democracy, though naturally this only captures one dimension of democratic regimes in practice.

Note also that Assumption 10.2 imposes some mild assumptions on m_G. In particular, less than $\bar{k}$ individuals is insufficient for a change to take place. This ensures that a rival government cannot take power without any support from other individuals (recall that $\bar{k}$ denotes the maximum size of the government, so the rival government must have no more than $\bar{k}$ members), and $m_G \leq n - \bar{k}$ individuals are sufficient to implement a change provided that l_G members of the current government are among them. For example, these requirements are naturally met when $\bar{k} < n/2$ and $m_G = \lfloor (n+1)/2 \rfloor$ (i.e., majority rule).

In addition to Assumptions 10.1 and 10.2, we also impose the following *genericity* assumption, which ensures that different governments have different competences. This assumption simplifies the notation and is without much loss of generality, since if it were not satisfied for a society, any small perturbation of competence levels would restore it.

Assumption 10.3. *For any $t \geq 0$ and any $G, H \in \mathcal{G}$ such that $G \neq H$, $\Gamma_G^t \neq \Gamma_H^t$.*

10.4.2. Political Equilibria in Nonstochastic Environments. Let us first focus on nonstochastic environments, where $\Gamma^t = \Gamma$ (or $\Gamma_G^t = \Gamma_G$ for all $G \in \mathcal{G}$). For these environments, let us introduce the equilibrium concept, *(Markov) political equilibrium*, and show that equilibria have a simple recursive characterization. To introduce this equilibrium concept more formally, let us first define the *transition rule* $\phi : \mathcal{G} \to \mathcal{G}$, which maps each feasible government G in power at time t to the government that would emerge in period $t+1$. Given ϕ, we can write the discounted utility implied by (10.50) for each individual $i \in \mathcal{I}$ starting from current government $G \in \mathcal{G}$ recursively as $V_i(G \mid \phi)$, given by

$$V_i(G \mid \phi) = w_i(G) + \beta V_i(\phi(G) \mid \phi) \text{ for all } G \in \mathcal{G}. \tag{10.53}$$

Intuitively, starting from $G \in \mathcal{G}$, individual $i \in \mathcal{I}$ receives a current payoff of $w_i(G)$. Then ϕ (uniquely) determines next period's government $\phi(G)$, and thus the continuation value of this individual, discounted to the current period, is $\beta V_i(\phi(G) \mid \phi)$.

A government G is *stable* given mapping ϕ if $\phi(G) = G$. In addition, we say that ϕ is *acyclic* if for any (possibly infinite) chain $H_1, H_2, \ldots \subset \mathcal{G}$ such that $H_{k+1} \in \phi(H_k)$, and any $a < b < c$, if $H_a = H_c$ then $H_a = H_b = H_c$.

Given (10.53), the next definition introduces the notion of a political equilibrium, which will be represented by the mapping ϕ provided that two conditions are met.

Definition 10.1. *A mapping $\phi : \mathcal{G} \to \mathcal{G}$ is a* (Markov) political equilibrium *if for any $G \in \mathcal{G}$, the following two conditions are satisfied:*

(i) either the set of players who prefer $\phi(G)$ to G (in terms of discounted utility) forms a winning coalition, i.e., $S = \{i \in \mathcal{I} : V_i(\phi(G) \mid \phi) > V_i(G \mid \phi)\} \in \mathcal{W}_G$, (or equivalently $|S| \geq m_G$ and $|S \cap G| \geq l_G$); or else, $\phi(G) = G$;

(ii) there is no alternative government $H \in \mathcal{G}$ that is preferred both to a transition to $\phi(G)$ and to staying in G permanently, i.e., there is no H such that $S'_H = \{i \in \mathcal{I} : V_i(H \mid \phi) > V_i(\phi(G) \mid \phi)\} \in \mathcal{W}_G$ and $S''_H = \{i \in \mathcal{I} : V_i(H \mid \phi) > w_i(G) / (1-\beta)\} \in \mathcal{W}_G$ (alternatively, for any alternative H, either $|S'_H| < m_G$, or $|S'_H \cap G| < l_G$, or $|S''_H| < m_G$, or $|S''_H \cap G| < l_G$).

This definition states that a mapping ϕ is a political equilibrium if it maps the current government G to alternative $\phi(G)$ that (unless it coincides with G) must be preferred to G

(taking continuation values into account) by a sufficient majority of the population and a sufficient number of current government members (so as not to be blocked). Note that in part (i), the set S can be equivalently written as $S = \{i \in \mathcal{I} : V_i (\phi (G) \mid \phi) > w_i (G) / (1 - \beta)\}$, since if this set is not a winning coalition, then $\phi (G) = G$ and thus $V_i (G \mid \phi) = w_i (G) / (1 - \beta)$. Part (ii) of the definition requires that there does not exist another alternative H that would have been a "more preferable" transition; that is, there should be no H that is preferred both to a transition to $\phi (G)$ and to staying in G forever by a sufficient majority of the population and a sufficient number of current government members. The latter condition is imposed, since if there exists a subset H that is preferred to a transition to $\phi (G)$ but not to staying in G forever, then at each stage a move to H can be blocked.

We use the definition of political equilibrium in Definition 10.1 throughout. The advantage of this definition is its simplicity. A disadvantage is that it does not explicitly specify how offers for different types of transitions are made and the exact sequences of events at each stage. However, the paper describes an infinite-horizon extensive-form game, where there is an explicit sequence in which proposals are made, votes are cast and transitions take place. We then characterize the Markov perfect equilibria (MPE) of this dynamic game and show that they are equivalent to political equilibria as defined in Definition 10.1. Briefly, in this extensive-form game, any given government can be either in a sheltered or unstable state. Sheltered governments cannot be challenged, but with some probability become unstable. When the incumbent government is unstable, all individuals (according to a pre-specified order) propose possible alternative governments. Primaries across these governments determine a challenger government, and then a vote between this challenger and the incumbent governments determines whether there is a transition to a new government (depending on whether those in support of the challenger form a winning coalition according to Assumption 10.2). New governments start out as unstable, and with some probability become sheltered. All votes are sequential. We prove that for a sufficiently high discount factor, MPE of this game does not depend on the sequence in which proposals are made, the protocols for primaries or the sequence in which votes are cast, and coincides with political equilibria described by Definition 10.1. This result justifies our focus on the much simpler notion of political equilibrium in the text. The fact that new governments start out as unstable provides a justification for part (ii) of Definition 10.1 that there should not exist another alternative H that is "more preferable" than $\phi (G)$ and than staying in G forever, otherwise there would be an immediate transition to H.

Let us now prove the existence and provide a characterization of political equilibria. We start with a recursive characterization of the mapping ϕ described in Definition 10.1. Let us enumerate the elements of the set $\mathcal{G}$ as $\{G_1, G_2, ..., G_{|\mathcal{G}|}\}$ such that $\Gamma_{G_x} > \Gamma_{G_y}$ whenever $x < y$. With this enumeration, G_1 is the most competent ("best") government, while $G_{|\mathcal{G}|}$ is the least competent government. In view of Assumption 10.3, this enumeration is well defined and unique.

Now, suppose that for some $q > 1$, we have defined ϕ for all G_j with $j < q$. Define the set

$$\mathcal{M}_q \equiv \{j : 1 \leq j < q,\ \{i \in \mathcal{I} : w_i(G_j) > w_i(G_q)\} \in \mathcal{W}_{G_q}, \text{ and } \phi(G_j) = G_j\}. \tag{10.54}$$

Note that this set depends simply on stage payoffs in (10.51), *not* on the discounted utilities defined in (10.53), which are "endogenous" objects. This set can thus be computed easily from the primitives of the model (for each q). Given this set, let the mapping ϕ be

$$\phi(G_q) = \begin{cases} G_q & \text{if } \mathcal{M}_q = \varnothing; \\ G_{\min\{j \in \mathcal{M}_q\}} & \text{if } \mathcal{M}_q \neq \varnothing. \end{cases} \tag{10.55}$$

Since the set $\mathcal{M}_q$ is well defined, the mapping ϕ is also well defined, and by construction it is single valued. Theorems 10.1 and 10.2 next show that, for sufficiently high discount factors, this mapping constitutes the unique acyclic political equilibrium and that, under additional mild conditions, it is also the unique political equilibrium (even considering possible cyclic equilibria).

THEOREM 10.1. *Suppose that Assumptions 10.1-10.3 hold and let $\phi : \mathcal{G} \to \mathcal{G}$ be as defined in* (10.55)*. Then there exists $\beta_0 < 1$ such that for any discount factor $\beta > \beta_0$, ϕ is the unique acyclic political equilibrium.*

Let us now illustrate the intuition for why the mapping ϕ constitutes a political equilibrium. Recall that G_1 is the most competent ("best") government. It is clear that we must have $\phi(G_1) = G_1$, since all members of the population that are not in G_1 will prefer it to any other $G' \in \mathcal{G}$ (from Assumption 10.1). Assumption 10.2 then ensures that there will not be a winning coalition in favor of a permanent move to G'. However, G' may not persist itself, and it may eventually lead to some alternative government $G'' \in \mathcal{G}$. But in this case, we can apply this reasoning to G'' instead of G', and thus the conclusion $\phi(G_1) = G_1$ applies. Next suppose we start with government G_2 in power. The same argument applies if G' is any one of $G_3, G_4, \ldots, G_{|\mathcal{G}|}$. One of these may eventually lead to G_1, thus for sufficiently high discount factors, a sufficient majority of the population may support a transition to such a G' in order to eventually reach G_1. However, discounting also implies that in this case, a sufficient majority would also prefer a direct transition to G_1 to this dynamic path (recall part (ii) of Definition 10.1). So the relevant choice for the society is between G_1 and G_2. In this comparison, G_1 will be preferred if it has sufficiently many supporters, that is, if the set of individuals preferring G_1 to G_2 is a winning coalition within G_2, or more formally if

$$\{i \in \mathcal{I} : w_i(G_1) > w_i(G_2)\} \in \mathcal{W}_{G_2}.$$

If this is the case, $\phi(G_2) = G_1$; otherwise, $\phi(G_2) = G_2$. This is exactly what the function ϕ defined in (10.55) stipulates. Now let us start from government G_3. We then only need to consider the choice between G_1, G_2, and G_3. To move to G_1, it suffices that a winning coalition within G_3 prefers G_1 to G_3. However, whether the society will transition to G_2 depends on the stability of G_2. In particular, we may have a situation in which G_2 is not a stable government, which, by necessity, implies that $\phi(G_2) = G_1$. Then a transition to G_2 will lead to a permanent transition to G_1 in the next period. However, this sequence may be non-desirable for some of those who prefer to move to G_2. In particular, there may exist a winning coalition in G_3 that prefers to stay in G_3 rather than transitioning permanently to G_1 (and as a consequence, there is no winning coalition that prefers such a transition), even though there also exists a winning coalition in G_3 that would have preferred a permanent move to G_2. Writing this more explicitly, we may have

$$\{i \in \mathcal{I} : w_i(G_2) > w_i(G_3)\} \in \mathcal{W}_{G_3},$$

but

$$\{i \in \mathcal{I} : w_i(G_1) > w_i(G_3)\} \notin \mathcal{W}_{G_3}.$$

If so, the transition from G_3 to G_2 may be blocked with the anticipation that it will lead to G_1 which does not receive the support of a winning coalition within G_3. This reasoning illustrates that for a transition to take place, not only should the target government be preferred to the current one by a winning coalition (starting from the current government), but also that the target government should be "stable," i.e., $\phi(G') = G'$. This is exactly the requirement in (10.55). In this light, the intuition for the mapping ϕ and thus for Theorem 10.1 is that a government G will persist in equilibrium (will be stable) if there does not exist another stable government receiving support from a winning coalition (a sufficient majority of the population and the required number of current members of government).

Theorem 10.1 states that ϕ in (10.55) is the unique acyclic political equilibrium. However, it does not rule out cyclic equilibria. Cyclic equilibria are unintuitive and "fragile". We next show that they can also be ruled out under a variety of relatively weak assumptions. The next theorem thus strengthens Theorem 10.1 so that ϕ in (10.55) is the unique equilibrium (among both cyclic and acyclic ones).

THEOREM 10.2. *The mapping* ϕ *defined in* (10.55) *is the unique political equilibrium (and hence in the light of Theorem 10.1, any political equilibrium is acyclic) if any of the following conditions holds:*

(1) *For any* $G \in \mathcal{G}$, $|G| = k$, $l_G = l$ *and* $m_G = m$ *for some* k, l *and* m.
(2) *For any* $G \in \mathcal{G}$, $l_G \geq 1$.
(3) *For any collection of different feasible governments* $H_1, \ldots, H_q \in \mathcal{G}$ *(for* $q \geq 2$*) and for all* $i \in \mathcal{I}$, *we have* $w_i(H_1) \neq \left(\sum_{p=1}^{q} w_i(H_p)\right)/q$.
(4) $\theta > \varepsilon \cdot |\mathcal{G}|$, *where* $\theta \equiv \min_{\{i\in\mathcal{I} \text{ and } G,H\in\mathcal{G}:\ i\in G\setminus H\}} \{w_i(G) - w_i(H)\}$ *and* $\varepsilon \equiv \max_{\{i\in\mathcal{I} \text{ and } G,H\in\mathcal{G}:\ i\in G\cap H\}} \{w_i(G) - w_i(H)\}$.

This theorem states four relatively mild conditions under which there are no cyclic equilibria (thus making ϕ in (10.55) the unique equilibrium). First, if all feasible governments have the same size, k, the same degree of incumbency veto power, l, and the same threshold for the required number of total votes for change, m, then all equilibria must be acyclic and thus ϕ in (10.55) is the unique political equilibrium. Second, the same conclusion applies if we always need the consent of at least one member of the current government for a transition to a new government. These two results imply that cyclic equilibria are only possible if starting from some governments, there is no incumbency veto power and either the degree of incumbency veto power or the vote threshold differs across governments. The third part of the theorem shows that there are also no acyclic political equilibria under a mild restriction on payoffs (which is a slight strengthening of Assumption 10.3 and holds generically, meaning that if it did not hold, a small perturbation of payoff functions would restore it). Finally, the fourth part of the theorem provides a condition on preferences that also rules out cyclic equilibria. In particular, this condition states that if each individual receives sufficiently high utility from being in government (greater than θ) and does not care much about the composition of the rest of the government (the difference in her utility between any two governments including her is always less than ε), then all equilibria must be acyclic.

10.4.3. Nonstochastic Transitions. The most interesting results involve the comparison of different political regimes in terms of their ability to select governments with high levels of competence. To simplify the exposition and focus on the more important interactions, we assume that all feasible governments have the same size, $k \in \mathbb{N}$, where $k < n/2$.

More formally, let us define

$$\mathcal{C}^k = \{Y \in \mathcal{C} : |Y| = k\}.$$

Then, $\mathcal{G} = \mathcal{C}^k$. In addition, we assume that for any $G \in \mathcal{G}$, $l_G = l \in \mathbb{N}$ and $m_G = m \in \mathbb{N}$, so that the set of winning coalitions can be simply expressed as

$$\mathcal{W}_G = \{X \in \mathcal{C} : |X| \geq m \text{ and } |X \cap G| \geq l\}, \tag{10.56}$$

where $0 \leq l \leq k < m \leq n - k$. If $l = 0$, then all individuals have equal weight and there is no incumbency veto power, thus we have a *perfect democracy*. In contrast, if $l > 0$, the consent of some of the members of the government is necessary for a change, thus there is some incumbency veto power. We have thus strengthened Assumption 10.2 to the following.

> **Assumption 2′:** *We have that $\mathcal{G} = \mathcal{C}^k$, and that there exist integers l and m such that the set of winning coalitions is given by (10.56).*

In view of part 1 of Theorem 10.2, Assumption 2′ ensures that the acyclic political equilibrium ϕ given by (10.55) is the unique equilibrium; naturally, we will focus on this equilibrium throughout the rest of the analysis. In addition, given this additional structure, the mapping ϕ can be written in a simpler form. Recall that governments are still ranked according to their level of competence, so that G_1 denotes the most competent government. Then we have:

$$\mathcal{M}_q = \{j : 1 \leq j < q,\ |G_j \cap G_q| \geq l, \text{ and } \phi(G_j) = G_j\}, \tag{10.57}$$

and, as before,

$$\phi(G_q) = \begin{cases} G_q & \text{if } \mathcal{M}_q = \varnothing; \\ G_{\min\{j \in \mathcal{M}_q\}} & \text{if } \mathcal{M}_q \neq \varnothing. \end{cases} \tag{10.58}$$

Naturally, the mapping ϕ is again well defined and unique. Finally, let us also define

$$\mathcal{D} = \{G \in \mathcal{G} : \phi(G) = G\}$$

as the set of stable governments (the fixed points of mapping ϕ). If $G \in \mathcal{D}$, then $\phi(G) = G$, and this government will persist forever if it is the initial government of the society.

We now investigate the structure of stable governments and how it changes as a function of the underlying political institutions, in particular, the extent of incumbency veto power, l. Throughout the remainder of the analysis of this model, we assume that Assumptions 1, 2′ and 3 hold, and we do not add these qualifiers to any of the propositions to economize on space.

Our first proposition provides an important technical result (part 1). It then uses this result to show that perfect democracy ($l = 0$) ensures the emergence of the best (most competent) government, but any departure from perfect democracy destroys this result and enables the emergence of highly incompetent/inefficient governments. It also shows that extreme dictatorship ($l = k$) makes all initial governments stable, regardless of how low their competences may be.

Proposition 10.3. *The set of stable feasible governments $\mathcal{D}$ satisfies the following properties.*

1. *If $G, H \in \mathcal{D}$ and $|G \cap H| \geq l$, then $G = H$. In other words, any two distinct stable governments may have at most $l - 1$ common members.*
2. *Suppose that $l = 0$. Then $\mathcal{D} = \{G_1\}$. In other words, starting from any initial government, the society will transition to the most competent government.*
3. *Suppose $l \geq 1$. Then there are at least two stable governments, i.e., $|\mathcal{D}| \geq 2$. Moreover, the least competent governments may be stable.*
4. *Suppose $l = k$. Then $\mathcal{D} = \mathcal{G}$, so any feasible government is stable.*

Proposition 10.3 shows the fundamental contrast between perfect democracy, where incumbents have no veto power, and other political institutions, which provide some additional power to "insiders" (current members of the government). Perfect democracy leads to the formation of the best government. With any deviation from perfect democracy, there will necessarily exist at least one other stable government (by definition less competent than the best), and even the worst government might be stable. The next example supplements Example 10.1 from the Introduction by showing a richer environment in which the least competent government is stable.

Example 10.3. *Suppose $n = 9$, $k = 3$, $l = 1$, and $m = 5$, so that a change in government requires support from a simple majority of the society, including at least one member of the current government. Suppose $\mathcal{I} = \{1, 2, \ldots, 9\}$, and that stage payoffs are given by (10.52) in Example 10.2. Assume also that $\Gamma_{\{i_1,i_2,i_3\}} = 1000 - 100i_1 - 10i_2 - i_3$ (for $i_1 < i_2 < i_3$). This implies that $\{1, 2, 3\}$ is the most competent government, and is therefore stable. Any other government that includes 1 or 2 or 3 is unstable. For example, the government $\{2, 5, 9\}$ will transit to $\{1, 2, 3\}$, as all individuals except 5 and 9 prefer the latter. However, government $\{4, 5, 6\}$ is stable: any government that is more competent must include 1 or 2 or 3, and therefore is either $\{1, 2, 3\}$ or will immediately transit to $\{1, 2, 3\}$, which means that any such transition will not receive support from any of the members of $\{4, 5, 6\}$. Now, proceeding inductively, we find that any government other than $\{1, 2, 3\}$ and $\{4, 5, 6\}$ that contains at least one individual $1, 2, \ldots, 6$ is unstable. Consequently, government $\{7, 8, 9\}$, which is the least competent government, is stable.*

Proposition 10.3 establishes that under any regime other than perfect democracy, there will necessarily exist stable inefficient/incompetent governments and these may in fact have quite low levels of competences. It does not, however, provide a characterization of when highly incompetent governments will be stable.

We next provide a systematic answer to this question focusing on societies with large numbers of individuals (i.e., n large). Before doing so, we introduce an assumption that will be used in the third part of the next proposition and in later results. In particular, in what follows we will sometimes suppose that each individual $i \in \mathcal{I}$ has a level of ability (or competence) given by $\gamma_i \in \mathbb{R}_+$ and that the competence of the government is a strictly increasing function of the abilities of its members. This is more formally stated in the next assumption.

Assumption 10.4. *Suppose $G \in \mathcal{G}$, and individuals $i, j \in \mathcal{I}$ are such that $i \in G$, $j \notin G$, and $\gamma_i \geq \gamma_j$. Then $\Gamma_G \geq \Gamma_{(G \setminus \{i\}) \cup \{j\}}$.*

The canonical form of the competence function consistent with Assumption 10.4 is

$$\Gamma_G = \sum\nolimits_{i \in G} \gamma_i, \tag{10.59}$$

though for most of our analysis, we do not need to impose this specific functional form.

Assumption 10.4 is useful because it enables us to rank individuals in terms of their "abilities". This ranking is strict, since Assumptions 10.3 and 10.4 together imply that $\gamma_i \neq \gamma_j$ whenever $i \neq j$. When we impose Assumption 10.4, we also enumerate individuals according to their abilities, so that $\gamma_i > \gamma_j$ whenever $i < j$.

The next proposition shows that for societies above a certain threshold of size (as a function of k and l), there always exist stable governments that contain no member of the ideal government and no member of any group of certain pre-specified sizes (thus, no member of groups that would generate a range of potentially high competence governments). Then, under Assumption 10.4, it extends this result, providing a bound on the percentile of the ability distribution such that there exist stable governments that do not include any individuals with competences above this percentile.

Proposition 10.4. *Suppose $l \geq 1$ (and as before, that Assumptions 10.1, 2′, and 10.3 hold).*

(1) *If*

$$n \geq 2k + k\,(k-l)\,\frac{(k-1)!}{(l-1)!\,(k-l)!}, \tag{10.60}$$

then there exists a stable government $G \in \mathcal{D}$ that contains no members of the ideal government G_1.

(2) *Take any $x \in \mathbb{N}$. If*

$$n \geq k + x + x\,(k-l)\,\frac{(k-1)!}{(l-1)!\,(k-l)!}, \tag{10.61}$$

then for any set of individuals X with $|X| \leq x$, there exists a stable government $G \in \mathcal{D}$ such that $X \cap G = \varnothing$ (so no member of set X belongs to G).

(3) *Suppose in addition that Assumption 10.4 holds and let*

$$\rho = \frac{1}{1 + (k-l)\,\frac{(k-1)!}{(l-1)!(k-l)!}}. \tag{10.62}$$

Then there exists a stable government $G \in \mathcal{D}$ that does not include any of the $\lfloor \rho n \rfloor$ highest ability individuals.

Let us provide the intuition for part 1 of Proposition 10.4 when $l = 1$. Recall that G_1 is the most competent government. Let G be the most competent government among those that do not include members of G_1 (such G exists, since $n > 2k$ by assumption). In this case, Proposition 10.4 implies that G is stable, that is, $G \in \mathcal{D}$. The reason is that if $\phi(G) = H \neq G$, then $\Gamma_H > \Gamma_G$, and therefore $H \cap G_1$ contains at least one element by construction of G. But then $\phi(H) = G_1$, as implied by (10.58). Intuitively, if $l = 1$, then once the current government contains a member of the most competent government G_1, this member will consent to (support) a transition to G_1, which will also receive the support of the population at large. She can do so, because G_1 is stable, thus there are no threats that further rounds of transitions will harm her. But then, as in Example 10.1 in the Introduction, G itself becomes stable, because any reform away from G will take us to an unstable government. Part 2 of the proposition has a similar intuition, but it states the stronger result that one can choose any subset of the society with size not exceeding the threshold defined in (10.61) such that there exist stable governments that do not include any member of this subset (which may be taken to include several of the most competent governments). Finally, part 3, which follows immediately from part 2 under Assumption 10.4, further strengthens both parts 1 and 2 of this proposition and also parts 3 and 4 of Proposition 10.3; it shows that there exist stable governments that do not include a certain fraction of the highest ability individuals. Interestingly, this fraction, given in (10.62), is non-monotonic in l, reaching its maximum at $l = k/2$, i.e., for an intermediate level of incumbency veto power. This partly anticipates the results pertaining to the relative successes of different regimes in selecting more competent governments, which we discuss in the next proposition.

Before providing a more systematic analysis of the relationship between political regimes and the quality of governments, we first extend Example 10.1 from the Introduction to show that, starting with the same government, the long-run equilibrium government may be worse when there is less incumbency veto power (as long as we are not in a perfect democracy).

Example 10.4. *Take the setup from Example 10.3 ($n = 9$, $k = 3$, $l = 1$, and $m = 5$), and suppose that the initial government is $\{4, 5, 6\}$. As we showed there, government $\{4, 5, 6\}$ is stable, and will therefore persist. Suppose, however, that $l = 2$ instead. In that case, $\{4, 5, 6\}$ is unstable, and $\phi(\{4, 5, 6\}) = \{1, 4, 5\}$; thus there will be a transition to $\{1, 4, 5\}$. Since $\{1, 4, 5\}$ is more competent than $\{4, 5, 6\}$, this is an example where the long-run equilibrium government is worse under $l = 1$ than under $l = 2$. Note that if $l = 3$, $\{4, 5, 6\}$ would be stable again.*

When either $k = 1$ or $k = 2$, the structure of stable governments is relatively straightforward. (Note that in this proposition, and in the examples that follow, a, b or c denote the indices of individuals, with our ranking that lower-ranked individuals have higher ability; thus $\gamma_a > \gamma_b$ whenever $a < b$.)

Proposition 10.5. *Suppose that Assumptions 10.1, 2′, 10.3 and 10.4 hold.*

(1) *Suppose that $k = 1$. If $l = 0$, then $\phi(G) = \{G_1\} = \{1\}$ for any $G \in \mathcal{G}$. If $l = k = 1$, then $\phi(G) = G$ for any $G \in \mathcal{G}$.*
(2) *Suppose that $k = 2$. If $l = 0$, then $\phi(G) = G_1 = \{1, 2\}$ for any $G \in \mathcal{G}$. If $l = 1$, then if $G = \{a, b\}$ with $a < b$, we have $\phi(G) = \{a - 1, a\}$ when a is even and $\phi(G) = \{a\ , a + 1\}$ when a is odd; in particular, $\phi(G) = G$ if and only if a is odd and $b = a\ + 1$. If $l = 2$, then $\phi(G) = G$ for any $G \in \mathcal{G}$.*

Proposition 10.5, though simple, provides an important insight about the structure of stable governments that will be further exploited below. When $k = 2$ and $l = 1$, the competence of the stable government is determined by the more able of the two members of the initial government. This means that, with rare exceptions, the quality of the initial government will improve to some degree, i.e., typically $\Gamma_{\phi(G)} > \Gamma_G$. However, this increase is generally limited; when $G = \{a, b\}$ with $a < b$, $\phi(G) = \{a - 1, a\}$ or $\phi(G) = \{a, a + 1\}$, so that at best the next highest ability individual is added to the initial government instead of the lower ability member. Therefore, summarizing these three cases, we can say that with a perfect democracy, the best government will arise; with an extreme dictatorship, there will be no improvement in the initial government; and in between this, there will be some limited improvements in the quality of the government.

When $k \geq 3$, the structure of stable governments is more complex, though we can still develop a number of results and insights about the structure of such governments. Naturally, the extremes with $l = 0$ and $l = 3$ are again straightforward. If $l = 1$ and the initial government is $G = \{a, b, c\}$, where $a < b < c$, then we can show that members ranked above $a - 2$ will never become members of the stable government $\phi(G)$, and the most competent member of G, a, is always a member of the stable government $\phi(G)$. Therefore, again with $l = 1$, only incremental improvements in the quality of the initial government are possible. This ceases to be the case when $l = 2$. In this case, it can be shown that whenever $G = \{a, b, c\}$, where $a + b < c$, $\phi(G) \neq G$; instead $\phi(G) = \{a, b, d\}$, where $d < c$ and in fact, $d \ll a$ is possible. This implies a potentially very large improvement in the quality of the government (contrasting with the incremental improvements in the case where $l = 1$). Loosely speaking, the presence of two veto players when $l = 2$ allows the initial government to import very high ability individuals without compromising stability. The next example illustrates this feature, which is at the root of the result highlighted in Example 10.4, whereby lesser incumbency veto power can lead to worse stable governments.

EXAMPLE 10.5. *Suppose $k = 3$, and first take the case where $l = 1$. Suppose $G = \{100, 101, 220\}$, meaning that the initial government consists of individuals ranked 100, 101, and 220 in terms of ability. Then $\phi_{l=1}(G) = \{100, 101, 102\}$ so that the third member of the government is replaced, but the highest and the second highest ability members are not. More generally, recall that only very limited improvements in the quality of the highest ability member are possible in this case. Suppose instead that $l = 2$. Then it can be shown that $\phi_{l=2}(G) = \{1, 100, 101\}$, so that now the stable government includes the most able individual in the society. Naturally if the gaps in ability at the top of the distribution are larger, implying that highest ability individuals have a disproportionate effect on government competence, this feature becomes particularly valuable.*

The following example extends the logic of Example 10.5 to any distribution and shows how expected competence may be higher under $l = 2$ than $l = 1$, and in fact, this result may hold under any distribution over initial (feasible) governments.

EXAMPLE 10.6. *Suppose $k = 3$, and fix a (any) probability distribution over initial governments with full support (i.e., with a positive probability of picking any initial feasible government). Assume that of players $i_1, \ldots, i_n$, the first q (where q is a multiple of 3 and $3 \leq q < n-3$) are "smart," while the rest are "incompetent," so that governments that include at least one of players $i_1, \ldots, i_q$ will have very high competence relative to governments that do not. Moreover, differences in competence among governments that include at least one of* *the players* $i_1, \ldots, i_q$ *and also among those that do not are small relative to the gap between the two groups of governments. Then it can be shown that the expected competence of the stable government* $\phi_{l=2}(G)$ *(under* $l = 2$*) is greater than that of* $\phi_{l=1}(G)$ *(under* $l = 1$*)—both expectations are evaluated according to the probability distribution fixed at the outset. This is intuitive in view of the structure of stable governments under the two political regimes. In particular, if* G *includes at least one of* $i_1, \ldots, i_q$*, so do* $\phi_{l=1}(G)$ *and* $\phi_{l=2}(G)$*. But if* G *does not, then* $\phi_{l=1}(G)$ *will not include them either, whereas* $\phi_{l=2}(G)$ *will include one with positive probability, since the presence of two veto players will allow the incorporation of one of the "smart" players without destabilizing the government.*

Conversely, suppose that Γ_G *is very high if all its players are from* $\{i_1, \ldots, i_q\}$*, and very low otherwise. In that case, the expected competence of* $\phi(G)$ *will be higher under* $l = 1$ *than under* $l = 2$*. Indeed, if* $l = 1$*, the society will end up with a competent government if at least one of the players is from* $\{i_1, \ldots, i_q\}$*, while if* $l = 2$*, because there are now two veto players, there needs to be at least two "smart" players for a competent government to form (though, when* $l = 2$*, this is not sufficient to guarantee the emergence of a competent government either).*

Examples 10.5 and 10.6 illustrate a number of important ideas. With greater incumbency veto power, in these examples with $l = 2$, a greater number of governments near the initial government are stable, and thus there is a higher probability of improvement in the competence of some of the members of the initial government. In contrast, with less incumbency veto power, in these examples with $l = 1$, fewer governments near the initial one are stable, thus incremental improvements are more likely. Consequently, when including a few high ability individuals in the government is very important, regimes with greater incumbency veto power perform better; otherwise, regimes with less incumbency veto power perform better. Another important implication of these examples is that the situations in which regimes with greater incumbency veto power may perform better are not confined to some isolated instances. This feature applies for a broad class of configurations and for expected competences, evaluated by taking uniform or nonuniform distributions over initial feasible governments. Nevertheless, we will see that in stochastic environments, there will be a distinct advantage to political regimes with less incumbency veto power or "greater degrees of democracy," a phenomenon the intuition of which will also be illustrated using Examples 10.5 and 10.6.

10.4.4. Royalty-Type Regimes. The analysis of the persistence of bad governments has so far focused on political institutions that are "junta-like" in the sense that no specific member is indispensable. In such an environment, the incumbency veto power takes the form of the requirement that some members of the current government must consent to change. The alternative is a "royalty-like" environment where one or several members of the government are irreplaceable (i.e., correspond to "individual veto players" in terms of Tsebelis', 2002, terminology). This can be conjectured to be a negative force, since it would mean that a potentially low ability person must always be part of the government. However, because such an irreplaceable member (the member of the "royalty") is also unafraid of changes, better governments may be more likely to arise under certain circumstances, whereas, as we have seen, junta members would resist certain changes because of the further transitions that these will unleash.

Let us change Assumption 10.2 and the structure of the set of winning coalitions $\mathcal{W}_G$ to accommodate royalty-like regimes. We assume that there are l members of the royalty whose votes are *always* necessary for a transition to be implemented (regardless of whether they are current government members). We denote the set of these individuals by Y. So, the new set of winning coalitions becomes

$$\mathcal{W}_G = \{X \in \mathcal{C} : |X| \geq m \text{ and } Y \subset X\}.$$

We also assume that all members of the royalty are in the initial government, that is, $Y \subset G^0$. Note that the interpretation of the parameter l is now different than it was for junta-like regimes. In particular, in junta-like regimes, l measured the incumbency veto power and could be considered as an inverse measure of (one dimension of) the extent of democracy. In contrast, in the case of royalty, $l = 1$ corresponds to a one-person dictatorship, while $l > 1$ could be thought of as a "more participatory" regime.

The next proposition compares royalty-like and junta-like institutions in terms of the expected competence of the equilibrium government, where, as in Example 10.6, the expectation is taken with respect to any full support probability distribution over the composition of the initial government.

PROPOSITION 10.6. *Suppose that we have a royalty-system with* $1 \leq l < k$ *and competences of governments are given by* (10.59)*, so that the* l *royals are never removed from the government. If* $\{a_1, ..., a_n\}$ *is sufficiently "convex" meaning that* $\frac{a_1 - a_2}{a_2 - a_n}$ *is sufficiently large, then the expected competence of the government under the royalty system is greater than under the original, junta-like system (with the same* l*). The opposite conclusion holds if* $\frac{a_1 - a_{n-1}}{a_{n-1} - a_n}$ *is sufficiently low and* $l = 1$.

Proposition 10.6 shows that royalty-like regimes achieve better expected performance than junta-like regimes provided that $\{a_1, ..., a_n\}$ is highly "convex" (such convexity implies that the benefit to society from having the highest ability individual in government is relatively high). As discussed above, juntas are unlikely to lead to such high quality governments because of the fear of a change leading to a further round of changes, excluding all initial members of the junta. Royalty-like regimes avoid this fear. Nevertheless, royalty-like regimes have a disadvantage in that the ability of royals may be very low or may change at some point and become very low, and the royals will always be part of the government. In this sense, royalty-like regimes create a clear disadvantage. However, this result shows that when $\{a_1, ..., a_n\}$ is sufficiently convex (so as to outweigh the loss of expected competence because of the presence of a potentially low ability royal), expected competence is nonetheless higher under the royalty-like system. This result is interesting because it suggests that different types of dictatorships may have distinct implications for long-run quality of government and performance, and regimes that provide security to certain members of the incumbent government may be better at dealing with changes and in ensuring relatively high-quality governments in the long run.

This proposition also highlights that, in contrast to existing results on veto players, a regime with individual veto players (members of royalty) can be less stable and more open to change. In particular, a junta-like regime with $l > 0$ has no individual veto players in the sense of Tsebelis (2002), whereas a royalty-like regime with the same l has such veto players, and Proposition 10.6 shows that the latter can lead to greater change in the composition of the government.

10.4.5. Equilibria in Stochastic Environments. Let us next introduce stochastic shocks to competences of different coalitions (or different individuals) in order to study the flexibility of different political institutions in their ability to adapt the nature and the composition of the government to changes in the underlying environment. Changes in the nature and structure of "high competence" governments may result from changes in the economic, political, or social environment, which may in turn require different types of government to deal with newly emerging problems. Our main results here establish the relationship between the extent of incumbency veto power (one aspect of the degree of democracy) and the flexibility to adapt to changing environments (measured by the probability that the most competent will come to power).

Changes in the environment are modeled succinctly by allowing changes in the function $\Gamma_G^t : \mathcal{G} \to \mathbb{R}$, which determines the competence associated with each feasible government.

Formally, we assume that at each t, with probability $1-\delta$, there is no change in Γ_G^t from Γ_G^{t-1}, and with probability δ, there is a shock and Γ_G^t may change. In particular, following such a shock we assume that there exists a set of distribution functions $F_\Gamma\left(\Gamma_G^t \mid \Gamma_G^{t-1}\right)$ that gives the conditional distribution of Γ_G^t at time t as functions of Γ_G^{t-1}. The characterization of political equilibria in this stochastic environment is a challenging task in general. However, when δ is sufficiently small so that the environment is stochastic but subject to infrequent changes, the structure of equilibria is similar to that in Theorem 10.1. We will exploit this characterization to illustrate the main implications of stochastic shocks on the selection of governments.

We now generalize the definition of (Markov) political equilibrium to this stochastic environment and generalize Theorems 10.1 and 10.2 (for δ small). We then provide a systematic characterization of political transitions in this stochastic environment and illustrate the links between incumbency veto power and institutional flexibility.

The structure of stochastic political equilibria is complicated in general because individuals need to consider the implications of current transitions on future transitions under a variety of scenarios. Nevertheless, when the likelihood of stochastic shocks, δ, is sufficiently small, as we have assumed here, then political equilibria must follow a logic similar to that in Definition 10.1 above. Motivated by this reasoning, we introduce a similar definition of stochastic political equilibria (with infrequent shocks).

To introduce the notion of *(stochastic Markov) political equilibrium*, let us first consider a set of mappings $\phi_{\{\Gamma_G\}}:\mathcal{G}\to\mathcal{G}$ defined as in (10.55), but now separately for each $\{\Gamma_G\}_{G\in\mathcal{G}}$. These mappings are indexed by $\{\Gamma_G\}$ to emphasize this dependence. Essentially, if the configuration of competences of different governments given by $\{\Gamma_G\}_{G\in\mathcal{G}}$ applied forever, we would be in a nonstochastic environment and $\phi_{\{\Gamma_G\}}$ would be the equilibrium transition rule, or simply the political equilibrium, as shown by Theorems 10.1 and 10.2. The idea underlying our definition for this stochastic environment with infrequent changes is that while the current configuration is $\{\Gamma_G\}_{G\in\mathcal{G}}$, $\phi_{\{\Gamma_G\}}$ will still determine equilibrium behavior, because the probability of a change in competences is sufficiently small. When the current configuration is $\{\Gamma_G\}_{G\in\mathcal{G}}$, $\phi_{\{\Gamma_G\}}$ will determine political transitions, and if $\phi_{\{\Gamma_G\}}(G)=G$, then G will remain in power as a stable government. However, when a stochastic shock hits and $\{\Gamma_G\}_{G\in\mathcal{G}}$ changes to $\{\Gamma'_G\}_{G\in\mathcal{G}}$, then political transitions will be determined by the transition rule $\phi_{\{\Gamma'_G\}}$, and unless $\phi_{\{\Gamma'_G\}}(G)=G$, following this shock, there will be a transition to a new government, $G'=\phi_{\{\Gamma'_G\}}(G)$.

Definition 10.2. *Let the set of mappings* $\phi_{\{\Gamma_G\}} : \mathcal{G} \to \mathcal{G}$ *(a separate mapping for each configuration* $\{\Gamma_G\}_{G\in\mathcal{G}}$*) be defined by the following two conditions. When the configuration of competences is given by* $\{\Gamma_G\}_{G\in\mathcal{G}}$*, we have that for any* $G \in \mathcal{G}$*:*

(i) the set of players who prefer $\phi_{\{\Gamma_G\}}(G)$ *to* G *(in terms of discounted utility) forms a winning coalition, i.e.,* $S = \left\{i \in \mathcal{I} : V_i\left(\phi_{\{\Gamma_G\}}(G) \mid \phi_{\{\Gamma_G\}}\right) > V_i\left(G \mid \phi_{\{\Gamma_G\}}\right)\right\} \in \mathcal{W}_G$*;*

(ii) there is no alternative government $H \in \mathcal{G}$ *that is preferred both to a transition to* $\phi_{\{\Gamma_G\}}(G)$ *and to staying in* G *permanently, i.e., there is no* H *such that* $S'_H = \left\{i \in \mathcal{I} : V_i\left(H \mid \phi_{\{\Gamma_G\}}\right) > V_i\left(\phi_{\{\Gamma_G\}}(G) \mid \phi_{\{\Gamma_G\}}\right)\right\} \in \mathcal{W}_G$ *and* $S''_H = \left\{i \in \mathcal{I} : V_i\left(H \mid \phi_{\{\Gamma_G\}}\right) > w_i(G) / (1-\beta)\right\} \in \mathcal{W}_G$ *(alternatively,* $|S'_H| < m_G$*, or* $|S'_H \cap G| < l_G$*, or* $|S''_H| < m_G$*, or* $|S''_H \cap G| < l_G$*).*

Then a set of mappings $\phi_{\{\Gamma_G\}} : \mathcal{G} \to \mathcal{G}$ *constitutes a* (stochastic Markov) political equilibrium *for an environment with sufficiently infrequent changes if there is a transition to government* G_{t+1} *at time* t *(starting with government* G_t*) if and only if* $\left\{\Gamma^t_G\right\}_{G\in\mathcal{G}} = \{\Gamma_G\}_{G\in\mathcal{G}}$ *and* $G_{t+1} = \phi_{\{\Gamma_G\}}(G_t)$*.*

Therefore, a political equilibrium with sufficiently infrequent changes involves the same political transitions (or the stability of governments) as those implied by the mappings $\phi_{\{\Gamma_G\}}$ defined in (10.55), applied separately for each configuration $\{\Gamma_G\}$.

The next theorem provides the general characterization of stochastic political equilibria in environments with sufficiently infrequent changes.

Theorem 10.3. *Suppose that Assumptions 10.1-10.3 hold, and let* $\phi_{\{\Gamma_G\}} : \mathcal{G} \to \mathcal{G}$ *be the mapping defined by* (10.55) *applied separately for each configuration* $\{\Gamma_G\}$*. Then there exist* $\beta_0 < 1$ *and* $\delta_0 > 0$ *such that for any discount factor* $\beta > \beta_0$ *and any positive probability of shocks* $\delta < \delta_0$*,* $\phi_{\{\Gamma_G\}}$ *is a unique acyclic political equilibrium.*

The intuition for this theorem is straightforward. When shocks are sufficiently infrequent, the same calculus that applied in the nonstochastic environment still determines preferences because all agents put most weight on the events that will happen before such a change. Consequently, a stable government will arise and will remain in place until a stochastic shock arrives and changes the configuration of competences. Following such a shock, the stable government for this new configuration of competences emerges. Therefore, Theorem 10.3 provides us with a tractable way of characterizing stochastic transitions. We will next use this result to study the links between different political regimes and institutional flexibility.

With this theorem in place, we can now compare different political regimes in terms of their flexibility (adaptability to stochastic shocks). Our main results will show that, even though limited incumbency veto power does not guarantee the emergence of more competent governments in the nonstochastic environment (nor does it guarantee greater expected competence), it does lead to greater "flexibility" and to better performance according to certain measures in the presence of shocks. In what follows, we always impose Assumptions 1, 2′, 3 and 10.4, which ensure that when the discount factor β is sufficiently large and the frequency of stochastic shocks δ is sufficiently small, there will be a unique (and acyclic) political equilibrium. Propositions 10.7, 10.8, and 10.9 describe properties of this equilibrium.

We also impose some additional structure on the distribution $F_\Gamma\left(\Gamma_G^t \mid \Gamma_G^{t-1}\right)$ by assuming that any shock corresponds to a rearrangement ("permutation") of the abilities of different individuals. Put differently, we assume throughout this subsection that there is a fixed vector of abilities, say $a = \{a_1, ..., a_n\}$, and the actual distribution of abilities across individuals at time t, $\left\{\gamma_j^t\right\}_{j=1}^n$, is given by some permutation φ^t of this vector a. We adopt the convention that $a_1 > a_2 > ... > a_n$. Intuitively, this captures the notion that a shock will change which individual is best placed to solve certain tasks and thus most effective in government functions.

We next characterize the "flexibility" of different political regimes. Throughout the rest of this section, our measure of flexibility is the probability with which the best government will be in power (either at given t or as $t \to \infty$). More formally, let $\pi_t\left(l, k, n \mid G, \{\Gamma_G\}\right)$ be the probability that in a society with n individuals under a political regime characterized by l (for given k), a configuration of competences given by $\{\Gamma_G\}$, and current government $G \in \mathcal{G}$, the most competent government will be in power at the time t. Given n and k, we will think of a regime characterized by l' as more flexible than one characterized by l if $\pi_t\left(l', k, n \mid G, \{\Gamma_G\}\right) > \pi_t\left(l, k, n \mid G, \{\Gamma_G\}\right)$ for all G and $\{\Gamma_G\}$ and for all t following a stochastic shock. Similarly, we can think of the regime as asymptotically more flexible than another, if $\lim_{t\to\infty} \pi_t\left(l', k, n \mid G, \{\Gamma_G\}\right) > \lim_{t\to\infty} \pi_t\left(l, k, n \mid G, \{\Gamma_G\}\right)$ for all G and $\{\Gamma_G\}$ (provided that these limits are well defined). Clearly, "being more flexible" is a partial order.

Proposition 10.7. (1) *If $l = 0$, then a shock immediately leads to the replacement of the current government by the new most competent government.*

(2) *If $l = 1$, the competence of the government following a shock never decreases further; instead, it increases with probability no less than*

$$1 - \frac{(k-1)!\,(n-k)!}{(n-1)!} = 1 - \binom{n-1}{k-1}^{-1}.$$

Starting with any G and $\{\Gamma_G\}$, the probability that the most competent government will ultimately come to power as a result of a shock is

$$\lim_{t\to\infty} \pi_t\left(l,k,n \mid G,\{\Gamma_G\}\right) = \pi\left(l,k,n\right) \equiv 1 - \binom{n-k}{k}\binom{n}{k}^{-1} < 1.$$

For fixed k as $n \to \infty$, $\pi\left(l,k,n\right) \to 0$.

(3) *If $l = k \geq 2$, then a shock never leads to a change in government. The probability that the most competent government is in power at any given period (any t) after the shock is*

$$\pi_t\left(l=k,k,n \mid \cdot,\cdot\right) = \binom{n}{k}^{-1}.$$

This probability is strictly less than $\pi_t\left(l=0,k,n \mid G,\{\Gamma_G\}\right)$ and $\pi_t\left(l=1,k,n \mid G,\{\Gamma_G\}\right)$ for any G and $\{\Gamma_G\}$.

Proposition 10.7 contains a number of important results. A perfect democracy ($l = 0$) does not create any barriers against the installation of the best government at any point in time. Hence, under a perfect democracy every shock is "flexibly" met by a change in government according to the wishes of the population at large (which here means that the most competent government will come to power). As we know from the above analysis, this is no longer true as soon as members of the governments have some veto power. In particular, we know that without stochastic shocks, arbitrarily incompetent governments may come to power and remain in power. However, in the presence of shocks the evolution of equilibrium governments becomes more complex.

Next consider the case with $l \geq 1$. Now, even though the immediate effect of a shock may be a deterioration in government competence, there are forces that increase government competence in the long run. This is most clearly illustrated in the case where $l = 1$. With this set of political institutions, there is zero probability that there will be a further decrease in government competence following a shock. Moreover, there is a positive probability that competence will improve and in fact a positive probability that, following a shock, the most competent government will be instituted. In particular, a shock may make the current government unstable, and in this case, there will be a transition to a new stable government. A transition to a less competent government would never receive support from the population. The change in competences may be such that the only stable government after the shock, starting with the current government, may be the best government. Proposition 10.7 also shows that when political institutions take the form of an extreme dictatorship, there will never be any transition, thus the current government can deteriorate following shocks (in fact, it can do so significantly).

Most importantly, Proposition 10.7, as well as Proposition 10.8 below, show that regimes with intermediate levels of incumbency veto power have a higher degree of *flexibility* than extreme dictatorship, ensuring better long-run outcomes (and naturally perfect democracy has the highest degree of flexibility). This unambiguous ranking in the presence of stochastic shocks (and its stronger version stated in the next proposition) contrasts with the results above, which showed that general comparisons between regimes with different degrees of incumbency veto power (beyond perfect democracy) are not possible in the nonstochastic case.

An informal intuition for the greater flexibility of regimes with more limited incumbency veto power in the presence of stochastic shocks can be obtained from Examples 10.5 and 10.6 above. Recall from these examples that an advantage of the less democratic regime, $l = 2$, is that the presence of two veto players makes a large number of governments near the initial one stable. But this implies that if the initial government is destabilized because of a shock, there will only be a move to a nearby government. In contrast, the more democratic regime, $l = 1$, often makes highly incompetent governments stable because there are no nearby stable governments (recall, for example, part 2 of Proposition 10.5). But this also implies that if a shock destabilizes the current government, a significant improvement in the quality of the government becomes more likely. Thus, at a broad level and contrasting with the presumption in the existing literature, regimes with greater incumbency veto power "create more stability," which facilitates small or moderate-sized improvements in initial government quality; but they do not create a large "basin of attraction" for the most competent government. In contrast, in regimes with less incumbency veto power, low competence governments are often made stable by the instability of nearby alternative governments; this instability can be a disadvantage in deterministic environments as illustrated previously, but turns into a significant flexibility advantage in the presence of stochastic shocks because it creates the possibility that, after a shock, there may be a jump to a very high competence government (in particular, to the best government, which now has a larger "basin of attraction").

The next proposition strengthens the conclusions of Proposition 10.7. In particular, it establishes that the probability of having the most competent government in power is decreasing in l (or in other words, it is increasing in this measure of the "degree of democracy").

Proposition 10.8. *The probability of having the most competent government in power after a shock (for any t),* $\pi_t(l, k, n \mid G, \{\Gamma_G\})$*, is decreasing in l for any k, n, G and* $\{\Gamma_G\}$*.*

Propositions 10.7 and 10.8 highlight a distinct flexibility advantage (in terms of the most competent government coming to power) of regimes with low incumbency veto power ("more democratic" regimes). These results can be strengthened further when shocks are "limited" in the sense that only the abilities of two (or in the second part of the proposition, of $x \geq 2$) individuals in the society are swapped. The next proposition contains these results.

Proposition 10.9. *Suppose that any shock permutes the abilities of x individuals in the society.*

(1) *If $x = 2$ (so that the abilities of two individuals are swapped at a time) and $l \leq k-1$, then the competence of the government in power is nondecreasing over time, that is, $\pi_t(l, k, n \mid G, \{\Gamma_G\})$ is nondecreasing in t for any l, k, n, G and $\{\Gamma_G\}$ such that $l \leq k-1$. Moreover, if the probability of swapping of abilities between any two individuals is positive, then the most competent government will be in power as $t \to \infty$ with probability 1, that is, $\lim_{t\to\infty} \pi_t(l, k, n \mid G, \{\Gamma_G\}) = 1$ (for any l, k, n, G and $\{\Gamma_G\}$ such that $l \leq k-1$).*
(2) *If $x > 2$, then the results in part 1 hold provided that $l \leq k - \lfloor x/2 \rfloor$.*

An interesting application of Proposition 10.9 is that when shocks are (relatively) rare and limited in their scope, relatively democratic regimes will gradually improve over time and install the most competent government in the long run. This is not true for the most autocratic governments, however. This proposition, therefore, strengthens the conclusions of Propositions 10.7 and 10.8 in highlighting the flexibility benefits of more democratic regimes.

10.5. References

(1) Acemoglu, Daron, James A. Robinson and Thierry Verdier (2004) "Kleptocracy and Divide-and-Rule: A Model of Personal Rule," The Alfred Marshall Lecture, *Journal of the European Economic Association Papers and Proceedings,* 2004, 162-192.
(2) Acemoglu, Daron, Georgy Egorov and Konstantin Sonin (2010) "Political Selection and Persistence of Bad Governments," forthcoming *Quarterly Journal of Economics.*
(3) Padro-i-Miguel, Gerard (2007) "The Control of Politicians in Divided Societies: Politics of Fear," *Review of Economic Studies.*

CHAPTER 11

Economic Institutions Under Elite Domination

11.1. Motivation

As noted in the first lecture, at a basic level we can distinguish between of democratic and non-democratic societies. And yet, it is difficult to determine what should really constitute a "democratic" society. Societies that hold rigged elections with non-representative parties would be far from the ideal of a democratic society. Perhaps an equally useful distinction might be between elite-dominated versus non-elite dominated politics.

Based on this distinction, here we start with a model of elite-dominated politics, and turn to "democratic" politics in the next lecture. The purpose of this lecture (and the model presented here) is to develop a simple framework for the determination of institutions and their impact on economic growth under elite domination. Consequently, the basic setup is one in which an existing elite is in control of political power, and uses their monopoly of political power for their own interests even when this is costly for the society at large.

Although it may be more natural to start with democratic politics (especially for those of us used to living in democratic countries), this model of politics under elites domination is highly tractable and gives a number of useful insights, whereas, as we will see in the next lecture, there are difficult unresolved issues about how to model the basic decision-making process under democratic institutions.

Since the key players here are going to be social groups (in particular, the elite, the middle class and citizens), it is useful to think of mapping these broad categories to reality in specific examples. We will discuss some specific examples as we go along, but in the context of economic development during the 19th century, which is important for understanding the current distribution of income across countries, we can think of the elite as landowners, the middle class as new small or medium-scale industrialists, and citizens as workers and peasants. In other contexts, for example, in 20th-century Latin America, however, the elite may correspond to segments of industrialists or bankers with monopoly position. Finally, in some instances, the elite may simply be groups in control of the state apparatus, as in the experience of a number of sub-Saharan African countries after the end of colonialism.

The simple model here will highlight various sources of inefficiencies in policies, which in turn will translate into inefficient (non-growth enhancing) institutions. It should be noted at this point, however, that the concept of inefficiency here is not that of Pareto inefficiency, since when distributional issues are important, Pareto efficiency is not a strong enough concept. An economy in which all of the resources are allocated to a single individual who has no investment opportunities, thus growth is stifled, may nevertheless be Pareto efficient. Thus the concept of "inefficiency" here is being used in the sense of "non-growth enhancing" or "non-surplus maximizing". Nevertheless, throughout the analysis of this simple model, and more generally, throughout this course, it is always useful to bear in mind the benchmark public policies that would apply in a world with a benevolent government (with or without certain restrictions on policy instruments). A comparison of the policies in models with political economy considerations to those with a benevolent planner is useful in delineating the role of political economy in creating distortions.

In the model we will analyze here, the sources of inefficiencies in policies will be:

1. *Revenue extraction:* the group in power—the elite—will set high taxes on middle class producers in order to extract resources from them. These taxes are distortionary. This source of inefficiency results from the absence of non-distortionary taxes, which implies that the distribution of resources cannot be decoupled from efficient production.

2. *Factor price manipulation:* the group in power may want to tax middle class producers in order to reduce the prices of the factors they use in production. This inefficiency arises because the elite and middle class producers compete for factors (here labor). By taxing middle class producers, the elite ensure lower factor prices and thus higher profits for themselves.

3. *Political consolidation:* to the extent that the political power of the middle class depends on their economic resources, greater middle class profits reduce the elite's political power and endanger their future rents. The elite will then want to tax the middle class in order to impoverish them and consolidate their political power.

Although all three inefficiencies in policies arise because of the desire of the elite to extract rents from the rest of the society, the analysis will reveal that of the three sources of inefficiency, the revenue extraction is typically the least harmful, since, in order to extract revenues, the elite need to ensure that the middle class undertakes efficient investments. In contrast, the factor price manipulation and political consolidation mechanisms encourage the elite to directly impoverish the middle class. An interesting comparative static result is that greater state capacity shifts the balance towards the revenue extraction mechanism, and thus, by allowing the elite to extract resources more efficiently from other groups, may improve the allocation of resources.

Additional inefficiencies arise when there are "commitment problems" on the part of the elites, in the sense that they may renege on policy promises once key investments are made. Following the literature on organizational economics, I refer to this as a *holdup problem*. With holdup, taxes are typically higher and more distortionary. Holdup problems, in turn, are likely to be important, for example, when the relevant investment decisions are long-term, so that a range of policies will be decided after these investments are undertaken.

The inefficiencies in policies translate into inefficient institutions. Institutions determine the framework for policy determination, and economic institutions determine both the limits of various redistributive policies and other rules and regulations that affect the economic transactions and productivity of producers. In the context of the simple model here, I associate economic institutions with two features:

(1) limits on taxation and redistribution, and
(2) regulation on the technology used by middle class producers.

The same forces that lead to inefficient policies imply that there will be reasons for the elite to choose inefficient economic institutions. In particular, they may not want to guarantee enforcement of property rights for middle class producers or they may prefer to block technology adoption by middle class producers. Holdup problems, which imply equilibrium taxes even higher than those preferred by the elite, create a possible exception, and may encourage the elite to use economic institutions to place credible limits on their own future policies (taxes). This suggests that economic institutions that (endogenously) restrict future policies may be more likely to arise in economies in which there are more longer-term investments and thus more room for holdup.

The model also sheds light on the conditions under which economic institutions discourage or block technology adoption. If the source of inefficiencies in policies is revenue extraction, the elite always wish to encourage the adoption of the most productive technologies by the middle class. However, when the source of inefficiencies in policies is factor price manipulation or political consolidation, the elite may want to *block* the adoption of more efficient technologies, or at the very least, they would choose not to invest in activities that would increase the productivity of middle class producers. This again reiterates that when the factor price manipulation and political consolidation mechanisms are at work, significantly more inefficient outcomes can emerge.

While economic institutions regulate fiscal policies and technology choices, political institutions govern the process of collective decision-making in society. In the baseline model, the elite have *de jure political power*, which means that they have the formal right to make policy choices and influence economic decisions. To understand the inefficiencies in the institutional framework, we need to investigate *the induced preferences* of different groups over institutions. In the context of political institutions, this means asking whether the elite wish to change the institutional structure towards a more equal distribution of political power. The same forces that make the elite choose inefficient policies also imply that the answer to this question is no. Consequently, despite the inefficiencies that follow, the institutional structure with elite control tends to persist.

The framework also enables us to discuss issues of appropriate and inappropriate institutions. Concentrating political power in the hands of the elite may have limited costs (may even be "efficient"), if the elite are sufficiently productive (more productive than the middle class). However, a change in the productivity of the elite relative to the middle class could make a different distribution of political power more beneficial. In this case, existing institutions, which may have previously functioned relatively well, become inappropriate to the new economic environment. Yet there is no guarantee that there will be a change in institutions in response to the change in environment.

At the end of the lecture, time permitting, I will discuss a framework that models the process of political consolidation of the elite in power in greater detail.

11.2. Baseline Model

11.2.1. The Environment. Consider an infinite horizon economy populated by a continuum $1+\theta_e+\theta_m$ of risk neutral agents, each with a discount factor equal to $\beta<1$. There is a unique non-storable final good denoted by y. The expected utility of agent j at time 0 is given by:

$$U_0^j = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t c_t^j, \tag{11.1}$$

where $c_t^j \in \mathbb{R}$ denotes the consumption of agent j at time t and $\mathbb{E}_t$ is the expectations operator conditional on information available at time t.

Agents are in three groups. The first are workers, whose only action in the model is to supply their labor inelastically. There is a total mass 1 of workers. The second is the elite, denoted by e, who initially hold political power in this society. There is a total of θ^e elites. Finally, there are θ^m "middle class" agents, denoted by m. The sets of elite and middle class producers are denoted by S^e and S^m respectively. With a slight abuse of notation, I will use j to denote either individual or group.

Each member of the elite and middle class has access to production opportunities, represented by the production function

$$y_t^j = \frac{1}{1-\alpha}(A_t^j)^{\alpha}(k_t^j)^{1-\alpha}(l_t^j)^{\alpha}, \tag{11.2}$$

where k denotes capital and l labor. Capital is assumed to depreciate fully after use. The Cobb-Douglas form is adopted for simplicity.

The key difference between the two groups is in their productivity. To start with, let us assume that the productivity of each elite agent is A^e in each period, and that of each middle class agent is A^m. Productivity of the two groups differs, for example, because they are engaged in different economic activities (e.g., agriculture versus manufacturing, old versus new industries, etc.), or because they have different human capital or talent.

On the policy side, we model potentially distortionary (redistributive) taxation as follows: there are activity-specific tax rates on production, τ^e and τ^m, which are constrained to be nonnegative, i.e., $\tau^e \geq 0$ and $\tau^m \geq 0$. There are no other fiscal instruments (in particular, no lump-sum non-distortionary taxes). In addition there is a total income (rent) of R from natural resources. The proceeds of taxes and revenues from natural resources can be redistributed as nonnegative lump-sum transfers targeted towards each group, $T^w \geq 0$, $T^m \geq 0$ and $T^e \geq 0$.

Let us also introduce a parameter $\phi \in [0,1]$, which measures how much of the tax revenue can be redistributed. This parameter, therefore, measures "state capacity," i.e., the ability of the states to penetrate and regulate the production relations in society (though it does so in a highly "reduced-form" way). When $\phi = 0$, state capacity is limited all tax revenue gets lost, whereas when $\phi = 1$ we can think of a society with substantial state capacity that is able to raise taxes and redistribute the proceeds as transfers. The government budget constraint is

$$T_t^w + \theta^m T_t^m + \theta^e T_t^e \leq \phi \int_{j \in S^e \cup S^m} \tau_t^j y_t^j dj + R. \tag{11.3}$$

Let us also assume that there is a maximum scale for each firm, so that $l_t^j \leq \lambda$ for all j and t. This prevents the most productive agents in the economy from employing the entire labor force. Since only workers can be employed, the labor market clearing condition is

$$\int_{j \in S^e \cup S^m} l_t^j dj \leq 1, \tag{11.4}$$

with equality corresponding to full employment. Since $l_t^j \leq \lambda$, (11.4) implies that if

$$\theta^e + \theta^m \leq \frac{1}{\lambda}, \tag{ES}$$

there can never be full employment. Consequently, depending on whether Condition (ES) holds, there will be excess demand or excess supply of labor in this economy. Throughout, I assume that

ASSUMPTION 11.1.

$$\theta^e \leq \frac{1}{\lambda} \text{ and } \theta^m \leq \frac{1}{\lambda},$$

This assumption ensures that neither of the two groups will create excess demand for labor by itself. Assumption 11.1 is adopted only for convenience and simplifies the notation (by reducing the number of cases that need to be studied).

11.2.2. Economic Equilibrium. I first characterize the economic equilibrium for a given sequence of taxes, $\{\tau_t^e, \tau_t^m\}_{t=0,1,\ldots,\infty}$ (the transfers do not affect the economic equilibrium). An *economic equilibrium* is defined as a sequence of wages $\{w_t\}_{t=0,1,\ldots,\infty}$, and investment and employment levels for all producers, $\left\{\left[k_t^j, l_t^j\right]_{j\in S^e\cup S^m}\right\}_{t=0,1,\ldots,\infty}$ such that given $\{\tau_t^e, \tau_t^m\}_{t=0,1,\ldots,\infty}$ and $\{w_t\}_{t=0,1,\ldots,\infty}$, all producers choose their investment and employment optimally and the labor market clears.

Each producer (firm) takes wages, denoted by w_t, as given. Finally, given the absence of adjustment costs and full depreciation of capital, firms simply maximize current net profits. Consequently, the optimization problem of each firm can be written as

$$\max_{k_t^j, l_t^j} \frac{1-\tau_t^j}{1-\alpha}(A^j)^\alpha (k_t^j)^{1-\alpha}\left(l_t^j\right)^\alpha - w_t l_t^j - k_t^j,$$

where $j \in S^e \cup S^m$. This maximization yields

$$k_t^j = (1-\tau_t^j)^{1/\alpha} A^j l_t^j, \text{ and} \tag{11.5}$$

$$l_t^j \begin{cases} = 0 & \text{if } w_t > \frac{\alpha}{1-\alpha}(1-\tau_t^j)^{1/\alpha}A^j \\ \in [0,\lambda] & \text{if } w_t = \frac{\alpha}{1-\alpha}(1-\tau_t^j)^{1/\alpha}A^j \\ = \lambda & \text{if } w_t < \frac{\alpha}{1-\alpha}(1-\tau_t^j)^{1/\alpha}A^j \end{cases}. \tag{11.6}$$

A number of points are worth noting. First, in equation (11.6), the expression $\alpha(1-\tau_t^j)^{1/\alpha}A^j/(1-\alpha)$ is the net marginal product of a worker employed by a producer of group j. If the wage is above this amount, this producer would not employ any workers, and if it is below, he or she would prefer to hire as many workers as possible (i.e., up to the maximum, λ). Second, equation (11.5) highlights the source of potential inefficiency in this economy.

Producers invest in physical capital but only receive a fraction $(1 - \tau_t^j)$ of the revenues. Therefore, taxes discourage investments, creating potential inefficiencies.

Combining (11.6) with (11.4), equilibrium wages are obtained as follows:

(i): If Condition (ES) holds, there is excess supply of labor and $w_t = 0$.

(ii): If Condition (ES) does not hold, then there is "excess demand" for labor and the equilibrium wage is

$$w_t = \min \left\langle \frac{\alpha}{1-\alpha}(1-\tau_t^e)^{1/\alpha} A^e, \frac{\alpha}{1-\alpha}(1-\tau_t^m)^{1/\alpha} A^m \right\rangle. \tag{11.7}$$

The form of the equilibrium wage is intuitive. Labor demand comes from two groups, the elite and middle class producers, and when condition (ES) does not hold, their total labor demand exceeds available labor supply, so the market clearing wage will be the minimum of their net marginal product.

One interesting feature, which will be used below, is that when Condition (ES) does not hold, the equilibrium wage is equal to the net productivity of one of the two groups of producers, so either the elite or the middle class will make zero profits in equilibrium.

Finally, equilibrium level of aggregate output is

$$Y_t = \frac{1}{1-\alpha}(1-\tau_t^e)^{(1-\alpha)/\alpha} A^e \int_{j \in S^e} l_t^j dj + \frac{1}{1-\alpha}(1-\tau_t^m)^{(1-\alpha)/\alpha} A^m \int_{j \in S^m} l_t^j dj + R. \tag{11.8}$$

The equilibrium is summarized in the following proposition:

PROPOSITION 11.1. *Suppose Assumption 11.1 holds. Then for a given sequence of taxes* $\{\tau_t^e, \tau_t^m\}_{t=0,1,\ldots,\infty}$, *the equilibrium takes the following form: if Condition (ES) holds, then* $w_t = 0$, *and if Condition (ES) does not hold, then* w_t *is given by (11.7). Given the wage sequence, factor demands are given by (11.5) and (11.6), and aggregate output is given by (11.8).*

11.2.3. Inefficient Policies. Now I use the above economic environment to illustrate a number of distinct sources of inefficient policies. In this section, political institutions correspond to "the dictatorship of the elite" in the sense that they allow the elite to decide the policies, so the focus will be on the elite's desired policies. The main (potentially inefficient) policy will be a tax on middle class producers, though more generally, this could correspond to expropriation, corruption or entry barriers. As discussed in the introduction, there will be three mechanisms leading to inefficient policies; (1) Revenue Extraction; (2) Factor Price Manipulation; and (3) Political Consolidation.

To illustrate each mechanism in the simplest possible way, I will focus on a subset of the parameter space and abstract from other interactions. Throughout, I assume that there is an upper bound on taxation, so that $\tau_t^m \leq \bar{\tau}$ and $\tau_t^e \leq \bar{\tau}$, where $\bar{\tau} \leq 1$. This limit can be institutional, or may arise because of the ability of producers to hide their output or shift into informal production.

The timing of events within each period is as follows: *first*, taxes are set; *then,* investments are made. This removes an additional source of inefficiency related to the holdup problem whereby groups in power may seize all of the output of other agents in the economy once it has been produced. Holdup will be discussed below.

To start with, I focus on Markov Perfect Equilibria (MPE) of this economy, where strategies are only dependent on payoff-relevant variables. In this context, this means that strategies are independent of past taxes and investments (since there is full depreciation). In the dictatorship of the elite, policies will be chosen to maximize the elite's utility. Hence, a *political equilibrium* is given by a sequence of policies $\{\tau_t^e, \tau_t^m, T_t^w, T_t^m, T_t^e\}_{t=0,1,\dots,\infty}$ (satisfying (11.3)) which maximizes the elite's utility, taking the economic equilibrium as a function of the sequence of policies as given.

More specifically, substituting (11.5) into (11.2), we obtain elite consumption as

$$c_t^e = \left[\frac{\alpha}{1-\alpha}(1-\tau_t^e)^{1/\alpha} A^e - w_t\right] l_t^e + T_t^e, \tag{11.9}$$

with w_t given by (11.7). This expression follows immediately by recalling that the first term in square brackets is the after-tax profits per worker, while the second term is the equilibrium wage. Total per elite consumption is given by their profits plus the lump sum transfer they receive. Then the political equilibrium, starting at time $t = 0$, is simply given by a sequence of $\{\tau_t^e, \tau_t^m, T_t^w, T_t^m, T_t^e\}_{t=0,1,\dots,\infty}$ that satisfies (11.3) and maximizes the discounted utility of the elite, $\sum_{t=0}^{\infty} \beta^t c_t^e$.

The determination of the political equilibrium is simplified further by the fact that in the MPE with full capital depreciation, this problem is simply equivalent to maximizing (11.9). We now characterize this political equilibrium under a number of different scenarios.

11.2.4. Revenue Extraction. To highlight this mechanism, suppose that Condition (ES) holds, so wages are constant at zero. This removes any effect of taxation on factor prices. In this case, from (11.6), we also have $l_t^j = \lambda$ for all producers. Also assume that $\phi > 0$ (for example, $\phi = 1$).

It is straightforward to see that the elite will never tax themselves, so $\tau_t^e = 0$, and will redistribute all of the government revenues to themselves, so $T_t^w = T_t^m = 0$. Consequently taxes will be set in order to maximize tax revenue, given by

$$\text{Revenue}_t = \frac{\phi}{1-\alpha}\tau_t^m (1-\tau_t^m)^{(1-\alpha)/\alpha} A^m \lambda \theta^m + R \tag{11.10}$$

at time t, facedownwhere the first term is obtained by substituting for $l_t^m = \lambda$ and for (11.5) into (11.2) and multiplying it by τ_t^m, and taking into account that there are θ^m middle class producers and a fraction ϕ of tax revenues can be redistributed. The second term is simply the revenues from natural resources. It is clear that tax revenues are maximized by $\tau_t^m = \alpha$. In other words, this is the tax rate that puts the elite at the peak of their Laffer curve. In contrast, output maximization would require $\tau_t^m = 0$. However, the output-maximizing tax rate is not an equilibrium because, despite the distortions, the elite would prefer a higher tax rate to increase their own consumption.

At the root of this inefficiency is a limit on the tax instruments available to the elite. If they could impose lump-sum taxes that would not distort investment, these would be preferable. Inefficient policies here result from the redistributive desires of the elite coupled with the absence of lump-sum taxes.

It is also interesting to note that as α increases, the extent of distortions are reduced, since there are greater diminishing returns to capital and investment will not decline much in response to taxes.

Even though $\tau_t^m = \alpha$ is the most preferred tax for the elite, the exogenous limit on taxation may become binding, so the equilibrium tax is

$$\tau_t^m = \tau^{RE} \equiv \min\{\alpha, \bar{\tau}\} \tag{11.11}$$

for all t. In this case, equilibrium taxes depend only on the production technology (in particular, how distortionary taxes are) and on the exogenous limit on taxation. For example, as α decreases and the production function becomes more linear in capital, equilibrium taxes decline.

This discussion is summarized in the following proposition (proof in the text):

Proposition 11.2. *Suppose Assumption 11.1 and Condition (ES) hold and $\phi > 0$, then the unique political equilibrium features $\tau_t^m = \tau^{RE} \equiv \min\{\alpha, \bar{\tau}\}$ for all t.*

11.2.5. Factor Price Manipulation. I now investigate how inefficient policies can arise in order to manipulate factor prices. To highlight this mechanism in the simplest possible way, let us first assume that $\phi = 0$ so that there are no direct benefits from taxation for the elite. There are indirect benefits, however, because of the effect of taxes on factor prices, which will be present as long as the equilibrium wage is positive. For this reason, I now suppose that Condition (ES) does not hold, so that equilibrium wage is given by (11.7).

Inspection of (11.7) and (11.9) then immediately reveals that the elite prefer high taxes in order to reduce the labor demand from the middle class, and thus wages, as much as possible. The desired tax rate for the elite is thus $\tau_t^m = 1$. Given constraints on taxation, the equilibrium tax is $\tau_t^m = \tau^{FPM} \equiv \bar{\tau}$ for all t. We therefore have:

PROPOSITION 11.3. *Suppose Assumption 11.1 holds, Condition (ES) does not hold, and $\phi = 0$, then the unique political equilibrium features $\tau_t^m = \tau^{FPM} \equiv \bar{\tau}$ for all t.*

This result suggests that the factor price manipulation mechanism generally leads to higher taxes than the pure revenue extraction mechanism. This is because, with the factor price manipulation mechanism, the objective of the elite is to reduce the profitability of the middle class as much as possible, whereas for revenue extraction, the elite would like the middle class to invest and generate revenues. It is also worth noting that, differently from the pure revenue extraction case, the tax policy of the elite is not only extracting resources from the middle class, but it is also doing so indirectly from the workers, whose wages are being reduced because of the tax policy.

The role of $\phi = 0$ also needs to be emphasized. Taxing the middle class at the highest rate is clearly inefficient. Why is there not a more efficient way of transferring resources to the elite? The answer relates to the limited fiscal instruments available to the elite. In particular, $\phi = 0$ implies that they cannot use taxes at all to extract revenues from the middle class, so they are forced to use inefficient means of increasing their consumption, by directly impoverishing the middle class. In the next subsection, I discuss how the factor price manipulation mechanism works in the presence of an instrument that can directly raise revenue from the middle class. This will illustrate that the absence of any means of transferring resources from the middle class to the elite is not essential for the factor price manipulation mechanism (though the absence of non-distortionary lump-sum taxes is naturally important).

An interesting example of the factor manipulation effect comes from Iceland's economy between the 16th and 19th centuries. Though quite prosperous today, during this time period, Iceland was relatively backward, both in terms of income level and technologically. It also experienced frequent famines, the health level of its population was low and there was very little industrialization. Part of the problem was that much of the labor force in Iceland remained in agriculture, while the fishing industry seemed much more promising and profitable. For example, Eggertsson (2005, p. 102) writes: "The central paradox in Iceland's economic history is Icelanders' failure to develop a specialized fishing industry and exploit on a large scale the country's famous fisheries."

The answer for this lack of development of the specialized fishing industry seems to be related to the factor manipulation effect. The political elite in Iceland were the agricultural landlords, and they were opposed to the development of the fisheries. Eggertsson (2005, p. 111) again argues that "... [landlords]... realized that the development of a specialized fishing industry would draw farm workers away, substantially increasing labor costs," and continues "The farm community was conscious of latent upward pressures on labor costs and fought those pressures. When the pull of the fisheries was relatively strong, courts reaffirmed the regulations in the labor market, and authorities tightened enforcement." This is therefore a clear example of how policies to retard development in a potentially productive sector may be useful for the political elites because of their implications on equilibrium (factor) prices.

Another example is the South African economy before the end of Apartheid, where the white elite chose many of its policies, including those related to the location and regulations on the mobility of the black population, in order to ensure low wages for blacks (see, for example, Feinstein, 2005).

11.2.6. Revenue Extraction and Factor Price Manipulation Combined. I now combine the two effects isolated in the previous two subsections. By itself the factor price manipulation effect led to the extreme result that the tax on the middle class should be as high as possible. Revenue extraction, though typically another motive for imposing taxes on the middle class, will serve to reduce the power of the factor price manipulation effect. The reason is that high taxes also reduce the revenues extracted by the elite (moving the economy *beyond the peak* of the Laffer curve), and are costly to the elite.

To characterize the equilibrium in this case again necessitates the maximization of (11.9). This is simply the same as maximizing transfers minus wage bill for each elite producer. As before, transfers are obtained from (11.10), while wages are given by (11.7). When Condition (ES) holds and there is excess supply of labor, wages are equal to zero, and we obtain the same results as in the case of pure revenue extraction.

The interesting case is the one where (ES) does not hold, so that wages are not equal to zero, and are given by the minimum of the two expressions in (11.7). Incorporating the fact that the elite will not tax themselves and will redistribute all the revenues to themselves, the maximization problem can be written as

$$\max_{\tau_t^m} \left[\frac{\alpha}{1-\alpha}A^e - w_t\right] l_t^e + \frac{1}{\theta^e}\left[\frac{\phi}{1-\alpha}\tau_t^m (1-\tau_t^m)^{(1-\alpha)/\alpha} A^m l_t^m \theta^m + R\right], \tag{11.12}$$

subject to (11.7) and

$$\theta^e l_t^e + \theta^m l_t^m = 1, \text{ and} \tag{11.13}$$

$$l_t^m = \lambda \text{ if } \quad (1-\tau_t^m)^{1/\alpha} A^m \geq A^e. \tag{11.14}$$

The first term in (11.12) is the elite's net revenues and the second term is the transfer they receive. Equation (11.13) is the market clearing constraint, while (11.14) ensures that middle class producers employ as much labor as they wish provided that their net productivity is greater than those of elite producers.

The solution to this problem can take two different forms depending on whether (11.14) holds in the solution. If it does, then $w = \alpha A^e / (1-\alpha)$, and elite producers make zero profits and their only income is derived from transfers. Intuitively, this corresponds to the case where the elite prefer to let the middle class producers undertake all of the profitable activities and maximize tax revenues. If, on the other hand, (11.14) does not hold, then the elite generate revenues both from their own production and from taxing the middle class producers. In this case $w = \alpha(1-\tau^m)^{1/\alpha} A^m / (1-\alpha)$. Rather than provide a full taxonomy, I impose the following additional assumption:

Assumption 11.2.

$$A^e \geq \phi(1-\alpha)^{(1-\alpha)/\alpha} A^m \frac{\theta^m}{\theta^e}.$$

This assumption ensures that the solution will always take the latter form (i.e., (11.14) does not hold). Intuitively, this condition makes sure that the productivity gap between the middle class and elite producers is not so large as to make it attractive for the elite to make zero profits themselves (recall that $\phi(1-\alpha)^{(1-\alpha)/\alpha} < 1$, so if $\theta^e = \theta^m$ and $A^e = A^m$, this condition is always satisfied).

Consequently, when Assumption 11.2 holds, we have $w_t = \alpha(1-\tau_t^m)^{1/\alpha} A^m \ \tau_t^m / (1-\alpha)$, and the elite's problem simply boils down to choosing τ_t^m to maximize

$$\frac{1}{\theta^e}\left[\frac{\phi}{1-\alpha}\tau_t^m (1-\tau_t^m)^{(1-\alpha)/\alpha} A^m l^m \theta^m + R\right] - \frac{\alpha}{1-\alpha}(1-\tau_t^m)^{1/\alpha} A^m \lambda, \tag{11.15}$$

where I have used the fact that all elite producers will employ λ employees, and from (11.13), $l_m = (1-\lambda\theta^e)/\theta^m$.

The maximization of (11.15) gives

$$\frac{\tau_t^m}{1-\tau_t^m} = \kappa(\lambda, \theta^e, \alpha, \phi) \equiv \frac{\alpha}{1-\alpha}\left(1 + \frac{\lambda\theta^e}{(1-\lambda\theta^e)\phi}\right).$$

The first interesting feature is that $\kappa(\lambda, \theta^e, \alpha, \phi)$ is always less than ∞. This implies that τ_t^m is always less than 1, which is the desired tax rate in the case of pure factor price manipulation. Moreover, $\kappa(\lambda, \theta^e, \alpha, \phi)$ is strictly greater than $\alpha/(1-\alpha)$, so that τ_t^m is always greater than α, the desired tax rate with pure revenue extraction. Therefore, the factor price manipulation motive always increases taxes above the pure revenue maximizing level (beyond the peak of the Laffer curve), while the revenue maximization motive reduces taxes relative to the pure factor price manipulation case. Naturally, if this level of tax is greater than $\bar{\tau}$, the equilibrium tax will be $\bar{\tau}$, i.e.,

$$\tau_t^m = \tau^{COM} \equiv \min \left\{ \frac{\kappa(\lambda, \theta^e, \alpha, \phi)}{1 + \kappa(\lambda, \theta^e, \alpha, \phi)}, \bar{\tau} \right\}. \tag{11.16}$$

It is also interesting to look at the comparative statics of this tax rate. First, as ϕ increases, taxation becomes more beneficial (generates greater revenues), but τ^{COM} declines. This might at first appear paradoxical, since one may have expected that as taxation becomes less costly, taxes should increase. Intuition for this result follows from the observation that an increase in ϕ raises the importance of revenue extraction, and as commented above, in this case, revenue extraction is a force towards lower taxes (it makes it more costly for the elite to move beyond the peak of the Laffer curve). Since the parameter ϕ is related, among other things, to state capacity, this comparative static result suggests that higher state capacity will translate into lower taxes, because greater state capacity enables the elite to extract revenues from the middle class through taxation, without directly impoverishing them. In other words, greater state capacity enables more efficient forms of revenue extraction by the groups holding political power.

Second, as θ^e increases and the number of elite producers increases, taxes also increase. The reason for this effect is again the interplay between the revenue extraction and factor price manipulation mechanisms. When there are more elite producers, reducing factor prices becomes more important relative to gathering tax revenue. One interesting implication of this discussion is that when the factor price manipulation effect is more important, there will typically be greater inefficiencies. Finally, an increase in α raises taxes for exactly the same reason as above; taxes create fewer distortions and this increases the revenue-maximizing tax rate.

Once again summarizing the analysis:

PROPOSITION 11.4. *Suppose Assumptions 11.1 and 11.2 hold, Condition (ES) does not hold, and $\phi > 0$. Then the unique political equilibrium features $\tau_t^m = \tau^{COM}$ as given by (11.16) for all t. Equilibrium taxes are increasing in θ^e and α and decreasing in ϕ.*

11.2.7. Political Consolidation. I now discuss another reason for inefficient taxation, the desire of the elite to preserve their political power. This mechanism has been absent so far, since the elite were assumed to always remain in power. To illustrate it, the model needs to be modified to allow for endogenous switches of power. Institutional change will be discussed in greater detail later. For now, let us assume that there is a probability p_t in period

t that political power permanently shifts from the elite to the middle class. Once they come to power, the middle class will pursue a policy that maximizes their own utility. When this probability is exogenous, the previous analysis still applies. Interesting economic interactions arise when this probability is endogenous. Here I will use a simple (reduced-form) model to illustrate the trade-offs and assume that this probability is a function of the income level of the middle class agents, in particular

$$p_t = p\left(\theta^m c_t^m\right) \in [0,1], \tag{11.17}$$

where I have used the fact that income is equal to consumption. Let us assume that p is continuous and differentiable with $p' > 0$, which captures the fact that when the middle class producers are richer, they have greater de facto political power. This reduced-form formulation might capture a variety of mechanisms. For example, when the middle class are richer, they may be more successful in solving their collective action problems or they may increase their military power.

This modification implies that the fiscal policy that maximizes current consumption may no longer be optimal. To investigate this issue we now write the utility of elite agents recursively, and denote it by $V^e(E)$ when they are in power and by $V^e(M)$ when the middle class is in power. Naturally, we have

$$V^e(E) = \max_{\tau_t^m} \left\{ \begin{array}{c} \left[\frac{\alpha}{1-\alpha}A^e - w_t\right] l_t^e + \frac{1}{\theta^e}\left[\frac{\phi}{1-\alpha}\tau_t^m(1-\tau_t^m)^{(1-\alpha)/\alpha}A^m l_t^m \theta^m + R\right] \\ +\beta\left[(1-p_t)V^e(E) + p_t V^e(M)\right] \end{array} \right\}$$

subject to (11.7), (11.13), (11.14) and (11.17), with $p_t = p\left(\frac{\alpha}{1-\alpha}(1-\tau_t^m)^{1/\alpha}A^m l_t^m \theta^m - w_t l_t^m \theta^m\right)$. I wrote $V^e(E)$ and $V^e(M)$ not as functions of time, since the structure of the problem makes it clear that these values will be constant in equilibrium.

The first observation is that if the solution to the static problem involves $c_t^m = 0$, then the same fiscal policy is optimal despite the risk of losing power. This implies that, as long as Condition (ES) does not hold and Assumption 11.2 holds, the political consolidation mechanism does not add an additional motive for inefficient taxation.

To see the role of the political consolidation mechanism, suppose instead that Condition (ES) holds. In this case, $w_t = 0$ and the optimal static policy is $\tau_t^m = \tau^{RE} \equiv \min\{\alpha, \bar{\tau}\}$ as discussed above and implies positive profits and consumption for middle class agents. The dynamic maximization problem then becomes

(11.18)

$$V^e(E) = \max_{\tau_t^m} \left\{ \begin{array}{c} \frac{\alpha}{1-\alpha} A^e \lambda + \frac{1}{\theta^e} \left[\frac{\phi}{1-\alpha} \tau_t^m (1-\tau_t^m)^{(1-\alpha)/\alpha} A^m \lambda \theta^m + R \right] \\ +\beta \left[V^e(E) - p\left(\frac{\alpha}{1-\alpha}(1-\tau_t^m)^{1/\alpha} A^m \theta^m \lambda \right) (V^e(E) - V^e(M)) \right] \end{array} \right\}.$$

The first-order condition for an interior solution can be expressed as

$$\phi - \phi \frac{1-\alpha}{\alpha} \frac{\tau_t^m}{1-\tau_t^m} + \beta \theta^e p' \left(\frac{\alpha}{1-\alpha} (1-\tau_t^m)^{1/\alpha} A^m \theta^m \lambda \right) (V^e(E) - V^e(M)) = 0.$$

It is clear that when $p'(\cdot) = 0$, we obtain $\tau_t^m = \tau^{RE} \equiv \min\{\alpha, \bar{\tau}\}$ as above. However, when $p'(\cdot) > 0$, $\tau_t^m = \tau^{PC} > \tau^{RE} \equiv \min\{\alpha, \bar{\tau}\}$ as long as $V^e(E) - V^e(M) > 0$. That $V^e(E) - V^e(M) > 0$ is the case is immediate since when the middle class are in power, they get to tax the elite and receive all of the transfers.

Intuitively, as with the factor price manipulation mechanism, the elite tax *beyond the peak* of the Laffer curve, yet now not to increase their revenues*, but to consolidate their political power*. These high taxes reduce the income of the middle class and their political power. Consequently, there is a higher probability that the elite remain in power in the future, enjoying the benefits of controlling the fiscal policy.

An interesting comparative static is that as R increases, the gap between $V^e(E)$ and $V^e(M)$ increases, and the tax that the elite sets increases as well. Intuitively, the party in power receives the revenues from natural resources, R. When R increases, the elite become more willing to sacrifice tax revenue (by overtaxing the middle class) in order to increase the probability of remaining in power, because remaining in power has now become more valuable. This contrasts with the results so far where R had no effect on taxes. More interestingly, a higher ϕ, i.e., greater state capacity, also increases the gap between $V^e(E)$ and $V^e(M)$ (because this enables the group in power to raise more tax revenues) and thus implies a higher tax rate on the middle class. Intuitively, when there is no political competition, greater state capacity, by allowing more efficient forms of transfers, improves the allocation of resources. But in the presence of political competition, by increasing the *political stakes*, it leads to greater conflict and more distortionary policies.

Summarizing this discussion:

Proposition 11.5. *Consider the economy with political replacement. Suppose also that Assumption 11.1 and Condition (ES) hold and $\phi > 0$, then the political equilibrium features $\tau_t^m = \tau^{PC} > \tau^{RE}$ for all t. This tax rate is increasing in R and ϕ.*

Perhaps the best example of blocking of economic development because of political consolidation is provided by the experiences of Russian and Austria-Hungarian empires during the 19th century. While much of the rest of Europe was industrializing, the political elites in these countries actively blocked development, for example, by preventing the building of railways, because they feared technological and institutional change that would be brought about by industrialization. An important factor in both Russia and Austria-Hungary was that land/labor relations were still feudal or quasi-feudal well and many of the remnants of the absolutist monarchy still persisted well into the 19th century, increasing the political stakes for the landed aristocracy, encouraging them to oppose industrialization to protect their rents.

11.2.8. Subgame Perfect Versus Markov Perfect Equilibria. I have so far focused on Markov perfect equilibria (MPE). In general, such a focus can be restrictive. In this case, however, it can be proved that subgame perfect equilibria (SPE) coincide with the MPE. This will not be true in the next subsection, so it is useful to briefly discuss why it is the case here.

MPE are a subset of the SPE. Loosely speaking, SPEs that are not Markovian will be supported by some type of "history-dependent punishment strategies". If there is no room for such history dependence, SPEs will coincide with the MPEs.

In the models analyzed so far, such punishment strategies are not possible even in the SPE. Intuitively, each individual is infinitesimal and makes its economic decisions to maximize profits. Therefore, (11.5) and (11.6) determine the factor demands uniquely in any equilibrium. Given the factor demands, the payoffs from various policy sequences are also uniquely pinned down. This means that the returns to various strategies for the elite are *independent of history*. Consequently, there cannot be any SPEs other than the MPE characterized above. Therefore, we have:

Proposition 11.6. *The MPEs characterized in Propositions 11.2-11.5 are the unique SPEs.*

11.2.9. Lack of Commitment—Holdup. The models discussed so far featured full commitment to taxes by the elites. Using a term from organizational economics, this corresponds to the situation without any "*holdup*". Holdup (lack of commitment to taxes or policies) changes the qualitative implications of the model; if expropriation (or taxation) happens after investments, revenues generated by investments can be *ex post* captured by others. These types of holdup problems are likely to arise when the key investments are long-term, so that various policies will be determined and implemented after these investments are made (and sunk).

The problem with holdup is that the elite will be unable to commit to a particular tax rate before middle class producers undertake their investments (taxes will be set after investments). This lack of commitment will generally increase the amount of taxation and inefficiency. To illustrate this possibility, I consider the same model as above, but change the timing of events such that first individual producers undertake their investments and then the elite set taxes. The economic equilibrium is unchanged, and in particular, (11.5) and (11.6) still determine factor demands, with the only difference that τ^m and τ^e now refer to "expected" taxes. Naturally, in equilibrium expected and actual taxes coincide.

What is different is the calculus of the elite in setting taxes. Previously, they took into account that higher taxes would discourage investment. Since, now, taxes are set after investment decisions, this effect is absent. As a result, in the MPE, the elite will always want to tax at the maximum rate, so in all cases, there is a unique MPE where $\tau_t^m = \tau^{HP} \equiv \bar{\tau}$ for all t. This establishes (proof in the text):

Proposition 11.7. *With holdup, there is a unique political equilibrium with $\tau_t^m = \tau^{HP} \equiv \bar{\tau}$ for all t.*

It is clear that this holdup equilibrium is more inefficient than the equilibria characterized above. For example, imagine a situation in which Condition (ES) holds so that with the original timing of events (without holdup), the equilibrium tax rate is $\tau_t^m = \alpha$. Consider the extreme case where $\bar{\tau} = 1$. Now without holdup, $\tau_t^m = \alpha$ and there is positive economic activity by the middle class producers. In contrast, with holdup, the equilibrium tax is $\tau_t^m = 1$ and the middle class stop producing. This is naturally very costly for the elite as well since they lose all their tax revenues.

In this model, it is no longer true that the MPE is the only SPE, since there is room for an implicit agreement between different groups whereby the elite (credibly) promise a different tax rate than $\bar{\tau}$. To illustrate this, consider the example where Condition (ES) holds and $\bar{\tau} = 1$. Recall that the history of the game is the complete set of actions taken up to that point. In the MPE, the elite raise no tax revenue from the middle class producers. Instead, consider the following trigger-strategy combination: the elite always set $\tau^m = \alpha$ and the middle class producers invest according to (11.5) with $\tau^m = \alpha$ as long as the history consists of $\tau^m = \alpha$ and investments have been consistent with (11.5). If there is any other action in the history, the elite set $\tau^m = 1$ and the middle class producers invest zero. With this strategy profile, the elite raise a tax revenue of $\phi\alpha(1-\alpha)^{(1-\alpha)/\alpha} A^m \lambda \theta^m / (1-\alpha)$ in every period, and receive transfers worth

$$\frac{\phi}{(1-\beta)(1-\alpha)}\alpha(1-\alpha)^{(1-\alpha)/\alpha}A^m\lambda\theta^m. \tag{11.19}$$

If, in contrast, they deviate at any point, the most profitable deviation for them is to set $\tau^m = 1$, and they will raise

$$\frac{\phi}{1-\alpha}(1-\alpha)^{(1-\alpha)/\alpha}A^m\lambda\theta^m. \tag{11.20}$$

The trigger-strategy profile will be an equilibrium as long as (11.19) is greater than or equal to (11.20), which requires $\beta \geq 1-\alpha$. Therefore we have:

Proposition 11.8. *Consider the holdup game, and suppose that Assumption 11.1 and Condition (ES) hold and $\bar{\tau} = 1$. Then for $\beta \geq 1-\alpha$, there exists a subgame perfect equilibrium where $\tau_t^m = \alpha$ for all t.*

An important implication of this result is that in societies where there are greater holdup problems, for example, because typical investments involve longer horizons, there is room for coordinating on a subgame perfect equilibrium supported by an implicit agreement (trigger strategy profile) between the elite and the rest of the society.

11.2.10. Technology Adoption and Holdup. Suppose now that taxes are set before investments, so the source of holdup in the previous subsection is absent. Instead, suppose that at time $t = 0$ before any economic decisions or policy choices are made, middle class agents can invest to increase their productivity. In particular, suppose that there is a cost $\Gamma(A^m)$ of investing in productivity A^m. The function Γ is non-negative, continuously differentiable and convex. This investment is made once and the resulting productivity A^m applies forever after.

Once investments in technology are made, the game proceeds as before. Since investments in technology are sunk after date $t = 0$, the equilibrium allocations are the same as in the results presented above. Another interesting question is whether, if they could, the elite would prefer to commit to a tax rate sequence at time $t = 0$.

The analysis of this case follows closely that of the baseline model, and I simply state the results (without proofs to save space):

Proposition 11.9. *Consider the game with technology adoption and suppose that Assumption 11.1 holds, Condition (ES) does not hold, and $\phi = 0$, then the unique political equilibrium features $\tau_t^m = \tau^{FPM} \equiv \bar{\tau}$ for all t. Moreover, if the elite could commit to a tax sequence at time $t = 0$, then they would still choose $\tau_t^m = \tau^{FPM} \equiv \bar{\tau}$.*

That this is the unique MPE is quite straightforward. It is also intuitive that it is the unique SPE. In fact, the elite would choose exactly this tax rate even if they could commit at time $t = 0$. The reason is as follows: in the case of pure factor price manipulation, the only objective of the elite is to reduce the middle class' labor demand, so they have no interest in increasing the productivity of middle class producers.

For contrast, let us next consider the pure revenue extraction case with Condition (ES) satisfied. Once again, the MPE is identical to before. As a result, the first-order condition for an interior solution to the middle class producers' technology choice is:

$$\Gamma'(A^m) = \frac{1}{1-\beta}\frac{\alpha}{1-\alpha}(1-\tau^m)^{1/\alpha} \tag{11.21}$$

where τ^m is the constant tax rate that they will face in all future periods. In the pure revenue extraction case, recall that the equilibrium is $\tau^m = \tau^{RE} \equiv \min\{\alpha, \bar{\tau}\}$. With the same arguments as before, this is also the unique SPE. Once the middle class producers have made their technology decisions, there is no history-dependent action left, and it is impossible to create history-dependent punishment strategies to support a tax rate different than the static optimum for the elite. Nevertheless, this is not necessarily the allocation that the elite prefer. If the elite could commit to a tax rate sequence at time $t = 0$, they would choose lower taxes. To illustrate this, suppose that they can commit to a constant tax rate (it is straightforward to show that they will in fact choose a constant tax rate even without this restriction, but this restriction saves on notation). Therefore, the optimization problem of the elite is to maximize tax revenues taking the relationship between taxes and technology as in (11.21) as given. In other words, they will solve: $\max \phi\tau^m(1-\tau^m)^{(1-\alpha)/\alpha}A^m\lambda\theta^m/(1-\alpha)$ subject to (11.21). The constraint (11.21) incorporates the fact that (expected) taxes affect technology choice.

The first-order condition for an interior solution can be expressed as

$$A^m - \frac{1-\alpha}{\alpha}\frac{\tau^m}{1-\tau^m}A^m + \tau^m\frac{dA^m}{d\tau^m} = 0$$

where $dA^m/d\tau^m$ takes into account the effect of future taxes on technology choice at time $t = 0$. This expression can be obtained from (11.21) as:

$$\frac{dA^m}{d\tau^m} = -\frac{1}{1-\beta}\frac{1}{1-\alpha}\frac{(1-\tau^m)^{(1-\alpha)/\alpha}}{\Gamma''(A^m)} < 0.$$

This implies that the solution to this maximization problem satisfies $\tau^m = \tau^{TA} < \tau^{RE} \equiv \min\{\alpha, \bar{\tau}\}$. If they could, the elite would like to commit to a lower tax rate in the future in order to encourage the middle class producers to undertake technological improvements.

Their inability to commit to such a tax policy leads to greater inefficiency than in the case without technology adoption. Summarizing this discussion:

Proposition 11.10. *Consider the game with technology adoption, and suppose that Assumption 11.1 and Condition (ES) hold and $\phi > 0$, then the unique political equilibrium features $\tau_t^m = \tau^{RE} \equiv \min\{\alpha, \bar{\tau}\}$ for all t. If the elite could commit to a tax policy at time $t = 0$, they would prefer to commit to $\tau^{TA} < \tau^{RE}$.*

An important feature is that in contrast to the pure holdup problem where SPE could prevent the additional inefficiency (when $\beta \geq 1 - \alpha$, recall Proposition 11.8), with the technology adoption game, the inefficiency survives the SPE. The reason is that, since middle class producers invest only once at the beginning, there is no possibility of using history-dependent punishment strategies. This illustrates the limits of implicit agreements to keep tax rates low. Such agreements not only require a high discount factor ($\beta \geq 1 - \alpha$), but also frequent investments by the middle class, so that there is a credible threat against the elite if they deviate from the promised policies. When such implicit agreements fail to prevent the most inefficient policies, there is greater need for economic institutions to play the role of placing limits on future policies.

11.3. Inefficient Economic Institutions

The previous analysis shows how inefficient policies emerge out of the desire of the elite, which possesses political power, to redistribute resources towards themselves. I now discuss the implications of these mechanisms for inefficient institutions. Since the elite prefer to implement inefficient policies to transfer resources from the rest of the society (the middle class and the workers) to themselves, they will also prefer inefficient economic institutions that enable and support these inefficient policies.

To illustrate the main economic interactions, I consider two prototypical economic institutions: (1) *Security of property rights;* there may be constitutional or other limits on the extent of redistributive taxation and/or other policies that reduce profitability of producers' investments. In terms of the model above, we can think of this as determining the level of $\bar{\tau}$. (2) *Regulation of technology*, which concerns direct or indirect factors affecting the productivity of producers, in particular middle class producers.

As pointed out in the introduction, the main role of institutions is to provide the framework for the determination of policies, and consequently, preferences over institutions are derived from preferences over policies and economic allocations. Bearing this in mind, let us now discuss the determination of economic institutions in the model presented here. To simplify the discussion, for the rest of the analysis, and in particular, throughout this section, I focus on MPE, and start with security of property rights.

The environment is the same as in the previous section, with the only difference that at time $t = 0$, before any decisions are taken, the elite can reduce $\bar{\tau}$, say from $\bar{\tau}^H$ to some level in the interval $[0, \bar{\tau}^H]$, thus creating an upper bound on taxes and providing greater security of property rights to the middle class. The key question is whether the elite would like to do so, i.e., whether they prefer $\bar{\tau} = \bar{\tau}^H$ or $\bar{\tau} < \bar{\tau}^H$

The next three propositions answer this question:

Proposition 11.11. *Without holdup and technology adoption, the elite prefer $\bar{\tau} = \bar{\tau}^H$.*

The proof of this result is immediate, since without holdup or technology adoption, putting further restrictions on the taxes can only reduce the elite's utility. This proposition implies that if economic institutions are decided by the elite (which is the natural benchmark since they are the group with political power), they will in general choose not to provide additional security of property rights to other producers. Therefore, the underlying economic institutions will support the inefficient policies discussed above.

The results are different when there are holdup concerns. To illustrate this, suppose that the timing of taxation decision is after the investment decisions (so that there is the holdup problem), and consider the case with revenue extraction and factor price manipulation combined. In this case, the elite would like to commit to a lower tax rate than $\bar{\tau}^H$ in order to encourage the middle class to undertake greater investments, and this creates a useful role for economic institutions (to limit future taxes):

Proposition 11.12. *Consider the game with holdup and suppose Assumptions 11.1 and 11.2 hold, Condition (ES) does not hold, and $\phi > 0$, then as long as τ^{COM} given by (11.16) is less than $\bar{\tau}^H$, the elite prefer $\bar{\tau} = \tau^{COM}$.*

The proof is again immediate. While τ^{COM} maximizes the elite's utility, in the presence of holdup the MPE involves $\tau = \bar{\tau}^H$, and the elite can benefit by using economic institutions to manipulate equilibrium taxes.

This result shows that the elite may provide additional property rights protection to producers in the presence of holdup problems. The reason is that because of holdup, equilibrium taxes are too high even relative to those that the elite would prefer. By manipulating economic institutions, the elite may approach their desired policy (in fact, it can exactly commit to the tax rate that maximizes their utility).

Finally, for similar reasons, in the economy with technology adoption discussed above, the elite will again prefer to change economic institutions to restrict future taxes:

PROPOSITION 11.13. *Consider the game with holdup and technology adoption, and suppose that Assumption 11.1 and Condition (ES) hold and* $\phi > 0$*, then as long as* $\tau^{TA} < \bar{\tau}^{H}$*, the elite prefer* $\bar{\tau} = \tau^{TA}$*.*

As before, when we look at SPE, with pure holdup, there may not be a need for changing economic institutions, since credible implicit promises might play the same role (as long as $\beta \geq 1 - \alpha$ as shown in Proposition 11.8). However, parallel to the results above, in the technology adoption game, SPE and MPE coincide, so a change in economic institutions is necessary for a credible commitment to a low tax rate (here τ^{TA}).

Turning to the regulation of technology now, we see that economic institutions also have and major effect on the environment for technology adoption or more directly the technology choices of producers. For example, by providing infrastructure or protection of intellectual property rights, a society may improve the technology available to its producers. Conversely, the elite may want to *block*, i.e., take active actions against, the technological improvements of the middle class. Therefore the question is: do the elite have an interest in increasing the productivity of the middle class as much as possible?

Consider the baseline model. Suppose that there exists a government policy $g \in \{0, 1\}$, which influences only the productivity of middle class producers, i.e., $A^m = A^m(g)$, with $A^m(1) > A^m(0)$. Assume that the choice of g is made at $t = 0$ before any other decisions, and has no other influence on payoffs (and in particular, it imposes no costs on the elite). Will the elite always choose $g = 1$, increasing the middle class producers' productivity, or will they try to block technology adoption by the middle class?

When the only mechanism at work is revenue extraction, the answer is that the elite would like the middle class to have the best technology:

PROPOSITION 11.14. *Suppose Assumption 11.1 and Condition (ES) hold and* $\phi > 0$*, then* $w = 0$ *and the the elite always choose* $g = 1$*.*

The proof follows immediately since $g = 1$ increases the tax revenues and has no other effect on the elite's consumption. Consequently, in this case, the elite would like the producers to be as productive as possible, so that they generate greater tax revenues. Intuitively, there is no competition between the elite and the middle class (either in factor markets or in the political arena), and when the middle class is more productive, the elite generate greater tax revenues.

The situation is different when the elite wish to manipulate factor prices:

Proposition 11.15. *Suppose Assumption 11.1 holds, Condition (ES) does not hold, $\phi = 0$, and $\bar{\tau} < 1$, then the elite choose $g = 0$.*

Once again the proof of this proposition is straightforward. With $\bar{\tau} < 1$, labor demand from the middle class is high enough to generate positive equilibrium wages. Since $\phi = 0$, taxes raise no revenues for the elite, and their only objective is to reduce the labor demand from the middle class and wages as much as possible. This makes $g = 0$ the preferred policy for the elite. Consequently, the factor price manipulation mechanism suggests that, when it is within their power, the elite will choose economic institutions so as to reduce the productivity of competing (middle class) producers.

The next proposition shows that a similar effect is in operation when the political power of the elite is in contention.

Proposition 11.16. *Consider the economy with political replacement. Suppose also that Assumption 11.1 and Condition (ES) hold and $\phi = 0$, then the elite prefer $g = 0$.*

In this case, the elite cannot raise any taxes from the middle class since $\phi = 0$. But differently from the previous proposition, there are no labor market interactions, since there is excess labor supply and wages are equal to zero. Nevertheless, the elite would like the profits from middle class producers to be as low as possible so as to consolidate their political power. They achieve this by creating an environment that reduces the productivity of middle class producers.

Overall, this section has demonstrated how the elite's preferences over policies, and in particular their desire to set inefficient policies, translate into preferences over inefficient—non-growth enhancing—economic institutions. When there are no holdup problems, introducing economic institutions that limit taxation or put other constraints on policies provides no benefits to the elite. However, when the elite are unable to commit to future taxes (because of holdup problems), equilibrium taxes may be too high even from the viewpoint of the elite, and in this case, using economic institutions to manipulate future taxes may be beneficial. Similarly, the analysis reveals that the elite may want to use economic institutions to discourage productivity improvements by the middle class. Interestingly, this never happens when the main mechanism leading to inefficient policies is revenue extraction. Instead, when factor price manipulation and political consolidation effects are present, the elite may want to discourage or block technological improvements by the middle class.

11.4. Modeling Political Institutions

The above analysis characterized the equilibrium under "the dictatorship of the elite," a set of political institutions that gave all political power to the elite producers. An alternative is to have "the dictatorship of the middle class," i.e., a system in which the middle class makes the key policy decisions (this could also be a democratic regime with the middle class as the decisive voters). Finally, another possibility is democracy in which there is voting over different policy combinations. If $\theta^e + \theta^m < 1$, then the majority are the workers, and they will pursue policies to maximize their own income.

I now briefly discuss the possibility of a switch from the dictatorship of the elite to one of these two alternative regimes. It is clear that whether the dictatorship of the elite or that of middle class is more efficient depends on the relative numbers and productivities of the two groups, and whether elite control or democracy is more efficient depends on policies in democracy. Hence, this section will first characterize the equilibrium under these alternative political institutions. Moreover, for part of the analysis in this subsection, I simplify the discussion by imposing the following assumption:

Assumption 11.3.

$$\theta^m = \theta^e < \frac{1}{2},$$

This assumption ensures that the number of middle class and elite producers is the same, and they are in the minority relative to workers.

11.4.1. Dictatorship of the Middle Class. With the dictatorship of the middle class, the political equilibrium is identical to the dictatorship of the elite, with the roles reversed. To avoid repetition, I will not provide a full analysis. Instead, let me focus on the case, combining revenue extraction and factor price manipulation. The analog of Assumption 11.2 in this case is:

Assumption 11.4.

$$A^m \geq \phi(1-\alpha)^{(1-\alpha)/\alpha} A^e \frac{\theta^e}{\theta^m}.$$

Given this assumption, a similar proposition to that above immediately follows; the middle class will tax the elite and will redistribute the proceeds to themselves, i.e., $T_t^w = T_t^e = 0$, and moreover, the same analysis as above gives their most preferred tax rate as

$$\tau_t^e = \tilde{\tau}^{COM} \equiv \min\left\{\frac{\kappa(\lambda, \theta^m, \alpha, \phi)}{1+\kappa(\lambda, \theta^m, \alpha, \phi)}, \bar{\tau}\right\}. \tag{11.22}$$

PROPOSITION 11.17. *Suppose Assumptions 11.1 and 11.3 hold, Condition (ES) does not hold, and* $\phi > 0$*, then the unique political equilibrium with middle class control features* $\tau_t^e = \tilde{\tau}^{COM}$ *as given by (11.22) for all* t.

Comparing this equilibrium to the equilibrium under the dictatorship of the elite, it is apparent that the elite equilibrium will be more efficient when A^e and θ^e are large relative to A^m and θ^m, and the middle class equilibrium will be more efficient when the opposite is the case.

PROPOSITION 11.18. *Suppose Assumptions 11.1-11.4 hold, then aggregate output is higher with the dictatorship of the elite than the dictatorship of the middle class if* $A^e > A^m$ *and it is higher under the dictatorship of the middle class if* $A^m > A^e$.

Intuitively, the group in power imposes taxes on the other group (and since $\theta^m = \theta^e$, these taxes are equal) and not on themselves, so aggregate output is higher when the group with greater productivity is in power and is spared from distortionary taxation.

11.4.2. Democracy. Under Assumption (A4), workers are in the majority in democracy, and have the power to tax the elite and the middle class to redistribute themselves. More specifically, each worker's consumption is $c_t^w = w_t + T_t^w$, with w_t given by (11.7), so that workers care about equilibrium wages and transfers. Workers will then choose the sequence of policies $\{\tau_t^e, \tau_t^m, T_t^w, T_t^m, T_t^e\}_{t=0,1,\ldots,\infty}$ that satisfy (11.3) to maximize $\sum_{t=0}^{\infty} \beta^t c_t^w$.

It is straightforward to see that the workers will always set $T_t^m = T_t^e = 0$. Substituting for the transfers from (11.3), we obtain that democracy will solve the following maximization problem to determine policies:

$$\max_{\tau_t^e, \tau_t^m} w_t + \frac{\phi}{1-\alpha} \left[\tau_t^m (1-\tau_t^m)^{(1-\alpha)/\alpha} A^m l^m \theta^m + \tau_t^e (1-\tau_t^e)^{(1-\alpha)/\alpha} A^e l^e \theta^e \right] + R$$

with w_t given by (11.7).

As before, when Condition (ES) holds, taxes have no effect on wages, so the workers will tax at the revenue maximizing rate, similar to the case of revenue extraction for the elite above. This result is stated in the next proposition (proof omitted):

PROPOSITION 11.19. *Suppose Assumption 11.1 and Condition (ES) hold and* $\phi > 0$*, then the unique political equilibrium with democracy features* $\tau_t^m = \tau_t^e = \tau^{RE} \equiv \min\{\alpha, \bar{\tau}\}$.

Therefore, in this case democracy is more inefficient than both middle class and elite control, since it imposes taxes on both groups. The same is not the case, however, when Condition (ES) does not hold and wages are positive. In this case, workers realize that by taxing the marginal group they are reducing their own wages. In fact, taxes always reduce wages more than the revenue they generate because of their distortionary effects. As a result, workers will only tax the group with the higher marginal productivity. More specifically, for example, if $A^m > A^e$, we will have $\tau_t^e = 0$, and τ_t^m will be such that $(1 - \tau_t^m)^{1/\alpha} A^m = A^e$ or $\tau_t^m = \alpha$ and $(1 - \alpha)^{1/\alpha} A^m \geq A^e$. Therefore, we have:

Proposition 11.20. *Suppose Assumptions 11.1 and 11.4 hold and Condition (ES) does not hold. Then in the unique political equilibrium with democracy, if $A^m > A^e$, we will have $\tau_t^e = 0$, and $\tau_t^m = \tau^{Dm}$ will be such that $(1 - \tau^{Dm})^{1/\alpha} A^m = A^e$ or $\tau^{Dm} = \alpha$ and $(1 - \alpha)^{1/\alpha} A^m \geq A^e$. If $A^m < A^e$, we will have $\tau_t^m = 0$, and $\tau_t^e = \tau^{De}$ will be such that $(1 - \tau^{De})^{1/\alpha} A^e = A^m$ or $\tau^{De} = \alpha$ and $(1 - \alpha)^{1/\alpha} A^e \geq A^e$.*

The most interesting implication of this proposition comes from the comparison of the case with and without excess supply. While in the presence of excess labor supply, democracy taxes both groups of producers and consequently generates more inefficiency than the dictatorship of the elite or the middle class, when there is no excess supply, it is in general less distortionary than the dictatorship of the middle class or the elite. The intuition is that when Condition (ES) does not hold, workers understand that high taxes will depress wages and are therefore less willing to use distortionary taxes.

11.4.3. Inefficiency of Political Institutions and Inappropriate Institutions. Consider a society where Assumption 11.4 is satisfied and $A^e < A^m$ so that middle class control is more productive (i.e., generates greater output). Despite this, the elite will have no incentive, without some type of compensation, to relinquish their power to the middle class. In this case, political institutions that lead to more inefficient policies will persist even though alternative political institutions leading to better outcomes exist.

One possibility is a Coasian deal between the elite and the middle class. For example, perhaps the elite can relinquish political power and get compensated in return. However, such deals are in general not possible. To discuss why (and why not), let us distinguish between two alternative approaches.

First, the elite may relinquish power in return for a promise of future transfers. This type of solution will run into two difficulties. (i) such promises will not be credible, and once they have political power, the middle class will have no incentive to keep on making such transfers. (ii) since there are no other, less distortionary, fiscal instruments, to compensate the elite, the middle class will have to impose similar taxes on itself, so that the alternative political institutions will not be as efficient in the first place.

Second, the elite may relinquish power in return for a lump-sum transfer from the middle class. Such a solution is also not possible in general, since the net present value of the benefit of holding political power often exceeds any transfer that can be made. Consequently, the desire of the elite to implement inefficient policies also implies that they support political institutions that enable them to pursue these policies. Thus, in the same way as preferences over inefficient policies translate into preferences over inefficient economic institutions, they also lead to preferences towards inefficient political institutions. We will discuss how political institutions can change from the "ground-up" in later parts of the class.

Another interesting question is whether a given set of economic institutions might be "appropriate" for a while, but then become "inappropriate" and costly for economic activity later. This question might be motivated, for example, by the contrast of the Northeastern United States and the Caribbean colonies between the 17th and 19th centuries. The Caribbean colonies were clear examples of societies controlled by a narrow elite, with political power in the monopoly of plantation owners, and few rights for the slaves that made up the majority of the population. In contrast, Northeastern United States developed as a settler colony, approximating a democratic society with significant political power in the hands of smallholders and a broader set of producers. While in both the 17th and 18th centuries, the Caribbean societies were among the richest places in the world, and almost certainly richer and more productive than the Northeastern United States, starting in the late 18th century, they lagged behind the United States and many other more democratic societies, which took advantage of new investment opportunities, particularly in industry and commerce. This raises the question as to whether the same political and economic institutions that encouraged the planters to invest and generate high output in the 17th and early 18th centuries then became a barrier to further growth.

The baseline model used above suggests a simple explanation along these lines. Imagine an economy in which the elite are in power, Condition (ES) does not hold, ϕ is small, A^e is relatively high and A^m is relatively small to start with. The above analysis shows that the elite will choose a high tax rate on the middle class. Nevertheless, output will be relatively high, because the elite will undertake the right investments themselves, and the distortion on the middle class will be relatively small since A^m is small.

Consequently, the dictatorship of the elite may generate greater income per capita than an alternative society under the dictatorship of the middle class. This is reminiscent of the planter elite controlling the economy in the Caribbean.

However, if at some point the environment changes so that A^m increases substantially relative to A^e, the situation changes radically. The elite, still in power, will continue to impose high taxes on the middle class, but now these policies have become very costly because they distort the investments of the more productive group. Another society where the middle class have political power will now generate significantly greater output.

This simple example illustrates how institutions that were initially "appropriate" (i.e., that did not generate much distortion or may have even encouraged growth) later caused the society to fall substantially behind other economies.

11.5. Further Modeling of Political Consolidation

The above model showed how the political consolidation mechanism could be an important factor in leading to inefficient policies and institutions by creating an incentive for the elites to block the introduction of new technologies. But the modeling of political consolidation mechanism was very reduced form. The role of blocking of new technologies in retarding economic development is also emphasized by Parente and Prescott (1999), but their mechanism is similar to the simple "economic losers" idea discussed in the first texture, and the modeling is also very reduced form. Ideally, we would like to go beyond this, and have a more micro-based model of how economic progress affects the distribution of political power in society. Unfortunately, currently there is no satisfactory model of this. Nevertheless, I now present a somewhat less reduced-form model than the one above, which highlights how technological change can be blocked because it creates turbulence, and under some circumstances (but not always) increases the probability that the incumbent elite will be replaced by a new group. One advantage of this model will be that, relative to the more reduced-form model of political consolidation presented above, it will make more specific predictions. Most importantly, it will show that blocking of economic development (or of new technologies) is most likely when political competition is at intermediate levels. When political competition is very high, rulers that block new technologies cannot survive. When political competition is very low, there is no threat to the political power of the existing elite (or ruler), so they do not need to take distortionary actions to protect their political power. It is only when incumbents (the elites) are secure enough to be able to block technological change and still survive, but insecure enough that they fear change that blocking of new technologies and economic development is most likely.

11.5.1. Environment. Consider again an infinite horizon economy in discrete time consisting of a group of citizens, with mass normalized to 1, an incumbent ruler, and an infinite stream of potential new rulers. All agents are infinitely lived, maximize the net present discounted value of their income and discount the future with discount factor, β. While citizens are infinitely lived, an incumbent ruler may be replaced by a new ruler, and from then on receives no utility.

There is only one good in this economy, and each agent produces:

$$y_t = A_t,$$

where A_t is the state of "technology" at time t.

When there is beneficial ecological change, A increases to αA, where $\alpha > 1$. The cost of such change in is normalized to 0. In addition, if there is political change and the incumbent ruler is replaced, this also affects the output potential of the economy as captured by A. In particular, when the incumbent does not adopt a new technology, the "cost of political change"—that is, the cost of replacing the incumbent—is zA, while this cost is $z'A$ when there is technological innovation.

Therefore, more formally

$$A_t = A_{t-1}\left[1 + \left((1-p_t)\,x_t + p_t\widehat{x}_t\right)(\alpha - 1) - p_t x_t z' - p_t\,(1-x_t)\,z\right], \tag{11.23}$$

where $x_t = 1$ or 0 denotes whether the new technology is introduced ($x_t = 1$) or not ($x_t = 0$) at time t by the incumbent ruler, while $\widehat{x}_t = 1$ or 0 refers to the innovation decision of a new ruler. Also, $p_t = 1$ denotes that the incumbent is replaced, while $p_t = 0$ applies when the incumbent is kept in place. Therefore, the relevant decision is x_t when $p_t = 0$ and $\widehat{x}_t$ when $p_t = 1$.

The important assumption is that the extent of costs of political change depend on whether there is technological change. When there is innovation by the incumbent, i.e., $x_t = 1$, the cost of replacing the ruler (which corresponds to $p_t = 1$) is a different random variable, z' than the case in which there is no innovation by the incumbent (i.e., $x_t = 0$). This captures the notion that without innovation, the position of the incumbent is relatively secure, and it will be more costly to replace him. With innovation, there is political uncertainty and *turbulence*, and part of the advantages of the incumbent are eroded. As a result, the cost of replacing the incumbent may be lower.

More explicitly, assume that z and z' are random variables, enabling stochastic changes in rulers along the equilibrium path. z is drawn from the distribution F^N and z' is drawn from F^I, which is first-order stochastically dominated by F^N, capturing the notion that technological change, or changing general, erodes part of the incumbency advantage of the initial ruler.

To simplify the algebra, assume that F^I is uniform over $\left[\mu - \frac{1}{2}, \mu + \frac{1}{2}\right]$, while F^N is uniform over $\left[\gamma\mu - \frac{1}{2}, \gamma\mu + \frac{1}{2}\right]$, where $\gamma \geq 1$. In this formulation, μ is an inverse measure of the degree of political competition: when μ is low, incumbents have little advantage, and when μ is high, it is costly to replace the incumbent.

Note that μ can be less than $\frac{1}{2}$, and in fact, we will focus much of the discussion on the case in which $\mu < \frac{1}{2}$. This implies that sometimes it may be less costly to replace the incumbent ruler than keeping him in place (i.e., the "cost" of replacing the incumbent may be negative). The case of $\mu = 0$ is of particular interest, since it implies that there is no incumbency advantage, and z is symmetric around zero.

On the other hand, γ is a measure of how much the incumbency advantage is eroded by the introduction of a new technology: when $\gamma = 1$, the costs of replacing the ruler are identical irrespective of whether there is technological change or not. A new entrant becomes the incumbent ruler in the following period after he takes control, and it will, in turn, be costly to replace him.

Citizens replace the ruler if a new ruler provides them with higher utility.

Finally, rulers levy a tax T on citizens. We assume that when the technology is A, citizens have access to a non-taxable informal technology that produces $(1 - \tau)A$. This implies that it will never be optimal for rulers to impose a tax greater than τ.

The timing of events within the period is

(1) The period starts with A_t.
(2) The incumbent decides whether to undertake technological change, $x_t = 0$ or 1.
(3) The stochastic costs of replacement, z_t or z_t', are revealed.
(4) Citizens decide whether to replace the ruler, p_t.
(5) If they replace the ruler, a new ruler comes the power and decides whether to initiate technological change $\widehat{x}_t = 0$ or 1.
(6) The ruler in power decides the level of the tax rate, T_t.

11.5.2. The Socially-Optimal Allocation. First consider the innovation decisions that would be taken by an output-maximizing social planner.

This can be done by writing the end-of-period Bellman equation for the social planner, $S(A)$. (evaluated in after step 6 in the timing of events above). By standard arguments, this value function can be written as:

(11.24)

$$S(A) = A +$$

$$\beta \left[x^S \int \left[\begin{array}{c} \left(1-p_I^S(z')\right) S(\alpha A) \\ +p_I^S(z') \left(\widehat{x}^S S((\alpha - z')\, A) + (1-\widehat{x}^S) S((1-z')\, A)\right) \end{array} \right] dF^I + \right.$$

$$\left. (1-x^S) \int \left[\begin{array}{c} \left(1-p_N^S(z)\right) S(A) \\ +p_N^S(z) \left(\widehat{x}^S S((\alpha - z)\, A) + (1-\widehat{x}^S) S((1-z)\, A)\right) \end{array} \right] dF^N \right]$$

where x^S denotes whether the social planner dictates that the incumbent innovates while $\widehat{x}^S$ denotes the social planner's decision of whether to undertake the innovation with a new ruler (after replacing the incumbent). $p_I^S(z') \in \{0,1\}$ denotes whether the planner decides to replace an incumbent who has innovated when the realization of the cost of replacement is z', while $p_N^S(z) \in \{0,1\}$ is the decision to keep an incumbent who has not innovated as a function of the realization z.

Intuitively, when technology is given by A, the total output of the economy is A, and the continuation value depends on the innovation and the replacement decisions. If $x^S = 1$, the social planner induces the incumbent to innovate, and the social value when he is not replaced is $S(\alpha A)$. When the planner decides to replace the incumbent, then there is a new ruler and the social planner decides if he will change institutions, $\widehat{x}^S$. In this case, conditional on the cost realization, z', the social value is $S((\alpha - z')\, A)$ or $S((1-z')\, A)$ depending on whether the new technology is adopted. Notice that if $\widehat{x}^S = 1$ and the newcomer innovates, this affects the output potential of the economy immediately, hence the term $(\alpha - z')\, A$. The second line of (11.24) is explained similarly following a decision by the planner not to innovate. The important point in this case is that the cost of replacement is drawn from the distribution F^N not from F^I.

By standard arguments, $S\,(A)$ is strictly increasing in A. This immediately implies that $S((\alpha - z')\, A) > S((1-z')\, A)$ since $\alpha > 1$, so the planner will always choose $\widehat{x}^S = 1$.

The same reasoning implies that the social planner would like to replace an incumbent who has innovated when $S((\alpha - z')\, A) > S(\alpha A)$, i.e., when $z' < 0$. Similarly, she would like to replace an incumbent who has not innovated when $S((\alpha - z)\, A) > S(A)$, i.e., when $z < \alpha - 1$. Substituting for these decision rules in (11.24), the decision to innovate or not boils down to a comparison of

$$\text{Value from innovating} = \left(\int_0^{\mu+\frac{1}{2}} S(\alpha A)\, dz'\right) + \left(\int_{\mu-\frac{1}{2}}^{0} S\left(\left(\alpha - z'\right) A\right) dz'\right)$$

and

$$\text{Value from not innovating} = \left(\int_{\alpha-1}^{\gamma\mu+\frac{1}{2}} S(A)\, dz\right) + \left(\int_{\gamma\mu-\frac{1}{2}}^{\alpha-1} S\left(\left(\alpha - z\right) A\right) dz\right)$$

Inspection shows that that the first expression is always greater. Therefore, the social planner will always adopt the new technology or initiate the necessary ecological change. Intuitively, the society receives two benefits from innovating: first, output is higher, and second the expected cost of replacing the incumbent, if necessary, is lower. Both of these benefits imply that the social planner always strictly prefers to undertake technological change.

11.5.3. Equilibrium and Blocking of Progress. Next considered the Markov Perfect Equilibria (MPE) of this repeated game.

The strategy of the incumbent in each stage game is simply a technology adoption decision, $x \in [0,1]$, and a tax rate $T \in [0,1]$ when in power, the strategy of a new entrant is also similarly, an action, $\widehat{x} \in [0,1]$ and a tax rate $\widehat{T}$.

The strategy of the citizens consists of a replacement rule, $p(x, z, z') \in [0,1]$, with $p = 1$ corresponding to replacing the incumbent. The action of citizens is conditioned on x, because they move following the innovation decision by the incumbent. At this point, they observe z, which is relevant to their payoff, if $x = 0$, and z', if $x = 1$.

An MPE of this game consists of a strategy combination $\left\{x, T, \widehat{x}, \widehat{T}, p(x, z, z')\right\}$, such that all these actions are best responses to each other for all values of the state A.

Denote the end-of-period value function of citizens by $V(A)$ (once again this is evaluated after the innovation decisions, i.e., after step 6 in the timing of events), so A includes the improvement due to technological change or the losses due to turbulence and political change during this period. With a similar reasoning to the social planner's problem:

(11.25)

$$V(A) = A(1-T) +$$

$$\beta \left[x \int \left[\begin{array}{c} (1-p_I(z'))\,V(\alpha A) \\ +p_I(z')\,(\widehat{x}V((\alpha - z')\,A) + (1-\widehat{x})V((1-z')\,A)) \end{array} \right] dF^I + \right.$$

$$\left. (1-x) \int \left[\begin{array}{c} (1-p_N(z))\,V(A) \\ +p_N(z)\,(\widehat{x}V((\alpha - z)\,A) + (1-\widehat{x})V((1-z)\,A)) \end{array} \right] dF^N \right]$$

where $p_I(z')$ and $p_N(z)$ denote the decisions of the citizens to replace the incumbent as a function of his innovation decision and the cost realization.

Intuitively, the citizens produce A and pay a tax of TA. The next two lines of (11.25) give the continuation value of the citizens. This depends on whether the incumbent innovates or not, $x = 1$ or $x = 0$, and on the realization of the cost of replacing the incumbent. For example, following $x = 1$, citizens observe z', and decide whether to keep the incumbent. If they do not replace the incumbent, $p_I(z') = 0$, then there is no cost, and the value to the citizens is $V(\alpha A)$. In contrast, if they decide $p_I(z') = 1$, that is, they replace the incumbent, then the value is $V((\alpha - z')\,A)$ when the newcomer innovates, and $V((1-z')\,A)$ when he doesn't. The third line is explained similarly as the expected continuation value following a decision not to innovate by the incumbent.

The end-of-period value function for a ruler (again evaluated after step 6 in the timing of the game, so once he knows that he is in power) can be written as

$$W(A) = TA + \beta \left[\begin{array}{c} x \int (1-p_I(z'))\,W(\alpha A) dF^I \\ +(1-x) \int (1-p_N(z))\,W(A) dF^N \end{array} \right]. \tag{11.26}$$

The ruler receives tax revenue of TA, and receives a continuation value which depends on his innovation decisions x. This continuation value also depends on the draw z' or z, indirectly through the replacement decisions of the citizens, $p_I(z')$ and $p_N(z)$.

Standard arguments immediately imply that the value of the ruler is strictly increasing in T and A. Since, by construction, in an MPE the continuation value does not depend on T, the ruler will choose the maximum tax rate $T = \tau$.

Next, consider the innovation decision of a new ruler. Here, the decision boils down to the comparison of $W((1-z)\,A)$ and $W((\alpha - z)\,A)$. Now the strict monotonicity of (11.26) in A and the fact that $\alpha > 1$ imply that $\widehat{x} = 1$ is a dominant strategy for the entrants.

The citizens' decision of whether or not to replace the incumbent ruler is also simple. Again by standard arguments $V\,(A)$ is strictly increasing in A. Therefore, citizens will replace the incumbent ruler whenever $V(A) < V(A')$ where A is the output potential under the incumbent ruler and A' is the output potential under the newcomer.

Now consider a ruler who has innovated and drawn a cost of replacement z'. If citizens keep him in power, they will receive $V(\alpha A)$. If they replace him, taking into account that the new ruler will innovate, they will receive $V((\alpha - z')A)$. Then, their best response is:

$$p_I(z') = 0 \text{ if } z' \geq 0 \text{ and } p_I(z') = 1 \text{ if } z' < 0. \tag{11.27}$$

Next, following a decision not to innovate by the incumbent, citizens compare the value $V(A)$ from keeping the incumbent to the value of replacing the incumbent and having the new technology, $V((\alpha - z)A)$. So

$$p_N(z) = 0 \text{ if } z \geq \alpha - 1 \text{ and } p_N(z) = 0 \text{ if } z < \alpha - 1. \tag{11.28}$$

Finally, the incumbent will decide whether to innovate by comparing the continuation values. Using the decision rules of the citizens, the return to innovating is

$$\int_{\mu-\frac{1}{2}}^{\mu+\frac{1}{2}} \left(1 - p_I(z')\right) \cdot W(\alpha A) dF^I,$$

and the value to not innovating is given by the expression

$$\int_{\gamma\mu-\frac{1}{2}}^{\gamma\mu+\frac{1}{2}} \left(1 - p_N(z)\right) \cdot W(A) dF^N.$$

Now incorporating the decision rules (11.27) and (11.28), and exploiting the uniformity of the distribution function F^I gives the value of innovating as

$$\text{Value from innovating} = \left[1 - F^I(0)\right] W(\alpha A) = P\left[\frac{1}{2} + \mu\right] W(\alpha A) \tag{11.29}$$

where the function P is defined as follows: $P[h] = 0$ if $h < 0$, $P[h] = h$ if $h \in [0,1]$, and $P[h] = 1$ if $h > 1$, making sure that the first term is a cumulative probability (i.e., it does not become negative or greater than 1). Similarly, the value to the ruler of not innovating is

$$\begin{aligned} \text{Value from not innovating} &= \left[1 - F^N(\alpha - 1)\right] W(A) \\ &= P\left[\frac{1}{2} + \gamma\mu - (\alpha - 1)\right] W(A), \end{aligned} \tag{11.30}$$

which differs from (11.29) for two reasons: the probability of replacement is different, and the value conditional on no-replacement is lower.

It is straightforward to see that if $P\left[\frac{1}{2} + \gamma\mu - (\alpha - 1)\right] < P\left[\frac{1}{2} + \mu\right]$, so that the probability of replacement is higher after no-innovation than innovation, the ruler will always innovate—by innovating, he is increasing both his chances of staying in power and his returns conditional on staying in power. Therefore, there will only be blocking of technological change when

$$P\left[\frac{1}{2}+\gamma\mu-(\alpha-1)\right]>P\left[\frac{1}{2}+\mu\right], \tag{11.31}$$

i.e., when by innovating, the ruler creates "turbulence," which increases his chances of being replaced.

For future reference, define $\bar{\gamma}$ such that (11.31) holds only when $\gamma>\bar{\gamma}$. Therefore, as long as $\gamma\leq\bar{\gamma}$, there will be no blocking of innovation.

To fully characterize the equilibrium, conjecture that both value functions are linear, $V(A)=v(x)A$ and $W(A)=w(x)A$. The parameters $v(x)$ and $w(x)$ are conditioned on x, since the exact form of the value function will depend on whether there is innovation or not. Note however that $w(x)$ and $r(x)$ are simply parameters, independent of the state variable, A.

The condition for the incumbent to innovate, that is, for (11.29) to be greater than (11.30), is:

$$\begin{aligned} w(x)\,\alpha AP\left[\frac{1}{2}+\mu\right] &\geq w(x)\,AP\left[\frac{1}{2}+\gamma\mu-(\alpha-1)\right] \\ &\Longleftrightarrow \\ \alpha P\left[\frac{1}{2}+\mu\right] &\geq P\left[\frac{1}{2}+\gamma\mu-(\alpha-1)\right]. \end{aligned} \tag{11.32}$$

When will the incumbent embraced technological change? First, consider the case $\mu=0$, where there is no incumbency advantage (i.e., the cost of replacing the incumbent is symmetric around 0). In this case, there is "fierce" competition between the incumbent and the rival. Condition (11.32) then becomes $\alpha P\left[\frac{1}{2}\right]>P\left[\frac{1}{2}-(\alpha-1)\right]$, which is always satisfied since $\alpha>1$. Therefore, when $\mu=0$, the incumbent will always innovate, i.e., $x=1$. By continuity, for μ low enough, the incumbent will always innovate.

Intuitively, because the rival is as good as the incumbent, and citizens prefer better technology, they are quite likely to replace an incumbent who does not innovate. As a result, the incumbent innovates in order to increase his chances of staying in power. The more general implication of this result is that incumbents facing fierce political competition, with little incumbency advantage, are likely to innovate because they realize that if they do not innovate they will be replaced.

Next, consider the polar opposite case where $\mu\geq 1/2$, that is, there is a very high degree of incumbency advantage. In this situation $P\left[\mu+\frac{1}{2}\right]=1\geq P\left[\frac{1}{2}+\gamma\mu-(\alpha-1)\right]$, so there is no advantage from not innovating because the incumbent is highly *entrenched* and cannot lose power. This establishes that highly entrenched incumbents will also not block technological change.

The situation is different however when $\mu \in \left(0, \frac{1}{2}\right)$. Inspection of condition (11.32) shows that for μ small and γ large, incumbents will prefer not to innovate. This is because of the *political replacement effect* in the case where $\gamma > \bar{\gamma}$: technological change increases the likelihood that the incumbent will be replaced, effectively eroding his political rents (notice that this is the opposite of the situation with $\mu = 0$ when the incumbent innovated in order to increase his chances of staying in power).

As a result the incumbent may prefer not to innovate in order to increase the probability that he maintains power. The reasoning is similar to the replacement effect in industrial organization emphasized by Arrow (1962): incumbents are less willing to innovate than entrants since they will be partly replacing their own rents. Here this replacement refers to the political rents that the incumbent is destroying by increasing the likelihood that he will be replaced.

To determine the parameter region where blocking happens, note that there can only be blocking when both $P\left[\frac{1}{2} + \mu\right]$ and $P\left[\frac{1}{2} + \gamma\mu - (\alpha - 1)\right]$ are between 0 and 1, hence respectively equal to $\frac{1}{2} + \mu$ and $\frac{1}{2} + \gamma\mu - (\alpha - 1)$. Then from (11.32), there will be blocking when

$$\gamma > \alpha + \frac{3}{2}\frac{\alpha - 1}{\mu}. \tag{11.33}$$

Hence as $\alpha \to 1$, provided that $\gamma > 1$, i.e., provided that there is a loss of incumbency advantage, there will always be blocking. More generally, a lower gain from innovation, i.e., a lower α, makes blocking more likely.

It is also clear that a higher level of γ, i.e., higher erosion of the incumbency advantage, encourages blocking of technological change. This is intuitive: the only reason why incumbents resist innovation is the fear of replacement. In addition, in (11.33) a higher μ makes blocking more likely. However, note that, as discussed above, the effect of μ on blocking is non-monotonic. As μ increases further, we reach the point where $P\left[\frac{1}{2} + \gamma\mu - (\alpha - 1)\right] = 1$, and then, further increases in μ make blocking less likely—and eventually when $P\left[\frac{1}{2} + \mu\right] = 1$, there will never be blocking. This establishes:

Proposition 11.21. *When μ is sufficiently small or large (political competition very high or very low), the elites will always undertake ecological change. For intermediate values of μ, technological change may be blocked.*

As emphasized above, blocking will happen because of the political replacement effect: in the region where blocking is beneficial for the incumbent ruler, the probability that he will be replaced increases when there is technological change. This implies that the incumbent ruler fails to internalize future increases in output.

11.5.4. Political Rents and Technological Change. It is straightforward to add rents from holding political power, once again denoted by R. With an argument very similar to the one in the above analysis, a greater value of these rents, R, makes technological change less likely.

The intuition is simply that technological change is blocked when it leads a greater probably the of losing power. Greater rents make losing power more costly.

11.5.5. External Threats and Technological Change. External threats, or the threat of revolution, may force technological change. Here is a simple extension to illustrate this point.

Suppose that at time t, rulers find out that there is a one-period external threat at $t+1$, which was unanticipated before. In particular, another country (the perpetrator) with technology B_t may invade.

Whether this invasion will take place or not depends on the level of output in two countries, and on a stochastic shock, p_t. If $\phi B_t - p_t > A_t$, the perpetrator will successfully invade and if $\phi B_t - p_t \leq A_t$, there will be no invasion, so $\phi \geq 0$ parameterizes the extent of the external threat: when ϕ is low, there will only be a limited threat. This formulation also captures the notion that a more productive economy, which produces more output, will have an advantage in a conflict with less productive economy.

For simplicity, suppose that there will never be an invasion threat again the future, and assume that p_t is uniform between $[0,1]$. Suppose also that $B_t = \delta A_{t-1}$. This implies that there will be an invasion if

$$p_t \leq \phi\delta - 1 - x_t(\alpha - 1),$$

were recall that x_t is the decision of the incumbent to innovate. Using the fact that p_t is uniform over $[0,1]$, and the same definition of the function $P[\cdot]$, we have the probability that the ruler will not be invaded at time t, conditional on x_t, as

$$P[1 - \phi\delta + x_t(\alpha - 1)].$$

The important point here is that the probability of invasion is higher when $x_t = 0$ because output is lower.

The same reasoning as before immediately establishes that at time t the ruler will innovate if

$$\alpha P\left[\frac{1}{2}+\mu\right]P\left[1-\phi\delta+(\alpha-1)\right]\geq P\left[\frac{1}{2}+\gamma\mu-(\alpha-1)\right]P\left[1-\phi\delta\right]. \tag{11.34}$$

When $P[1-\phi\delta]\in(0,1)$, blocking new technologies becomes less attractive in the presence of the external threat, because a relatively backward technology increases the probability of foreign invasion. Therefore, in this extended model, the emergence of an external threat might induce innovation in an economy that was other was going to block technological change.

An increase in δ or ϕ will typically make blocking less likely. For example, when $\delta\to 0$ or $\phi\to 0$, $P[1-\phi\delta]\to 1$, threat of invasion disappears and we are back to condition (11.32). For future reference, we state this result as a proposition:

Proposition 11.22. *Political elites are less likely to block technological change when there is a severe external threat (high ϕ) and when the perpetrator is more developed (high δ).*

The intuition for both comparative statics is straightforward. With a more powerful external threat or a more developed perpetrator, the ruler will be "forced" to allow innovation so as to reduce for an invasion. Therefore, this extension shows how an external threat can induce technological change.

11.6. References

(1) Acemoglu, Daron (2005) "Modelling Inefficient Institutions," forthcoming *Advances in Economic Theory World Congress 2006,* http://econ-www.mit.edu/faculty/download_pdf.php?id=1214
(2) Acemoglu, Daron and James A. Robinson (2002) "Economic Backwardness in Political Perspective," NBER Working Paper #8831, forthcoming *American Political Science Review.*
(3) Eggertsson, Thrainn (2005) *Imperfect Institutions: Possibilities and Limits of Reform*, University of Michigan Press, Ann Arbor.
(4) Feinstein, Charles (2005) *An Economic History of South Africa: Conquest, Discrimination and Development*, Cambridge University Press, London UK.
(5) Parente, Stephen L. and Edward C. Prescott (1999) "Monopoly Rights: A Barrier to Riches," *American Economic Review*, 89, 1216-1233.

CHAPTER 12

Policy under Democratic Political Institutions

So far we examined models of democracy with few institutional details. Since we discussed political institutions as the "political rules of the game" we now turn to try to describe the implications of the rules in richer settings. In lecture 2 we saw that there are some interesting correlations in the data between various details of democratic political institutions (electoral systems, president versus parliament) and various policy outcomes, such as the size of government. What mechanisms might account for such outcomes? We follow Persson, Tabellini and Roland (1997, 2000).

The model has an infinite horizon with three groups of citizen-voters, $i = 1, 2, 3$. Each group has a continuum of citizens with unit mass. Time is discrete. Preferences of a member of i in period j are represented by the utility function

$$\sum_{t=j}^{\infty} \delta^{t-j} \left(c_t^i + H(g_t) \right)$$

where c_t^i is consumption of a unique consumption good and $H(g_t)$ is utility of public goods provided in period t. Apart from providing a public good the government can tax incomes with a lump-sum tax τ_t, make a group specific transfer r_t^i or a politician specific transfer s_t^l (how much rent each gets to steal). There are three politicians, one representing each of the groups. Each individual in the society has one unit of income per-period (exogenous) and thus faces a budget constraint

$$c_t^i = 1 - \tau_t + r_t^i.$$

A policy vector is denoted

$$\mathbf{p}_t = [\tau_t, g_t, \{r_t^i\}, \{s_t^l\}]$$

In each period the political system has to determine p_t - the tax on incomes, public good provision, transfers, and politician rents. This is done subject to the government budget constraint

$$3\tau_t = g_t + \sum_i r_t^i + \sum_l s_t^l = g_t + r_t + s_t. \tag{12.1}$$

Let's begin the analysis with what PRT call a "simple legislature" just to see how the model works. In this model each region i elects one legislator and separate elections take place under plurality rule in each district. In period j each incumbent legislator has preferences

$$\sum_{t=j}^{\infty} \delta^{t-j} s_t^l D_t^l$$

so that they get utility only from rents. $D_t^l = 1$ if such a legislator is in power in period t. If out of office a legislator gets zero utility and a legislator who is voted out of office is never re-elected.

The idea is that incumbents are accountable to their district and that voters within districts coordinate their voting strategies and set a particular reservation utility level of utility such that if they get this level of utility they re-elect the incumbent. If not they replace him with an alternative politician who is identical (there are a large number of these). Crucially however, voters in different groups choose their re-election strategies non-cooperatively with respect to the other groups.

The Timing of the stage game is like this. In period t the incumbent legislators elected at the end of period $t-1$ decide on policy in a Baron-Ferejohn type legislative bargaining model.

(1) Nature randomly chooses an agenda setter a and each politician has an equal chance of becoming a.
(2) Voters formulate their re-election strategies.
(3) The agenda setter proposes p_t. To do so he makes a take it or leave it offer.
(4) Legislature votes. If 2 legislators support p_t it is implemented. If not a default outcome is implemented $\tau = s^l = \sigma > 0$ and $g = r^i = 0$.
(5) Elections are held.

The re-election strategy of voters has the form

$$D_{t+1}^l = 1 \text{ if } c_t^i + H(g_t) \geq b_t^i.$$

Voters in different groups set their b_t^i non-cooperatively. Let $\mathbf{b}_t$ be the vector of reservation utilities. Note that since this part of the stage game takes place after nature has determined who is the agenda setter, voters will take this into account. In general accountability for a will be different from accountability for $l \neq a$.

We can now state what an equilibrium looks like in this model (superscript L is for the simple legislature). The idea is to look for MPEs where players cannot condition their strategies on the history of play but only the state. Note that all of the action is within the stage game.

Definition 12.1. *An MPE of the simple legislature is a vector of policies* $p_t^L(\mathbf{b}_t^L)$ *and a vector of reservation utilities* $\mathbf{b}_t^L$ *such that in any period* t *(1) for any given* $\mathbf{b}_t^L$*, at least one legislator* $l \neq a$ *prefers* $p_t^L(\mathbf{b}_t^L)$ *to the default outcome; (2) for any given* $\mathbf{b}_t^L$*, the agenda setter* a *prefers* $p_t^L(\mathbf{b}_t^L)$ *to any other policy satisfying (1); (3) the reservation utilities* $\mathbf{b}_t^L$ *are optimal for the voters in each district, taking into account that policies are chosen according to* $p_t^L(\mathbf{b}_t^L)$ *and taking the identities of the legislators and the other* b_t^{-iL} *as given.*

There is a unique MPE in this model which is stationary so we drop the time subscripts. It has the following form.

Proposition 12.1. *In the equilibrium of the simple legislature* $\tau^L = 1$*;* $s^L = 3(1-\delta)/(1-(\delta/3))$*;*

$$g^L = \min\left\{\hat{g}, \frac{2\delta}{1-(\delta/3)}\right\}$$

where $\hat{g}$ *satisfies* $H'(\hat{g}) = 1$*;* $r^{aL} = 2\delta/(1-(\delta/3)) - g^L$*, and* $r^{lL} = 0$ *for* $l \neq a$*;* $b^{aL} = H(g^L) - g^L + 2\delta/(1-(\delta/3))$*,* $b^{iL} = H(g^L)$ *for* $i \neq a$*.*

There are several interesting implications. First, public goods are underprovided relative to the Lindahl-Samuelson rule which here is $3H'(g) = 1$. Second, takes are maximal, though note that there are no distortions associated with taxation in this model. Third, s^L denotes the total amount taken by politicians, and $s^L > 0$ so that the politicians get rents. Finally, the constituents of the agenda setter benefit by getting transfers $b^{aL} > 0$, though no other group does so.

To see why the equilibrium has this form start with backward induction. After the agenda setter has been chosen and the re-election rules determined, legislative bargaining takes place. In this game the agenda setter aims to form a minimum winning coalition and thus wants to design a policy such that one other group supports it. The agenda setter makes a take-it-or leave it offer to the other legislator who is easiest to buy, where the price will be in terms of what the agenda setter has to offer the politician to get them to say yes. In turn this will be determined by what that legislator has to deliver to his voters to get re-elected.

First observe that the legislative bargaining game must have the outcome $r^m = r^n = 0$ for $m, n \neq a$. As noted above, how cheap a legislator is to buy depends on the reservation utilities of the districts. To buy a legislator's vote, a must make transfers to his district. If b^m, say, is set very high, then group m will be costly to buy for a and will be excluded from the winning coalition. Exclusion means no transfers for the district. This situation creates a Bertrand game between districts $m, n \neq a$ and imply that they underbid each other until $b^m = b^n = 1 - \tau + H(g)$ and $r^m = r^n = 0$.

Now define W to be the expected continuation value to any legislator from being elected. Note that since in the future, if a legislator is re-elected he has an equal change of becoming agenda setter, or of taking on each of the other two possible roles, each legislator has the same continuation value. We now argue that in equilibrium $s^L \geq 3 - 2\delta W$. First note that in forming the minimum winning coalition a gives rents only to the legislator he includes in the coalition, say legislator m. Moreover, he gives just enough to make m indifferent between accepting and saying no. Given that the excluded legislator says no, if m deviates they all get the status-quo payoff and are thrown out of office. Hence a must make an offer to m, s^m such that

$$s^m + \delta W = \sigma. \tag{12.2}$$

Now consider whether or not a wishes to be re-appointed. Alternatively, he can propose a p_t which will get him thrown out of office (subject to the constraint that one other legislator has to agree to it). The best such p_t involves $g = r = 0$ and $\tau = 1$. Here m gets s^m but a provides no public goods or transfers and sets s^a as high as possible. To avoid this, we require

$$s^a + \delta W \geq 3 - \sigma \tag{12.3}$$

where the deviation payoff is 3 (s^a=total tax revenue) minus the payment to m to get agreement. Combining (12.2) and (12.3) we see that a and m will choose a policy leading to re-election if and only if

$$s \equiv s^a + s^m + 2\delta W \geq 3 \tag{12.4}$$

as claimed. When this is satisfied all three legislators are re-elected.

Now consider the rest of the equilibrium. Note that the policy choices of a must be such as to maximize the utility of voters in the district that the agenda setter comes from, subject to the government budget constraint and (12.4). Thus the proposal of a solves the maximization problem

$$\max_{r,\tau,g} \; r + 1 - \tau + H(g) \text{ subject to (12.1) and (12.4).}$$

Combining (12.1) and (12.4) to eliminate s we find

$$3(\tau - 1) + 2\delta W \geq r + g$$

which will hold as an equality since voters do not want to concede any more rents to a than they need to. Note that we must have

$$\begin{aligned} W &= \frac{s^a}{3} + \frac{s^m}{3} + \delta W = \frac{s^L}{3} + \delta W \text{ hence} \\ W &= \frac{1}{1-(\delta/3)}. \end{aligned}$$

This follow from the fact that in the next period, each group has a probability $1/3$ of being the agenda setter and getting payoff s^a and a probability $1/3$ of being the other group included in the winning coalition and getting s^m we then use the fact that $s^a + s^m = 3 - 2\delta W$. We can substitute the constraint into the objective function, eliminating r to derive

$$\max_{r,\tau,g} 3(\tau - 1) + \frac{2\delta}{1-(\delta/3)} - g + 1 - \tau + H(g) \tag{12.5}$$

The Kuhn-Tucker conditions for (12.5) are

$$\begin{aligned} -1 + H'(g) &= 0 \text{ and } g > 0 \text{ or } -1 + H'(g) > 0 \text{ and } g = 2\delta W, \\ 3 - 1 &> 0 \text{ and } \tau = 1. \end{aligned}$$

with r determined residually by

$$r = \frac{2\delta}{1-(\delta/3)} - g.$$

This establishes the proposition. The final thing to note is that b^a is simply the utility of members of group a evaluated at the solution to (12.5).

The intuitions for the main results are as follows: Public goods are undersupplied because the Bertrand competition between the non-agenda setter groups means that the agenda setter only has to please voters in his own group. Thus he ignores the benefits to the other groups of providing public goods, while internalizing the full cost. The same logic implies that only voters in this group get redistribution. Finally the two legislators in the winning coalition get rents because citizens cannot punish them hard enough. As in efficiency wage models, when the stick is too small, the carrot has to be used and citizens have to concede rents to politicians to stop them deviating and grabbing all of the tax revenues.

12.0.1. Presidential-Congressional Institutions. We now enrich the model to see how a different formulation of institutions influences the equilibrium outcome. It is useful to think of the implications of the different institutional options in terms of how they reconcile three conflicts of interests. There is a conflict of interest between the citizens and the politicians (citizens want public goods and transfers, politicians want rents), between citizens (since more transfers for one district means less for another) and between politicians (more rents for one means less for another). Consider a situation where taxation and expenditure decisions are separated. Instead of there being one agenda setter, there are now two.

The Timing of the stage game is like this. In period t the incumbent legislators elected at the end of period $t-1$ decide on policy.

(1) Nature randomly chooses two agenda setters a_τ for the taxation decision, and a_g for the allocation of revenues. Each politician has an equal chance of becoming an agenda setter.
(2) Voters formulate their re-election strategies.
(3) Agenda setter a_τ proposes a taxation decision τ.
(4) Congress votes. If at least 2 legislators are in favor the policy is adopted. Otherwise a default tax rate $\tau = \sigma < 1$ is enacted.
(5) The agenda setter a_g proposes $[g, \{s^l\}, \{r^i\}]$ subject to $r + s + g \leq 3\tau$.
(6) Congress votes. If 2 legislators support the policy it is implemented. If not a default outcome is implemented with $\tau = s^l$ and $g = r^i = 0$.
(7) Elections are held.

Note here that what happens at stage 3 is binding subsequently. At stage 5 a_g tires to form a coalition which is optimal for him and we assume that if he is indifferent between the two other politicians then they each become part of the winning coalition with probability 1/2.

We now have the following result where superscript C indicates the equilibrium outcomes in the Presidential-Congressional game.

Proposition 12.2. *There is a unique MPE of the Presidential-Congressional Game with,*

$$\tau^C = \frac{1-(\delta/3)}{1+(2\delta/3)} < 1;\ s^C = 3\frac{1-\delta}{1+(2\delta/3)} < s^L;\ g^C = \min\left\{\hat{g}, \frac{2\delta}{1+(2\delta/3)}\right\} \leq g^L$$

and

$$\begin{aligned} r^{aC} &= \frac{2\delta}{1+(2\delta/3)} - g^C \leq r^{aL} \text{ and } r^{iC} = 0 \text{ for } i \neq a, \\ b^{aC} &= H(g^C) - g^C + \frac{2\delta}{1+(2\delta/3)} \text{ and } b^{iC} = H(g^C) \text{ for } i \neq a. \end{aligned}$$

To see why the equilibrium looks like this we again apply backward induction within the stage game. a_g takes τ as given and incentive compatibility implies that he will offer

$$s^{m_g} + \delta W = \tau$$

to the other partner in the winning coalition. This in turn implies that for re-election to be desired, a_g must get enough rent so that

$$s^{a_g} + \delta W \geq 2\tau$$

given that he has to give τ to m to get a yes vote. Hence total rents s must be such that $s + 2\delta W \geq 3\tau$. Using the government budget constraint incentive compatibility entails

$$g + r \leq 2\delta W. \tag{12.6}$$

As before, Bertrand competition between the non-agenda setter groups implies that they get no transfers. Thus the optimal allocation from the point of view of voters in the group with the agenda setter maximizes $r + H(g)$ subject to (12.6). This gives $g = \min[\hat{g}, 2\delta W]$, $r = 2\delta W - g$, and $s = 3\tau - 2\delta W$.

Now move back to the taxation decision noting that $a_\tau \neq a_g$. Note that the voters in the group of agenda setter a_τ will not benefit from high taxes since these will be allocated by a different legislator subsequently. Nevertheless, the re-election rule has to allow taxes to be sufficiently high to avoid a_τ deviating. Indeed we now show that $\tau^C \geq 1 - \delta W$. Note first that with probability one half, a_τ will be in the winning coalition when expenditure is decided. a_τ will not deviate from a tax proposal if

$$\frac{s^m}{2} + \delta W \geq v^d$$

where v^d is the deviation utility from some other tax proposal. The highest deviation payoff that a_τ could get would be by setting $\tau^d = 1$ since if he deviates then the players get the status quo payoffs $s^l = \tau$. Since a_τ is in the winning coalition with probability 1/2, the highest v^d is 1/2. Thus an incentive compatible τ^C must satisfy

$$\frac{s^{m_g}}{2} + \delta W \geq \frac{1}{2} \text{ or using } s^{m_g} \text{ derived above } \frac{\tau^C - \delta W}{2} + \delta W \geq \frac{1}{2}$$

which gives $\tau^C \geq 1 - \delta W$ as claimed.

Now if $\tau^C = 1 - \delta W$ is high enough to finance the maximum amount of incentive compatible public goods, the optimal voting rule for citizens of the group of a_τ would be to make him propose this τ^C. This will be supported by the third legislator (not a_g). From this and using the fact that W is defined as before the Proposition follows.

Compared to the simple legislature, taxes are lower as are rents. However, public goods are even further from the optimal level. Transfers are again concentrated to one specific group, here that represented by a_g. Here the separation of powers element allows the voters to restrict the amount of rents that the politicians can extract and also reduces taxes because taxes are set by one agent but allocated by another.

12.1. Parliamentary Institutions

Now we compare the outcomes to a different extensive form game with Persson, Roland and Tabellini argue captures some of the key institutional features of a parliamentary system.

The Timing of the stage game is like this. In period t the incumbent legislators elected at the end of period $t-1$ decide on policy.

(1) Nature randomly chooses coalition partners from amongst the incumbent legislators. One becomes the agenda setter a the other becomes her junior partner.
(2) Voters formulate their re-election strategies.
(3) Agenda setter a proposes a taxation decision $[\tau_a, \{r_a^i\}, g_a, \{s_a^l\}]$ subject to $r_a + g_a + s_a \leq 3\tau_a$.
(4) The junior partner can veto the proposal from stage 3. If approved the proposal is implemented and the game goes to stage 9. If not the government falls and the game goes to stage 5.
(5) Nature randomly selects a new agenda setter a' from the three legislators.
(6) Voters re-formulate their re-election strategies.
(7) The new agenda setter a' proposes an entire allocation $p_{a'}$.
(8) Parliament votes. If $p_{a'}$ is supported by two legislators it is implemented. If not a default outcome is implemented with $\tau = s^l = \sigma$ and $g = r^i = 0$.
(9) Elections are held.

The emphasis here is on the idea that a parliamentary government can fail if it looses a vote of confidence. Voting in Parliament is not sequential so that the model does not have the checks and balances and none of the separation of powers inherent in the previous game. If a government crisis occurs this wipes away the entire proposal, whereas before if an allocation of expenditure was defeated this did not undo the tax rate previously determined. The main result here is as follows.

Proposition 12.3. *In the parliamentary regime there is a continuum of equilibria such that* $\tau^P = 1 = \tau^L > \tau^C$;

$$s^P = 3\frac{1-\delta}{1-(\delta/3)} = s^L > s^C;\ s^{aP} = \frac{2}{3}s^P,\ s^{mP} = \frac{2}{3}s^P;$$

$\bar{g} \geq g^P > g^C$ *where* $H'(\bar{g}) = \frac{1}{2}$;

$$r^P = \frac{2\delta}{1-(\delta/3)} - g^P \geq 0;$$

$r^{iP} \geq 0$ if $i = a, m$; and $r^{iP} = 0$ if $i = n$. If $r^{iP} > 0$ for $i = a, m$ then $g^P = \bar{g}$, $b^{iP} = H(g^P) + r^{iP}$,

$$b^{a'P} = H(g') - g' + \frac{2\delta}{1-(\delta/3)}$$

and $b' = H(g')$ with

$$g' = \min\left\{\hat{g}, \frac{2\delta}{1-(\delta/3)}\right\},$$

all politicians are re-elected and a government crisis never occurs.

The proof of this follows a similar logic to the above. First note that if one of the governing coalition vetoes the initial proposal the legislators play the simple legislative bargaining model that we began with. This game has the same solution to the previous one and this will pin down the lowest possible payoff that the agenda setter can offer the junior partner. With some probability the junior partner can be chosen as agenda setter a', etc. This continuation game also pins down the s^P (total rents) that voters have to concede to the politicians. Now moving backward the key observation is that since the voters in the two groups that form the governing coalition simultaneously choose their reservation utility levels, there are multiple (a continuum) of Nash equilibria, i.e there are lots of pairs of (b^a, b^m) which are mutual best responses and which will map into different distributions of (r^a, r^m) between the coalition partners. The key observation is that in this model there is not a Bertrand game between the members of the government, so when g is chosen it will internalize the utility of both members of the coalition, hence the condition $2H'(\bar{g}) = 1$. Relative to the previous models this means that the supply of public goods will be larger. Hence also the fact that two groups of voters get transfers, rather than one as in the previous two models. However, since taxation and expenditure decisions are not decoupled now. This implies that the members of the governing coalition are residual claimants on taxation and wish to set $\tau = 1$ (to extract as many resources a possible from the third group). Note that since rents to politicians are pinned down by the simple legislature, they are the same as in the first model.

The model we looked at above is a seminal attempt to model the importance of the institutional details in democracy. The focus of these papers is very attractive for economists trained in the public finance tradition. As yet however, it is unclear if these models capture the first-order effects of these institutions. This is a very open and interesting research agenda.

12.2. Other Important Institutions

Maybe the first such idea about the impact of democratic institutions was Duverger's Law. Duverger (1954) argued that there was a close relationship between majoritarian electoral systems and two-party systems. When parties compete for office in electoral districts where one party wins, then three party equilibria cannot exist. The basic logic is that anyone who is voting for the party who comes first must be casting a 'wasted vote.' In response such a person will wish to vote strategically and switch her vote to whichever of the first two parties she prefers (see Austen-Smith Banks, 2004, Chapter 8 for some formalizations of this idea, or Cox, 1997, for evidence and extensive discussion). This is a clean prediction about the impact of electoral systems on the equilibrium number of parties.

More generally it would be nice to have models where the incentives of different electoral systems are modelled in detail. However, this is a very new and unsettled area of research. The details of actual electoral systems are very complex and it is hard to model their incentive effects in parsimonious ways that are also convincing. Persson and Tabellini (2000, Chapter 8) develop some simple approaches, but in their model of PR there are only two parties, which is probably not very reasonable. However, the study of models with many parties is in its infancy. One complication of such models is that sincere voting is typically not optimal with three or more parties. Nevertheless, the topic is very interesting. We saw already that there are some stylized facts and correlations that suggest electoral systems may matter (bearing in mind that so far there is not a convincing empirical strategy for estimating the causal effects). Austen-Smith (2000) was the first to point out that countries with PR seem to have more income redistribution, and see also Milesi-Ferretti, Perotti and Rostagno (2002). Austen-Smith's paper also proposes a formal approach to three party competition, see Baron and Diermeier (2001) and Austen-Smith and Banks (2004, Chapter 8).

The political science literature is full of a huge number of claims about the incentive effects of electoral systems and how they influence equilibrium political behavior. For instance in the literature on Latin America it is now very common to blame 'clientelism' or dysfunctional party systems on the form of the electoral system (Carey and Shugart, 1995, Ames, 2001).

Thus we are still at the early stages of trying to characterize what the first-order effects of these institutions are on incentives. Once we manage to do that we can then try to understand why countries have the political institutions they do. We shall return to some of the foolhardy attempts to think this issue through later in the course (e.g., Mazzuca and Robinson, 2005, Ticchi and Vindigni, 2005).

12.3. References

(1) Ames, Barry (2001) *The deadlock of democracy in Brazil*, Ann Arbor; University of Michigan Press.

(2) Austen-Smith, David (2000) "Redistributing Income under Proportional Representation," *Journal of Political Economy*, 108, 1235-1269.

(3) Austen-Smith, David and Jeffrey Banks (2004) *Positive Political Theory II: Strategy and Structure*, Ann Arbor; University of Michigan Press.

(4) Baron, David P. and Daniel Diermeier (2001) "Elections, Governments and Parliaments in Proportional Representation Systems," *Quarterly Journal of Economics*, 116, 933-967.

(5) Carey, John M. and Mathew S. Shugart (1995) "Incentives to Cultivate a Personal Vote: A Rank-Ordering of Electoral Formulas," *Electoral Studies*, 14, 417-439.

(6) Cox, Gary W. (1997) *Making votes count : strategic coordination in the world's electoral systems*, New York; Cambridge University Press.

(7) Duverger, Maurice (1954) *Political parties, their organization and activity in the modern state*, New York; Wiley.

(8) Mazzuca, Sebastián and James A. Robinson (2005) "Political Conflict and Power-sharing in the Origins of Modern Colombia," Unpublished, Department of Government, Harvard.

(9) Milesi-Ferretti, Gian Maria, Roberto Perotti and Massimo Rostagno (2002) "Electoral Systems And Public Spending," *Quarterly Journal of Economics*, 117, 609-657.

(10) Persson, Torsten, Gerard Roland and Guido Tabellini (1997) "Separation of Powers and Political Accountability" *Quarterly Journal of Economics*, 112, 1163-1202.

(11) Persson, Torsten, Gerard Roland and Guido Tabellini (2000) "Comparative Politics and Public Finance," *Journal of Political Economy*, 108, 1121-1161.

(12) Persson, Torsten and Guido Tabellini (2000) *Political Economics: Explaining Economic Policy*, Cambridge; The MIT Press, Chapter .

(13) Ticchi, Davide and Andrea Vindigni (2005) "Endogenous Constitutions," Unpublished, Department of Politics, Princeton University.

CHAPTER 13

The Form of Redistribution

At the heart of the models of politics and institutions that we have been developing is distributional conflict. Because there are conflicting interests and because institutions are crucial in determining the distribution of economic resources and political power, there is conflict over institutions. Conflict however is not necessarily inefficient. Cooperative or non-cooperative bargaining theory suggests that with complete information, two parties with completely opposed interests can nevertheless strike a deal to efficiently divide some income/rents. People don't want to throw rents away and recognizing how the structure of a bargaining situation maps into relative bargaining power, bargainers prefer to settle right away. In addition Coasian logic suggests that conflicts of interests can be settled by contracts or side-payments, whose direction again reflects some underlying structure of property rights or bargaining power. If you read the survey by Garfinkle and Skaperas (2006) on conflict we discussed earlier they have a nice section of "bargaining in the shadow of conflict" - anticipating that fighting is costly, people will want to find an agreement which respects an agent's subsequent fighting power.

We have discussed some ideas as to why we might not expect conflict to be efficient. In this lecture we look at another angle. A prime source of inefficiency in politics seems to be not just that redistribution takes place, but that it takes place in highly inefficient ways. The *locus classicus* of this Robert Bates' seminal book on the political economy of underdevelopment in Africa (1981) *Markets and States in Tropical Africa*. Bates argued that variation in agricultural performance after independence in Africa was due to the distribution of political power. In places where farmers had little power (Ghana) they were taxed at punitive rates by urban groups and agricultural output declined. In places where they had more power (Kenya) they avoided this and agriculture did much better. Why did farmers have power in some places rather than others? Bates' argument was this was because of collective action. In Ghana, for instance, the main export crop, cocoa, was grown by smallholders who could not solve the collective action problem. In Kenya export agriculture inherited the institutions of the white farmers of the "Happy Valley" who were well organized to further their interests. Other factors were also important, in Kenya farmers were primarily Kikuyu, which was the main ethnic group of the government of Jomo Kenyatta. In Ghana cocoa farmers were Ashanti, but the state was not.

But how come they killed the goose that laid the golden egg in Ghana? (see Figure) .Here Bates analyzes in detail the political rationality of the instruments that were used. The basic thrust of the argument is that efficient redistribution is not politically rational. More specifically, the ruling governments used a particular political strategy, which political scientists call "clientelism" to stay in power. This strategy has profound implications for the form redistribution takes. For instance, the you don't want to redistribute income using public goods, because these cannot be targeted to your supporters or withheld from your opponents. Another interesting argument is about rationing. He examines why exchange rates were so overvalued in African countries, with disastrous implications for exports. His explanation is that when you set a disequilibrium price for something (foreign exchange) you automatically create a scarce good whose allocation confers rents on people. Foreign exchange is a private good which can be allocated to supporters and withheld from opponents but you have to make it scarce for it to be valuable.

Of course these concerns are close to a large literature on US politics. Maybe the first economists who woke up and began to realize that political economy was important were those studying international trade. Here economists have such a strong sense of free trade being a good policy that it is very hard to concoct convincing market failure explanations for why tariff protection is so important. But then if deviations from free-trade are all about redistributive conflict between winners and losers from free trade, why is it that this redistribution takes such inefficient forms like tariffs (see Rodrik, 1995, for a discussion of the inefficiency of the form of redistribution in this context)?

The form of redistribution is thus a first-order topic in this class. Here our aim is to discuss some of the mechanisms that lead to inefficient redistribution (see Stiglitz, 1998, for some ruminations on this issue in the light of his experience on the Council of Economic Advisers to Clinton).

13.1. Inefficient Redistribution: General Issues

Most redistribution in practice takes an inefficient form, that is, redistribution hardly ever takes the form of lump sum transfers away from some groups towards others.

Part of this is because of informational problems. Lump sum transfers are often not possible (e.g., consider a situation in which agents differ according to their productivity, and these productivities are unobserved by the taxation authority. This will lead to a Mirrlees type optimal income taxation, and a greater level of taxation will increase distortions).

Nevertheless, there are many examples of redistribution that takes a clearly more inefficient form than some of the available alternatives.

A prominent example is farm subsidies in the form of price supports or input subsidies. These policies distort relative prices and discourage the reallocation of productive resources to other sectors, and the general perception among economists and political scientists is that these policies are mainly designed for transferring income to farmers.

Another example is trade policy which provides protection to domestic industries via tariffs and quotas, even when free trade is socially beneficial.

Why does redistributive policy take an inefficient form? There are many different types of answers in the literature only some of which have really been formalized (and only some of which may be coherent...):

- A commitment to inefficient methods of redistribution may be a way of constraining the overall amount of redistribution (Rodrik, 1986, Wilson, 1990, Becker and Mulligan, 1998).
- Voters are myopic, and are fooled by inefficient redistribution methods, whereas they would not be fooled by lump sum transfers.
- Voters are rational, and we are in an equilibrium in which political parties exploit their imperfect information to make transfers to their favorite groups under the disguise of "Pigovian" subsidies (Coate and Morris, 1995).
- Inefficient redistribution as a method of maintaining future power (Acemoglu and Robinson, 2001).
- By politicians because it helps them to create incentives for voters to support them (Bates, 1981, Persson and Tabellini, 1999, Lizzeri and Persico, 2000, Robinson and Torvik, 2005).
- By politicians because it allows them to take credit for policy and influence the beliefs of voters about their preferences.
- By interest groups because it influences the type of game (and therefore the terms of trade) between them and politicians (Dixit, Grossman and Helpman, 1997).
- By interest groups because it can help them solve the collective action problem.

Most of these explanations have some degree of group conflict embedded in them.

Now we will outline some simple models to give the basic ideas.

13.2. Inefficient Redistribution to Constrain the Amount of Redistribution

Consider a society that consists of two groups, the rich and the poor.

The poor form the majority, and will decide the amount of redistribution in society. But the rich design the institutions. In particular, suppose that a fraction λ of the agents is rich with income h^r, and the remaining agents are poor with income $h^p < h^r$.

Average income in the economy is

$$h = \lambda h^r + (1 - \lambda) h^p$$

There is a linear tax rate τ imposed on all incomes, and the proceeds are distributed lump sum to all agents.

Taxation creates a dead weight loss of

$$\mu c(\tau) h$$

where $c(\tau)$ is strictly increasing and convex, and μ parameterizes the degree of inefficiency of redistributive of taxation.

The overall amount of lump sum subsidy is

$$T = [\tau - \mu c(\tau)] h$$

Imagine the following timing of events:

- The rich choose μ from a set $\left[\underline{\mu}, \bar{\mu}\right]$.
- The poor decide τ.

Let us solve this game backwards:

For given μ, the utility of the poor agent is

$$(1 - \tau) h^p + [\tau - \mu c(\tau)] h,$$

hence, their preferred tax rate will satisfy the first-order condition

$$h^p = \left[1 - \mu c'(\tau^*(\mu))\right] h$$

where we wrote the preferred tax rate as $\tau^*(\mu)$ to emphasize that it is a function of the degree of inefficiency of the tax system.

Straightforward differentiation gives that

$$\frac{d\tau^*(\mu)}{d\mu} = -\frac{c'(\tau^*(\mu))}{\mu c''(\tau^*(\mu))} < 0,$$

that is, a greater inefficiency in the tax system reduces the optimal tax rate for the poor. Intuitively, as taxes become more inefficient, a given level of tax rate creates less redistribution towards the poor, and become less desirable.

Let us now look at the utility of a rich agent, anticipating the tax rate that will be chosen by the poor agents. This is

$$(1-\tau^*(\mu))\,h^r + [\tau^*(\mu) - \mu c(\tau^*(\mu))]\,h,$$

The rich agent will choose μ to maximize this utility.

The derivative of this utility evaluated at $\mu = \underline{\mu}$ is

$$\left[-h^r + \left[1 - \mu c'\left(\tau^*\left(\underline{\mu}\right)\right)\right] h\right] \frac{d\tau^*\left(\underline{\mu}\right)}{d\mu} - c\left(\tau^*\left(\underline{\mu}\right)\right)$$

If this derivative is positive, it means that the rich prefer a more inefficient tax system than the most efficient available alternative characterized by $\underline{\mu}$.

The first term is positive by virtue of the fact that there is redistribution away from the rich: a higher μ reduces the tax rate, and hence benefits rich agents

So if $c\left(\tau^*\left(\underline{\mu}\right)\right)$ is small, the rich would like to create some inefficiency in the tax system in order to reduce redistribution away from themselves towards the poor.

Intuitively, more inefficiency in the tax system reduces the desired tax rate of the poor, and may help the rich.

Is this story plausible as an explanation for why there are inefficient methods of redistribution in practice?

Possibly not: the argument requires that it is easy to commit to the form of redistribution, but not to the level of redistribution. It seems far-fetched to believe that the political system can commit to the form of redistribution, which is much harder to specify, but cannot commit to the level of redistribution. This is still a set of ideas waiting for a convincing application.

13.3. Inefficient Redistribution to Fool Voters

Perhaps the most common argument for why farmers receive price subsidies is that this is a "hidden" transfer to them, and the voters tolerate such hidden transfers.

If instead there were direct transfers to farmers, voters would object to it.

A simple form of this argument would build on myopia: voters do not understand that price subsidies are transfers, but would immediately understand when they see direct transfers.

But perhaps this argument views voters as too "irrational".

Can a similar argument be developed where voters are not irrational, but simply misinformed?

Coate and Morris develop such a model. Here is a simple version of their model.

Suppose there are two groups 1 and 2. There are three different policies possible:

- Do nothing, in which case the two groups receive their status quo payoffs a_1 and a_2.
- A price subsidy to group 1. In this case, the two groups receive $b_1+\varepsilon$ and $b_2+\varepsilon$, where ε is takes the value 0 with probability $1-p$, and the value $\varepsilon = e$ with probability p. Voters do not know the value of ε, but we will assume that politicians do. In addition, we have

$$\begin{aligned} b_1 &> a_1,\, b_2 < a_2 < b_2 + e \\ &\text{and} \\ b_1 + b_2 &< a_1 + a_2 < b_1 + b_2 + 2e \end{aligned}$$

  That is group 1 always gains as a result of this policy, and the society may gain or lose depending on the value of ε. In other words, with probability p, we can think of this price subsidy as a Pigovian subsidy correcting some market failure.
- Direct transfer to group 1, so that the payoffs are $a_1 + \tau$ and $a_2 - \tau$.

In addition, suppose that there are two different types of parties in power. With probability q, we can have a party that wants to maximize efficiency, and with probability $1-q$, we have a "partisan" party that wants to transfer resources to group 1. All parties receive some additional (small) utility from being in power.

Voters do not observe the identity of the party in power.

Finally, assume that group 2 is the majority, and they can vote the party out of power once they see its policy.

The timing of events is:

- Nature determines the identity of partying power and the value of ε.
- The party chooses its policy.
- Group 2 agents decide whether to kick this party out of power. If they do so, the status quo payoffs are realized.
- If the party is not kicked out of power, its policy gets implemented, and payoffs are realized.

Suppose the party turns out to be the one that favors group 1. Can it directly transfer resources to group 1? The answer is clearly no. If it did so, group 2 agents would immediately removed this party from power.

However, they can undertake hidden transfers by choosing the price subsidy policy.

Will this hidden transfer "fool" the voters? It depends on parameters.

The price subsidy policy could be observed in two different eventualities:

- Either the party in power is the social welfare maximizer, and in fact the policy is welfare improving. This has probability pq.
- Or it is the partisan party, wanting to transfer resources to group 1. This has probability $(1-q)$.

The expected return to group 2 voters from the price subsidy policy is therefore:

$$\frac{pq\left(b_2+e\right)+(1-q)\,b_2}{pq+(1-q)}$$

So they will prefer to accept this policy only if

$$\frac{pq\left(b_2+e\right)+(1-q)\,b_2}{pq+(1-q)}>a_2 \tag{13.1}$$

Therefore, when (13.1) is satisfied, an equilibrium in which the partisan party fools the rational voters is possible.

Is this a plausible explanation for inefficient redistribution?

Probably not. There are two problems:

- Only policies that have a sufficiently high ex ante probability of being efficient can be used for inefficient transfers. Subsidies to farmers do not seem to satisfy this requirement.
- If any value of τ is allowed in combination with price subsidies, a forward induction argument (e.g. The Intuitive Criterion of Cho and Kreps) would rule out inefficient redistribution. Intuitively, the party that cares about social welfare would combine the price subsidy with a negative value of τ to reveal its type, and this policy would of course not be desirable for the partisan party. Thus we would end up with a separating equilibrium, in which the voters would be able to tell apart the social-welfare-maximizing and partisan parties, and there would be no inefficient redistribution.

13.4. Inefficient Redistribution to Maintain Power

Another possibility is that inefficient redistribution arises as a method of ensuring future redistribution.

Imagine that efficient redistribution for the sector, say farming, would encourage some agents to leave farming, or at least young agents not to enter farming.

Suppose also that future redistribution depends positively on the number of farmers (in general, we can have situations in which a greater number of agents in a given sector may be an advantage or a disadvantage for political power, because of free rider problems).

In this case, current farmers may prefer inefficient redistribution in order to guarantee future redistribution.

The following model illustrates this point.

Consider the following two-period economy (periods 0 and 1) with a single consumption good produced by one of two sectors, farming and manufacturing.

In the first period there are $1-\delta$ agents with a fraction n_0 in farming and $1-n_0$ in manufacturing.

These agents cannot change sector.

All agents are risk neutral and discount the second period by a factor $\beta \in (0,1)$.

In each period, a farmer produces an output of B and a manufacturer produces output A, with $A > B$.

Farmers cannot be taxed (e.g. they can hide their output costlessly), while manufacturers can be taxed a maximum of T (e.g. they can hide their output at a cost of T) where $T < A$.

At the beginning of period 0, δ new agents arrive and choose which sector to enter. This decision is irreversible. There are no new agents in period 1.

Let τ_0 and τ_1 denote the tax on manufacturers in periods 0 and 1 respectively, where $\tau_t \in [0,T]$, for $t = 0,1$.

The tax revenue can be redistributed to farmers in two distinct forms.

- a transfer to agents who are in farming at the beginning of the period, denoted by $\theta_t \geq 0$, for $t = 0,1$.
- a general price subsidy which all farmers receive, denoted by $\mu_t \geq 0$.

The difference between μ_0 and θ_0 is that only those who were initially farmers at $t = 0$ receive θ_0, whereas μ_0 is also paid out to young agents who enter farming at time $t = 0$. θ_0 therefore approximates an efficient transfer as it is conditioned on characteristics outside the agents control. In contrast, μ_0 subsidizes farm output and encourages new agents to enter farming, and so, is an inefficient method of redistribution.

Ignoring political economy considerations, existing farmers prefer θ-transfers to μ-transfers, because they do not have to share the former with newly arriving farmers. However, it will turn out that political economy considerations may nonetheless encourage existing farmers to choose μ-transfers.

Consider the following reduced form political process which determines the current tax rate on manufacturers as a function of the number of farmers. More explicitly, the tax rates in the two periods are

$$\begin{aligned} \tau_0 &= \tau(n_0) \in [0, T] \\ \tau_1 &= \tau(n_1) \in [0, T]. \end{aligned} \tag{13.2}$$

The assumption $\tau \geq 0$ incorporates the fact that farmers cannot be taxed. Notice that the tax rate in period t is only a function of the fraction of the population in farming at the time.

Assume that the function τ satisfies the following two conditions.

if $n \leq n^-$ then $\tau(n) = 0$

if $n \geq n^+$ then $\tau(n) = T$

Assume also that the division of the tax revenue between θ- and μ- subsidies is decided only by farmers.

The timing of political and economic events is as follows.

- In period 0, the political economy process determines τ_0, then the farmers decide θ_0 and μ_0.
- Young agents are born, they observe the policy vector, and decide which sector to enter.
- Then production takes place and the policy is implemented.
- At the beginning of period 1, the political process determines τ_1 and θ_1.
- The model ends following production and implementation of the chosen policy.

Define x as the fraction of new agents going into farming at time $t = 0$. The government budget constraints in the two periods are

$$(1 - \delta)(1 - n_0)\tau_0 = (1 - \delta)n_0 (\theta_0 + \mu_0) + \delta\mu_0 x \tag{13.3}$$

$$(1 - n_1)\tau_1 = n_1\theta_1. \tag{13.4}$$

In equation (13.3), $(1 - \delta)(1 - n_0)\tau_0$ is total tax revenue, $(\theta_0 + \mu_0)$ is the total per capita transfer to the $(1 - \delta)n_0$ existing farmers, and μ_0 is the inefficient transfer that the δx newcomers who enter farming receive. In equation (13.4), $(1 - n_1)\tau_1$ is total tax revenue, and is distributed among n_1 farmers. [...Note that young agents who go into manufacturing do not get taxed in period 0, and they may also not receive any transfers when they go into farming (that is, if $\mu_0 = 0$)...]

Notice that although the political process can discriminate between young and old farmers in period 0, this is not possible in period 1.

Let V^f and V^m be the expected utilities (at time 0) of old farmers and manufacturers. Let W^f and W^m be the expected utilities (at time 0) of new agents who choose farming and manufacturing. Then,

$$V^f(\theta_0, \mu_0, \theta_1) = B + \theta_0 + \mu_0 + \beta\,[B + \theta_1]\,, \tag{13.5}$$

$$V^m(\tau_0, \tau_1) = A - \tau_0 + \beta\,[A - \tau_1]\,, \tag{13.6}$$

and

$$W^m(\tau_1) = (1+\beta)A - \beta\tau_1, \tag{13.7}$$

$$W^f(\mu_0, \theta_1) = (1+\beta)B + \mu_0 + \beta\theta_1. \tag{13.8}$$

Newcomers make their occupational choices after observing μ_0. Their strategy is therefore conditioned on μ_0, and we write the fraction of new agents who go into farming when the subsidy is μ as $x(\mu)$. Then the optimal sectoral choice of new agents in period 0 is:

$$\begin{aligned} x(\mu) &= 0 && \text{if } W^m(\tau_1) > W^f(\mu, \theta_1) \\ x(\mu) &= 1 && \text{if } W^m(\tau_1) < W^f(\mu, \theta_1) \\ x(\mu) &\in [0,1] && \text{if } W^m(\tau_1) = W^f(\mu, \theta_1). \end{aligned} \tag{13.9}$$

$x(\mu)$ defines the best response function (correspondence) of newcomers for all possible levels of subsidies.

The fraction of farmers in the population at time $t = 1$ is then

$$n_1 = (1-\delta)n_0 + \delta x, \tag{13.10}$$

A pure strategy subgame perfect Nash equilibrium is a tuple,
$\{x(\mu), n_1, \tau_0, \theta_0, \mu_0, \tau_1, \theta_1\}$ such that equations (13.3), (13.4), (13.10) hold, $\tau_0 = \tau(n_0)$, $\tau_1 = \tau(n_1)$, and the function $x(\mu)$ is defined by (13.9), $\{\theta_0, \mu_0\}$ maximizes V^f.

Notice that $\tau_1 = \tau(n_1)$ at time 1, which implies that the political system cannot commit to future redistribution. This is a crucial ingredient in this theory for inefficient redistribution because it provides a reason for the farmers to wish to increase their numbers in period 1 to achieve greater political power.

For simplicity, assume;

$(1+\beta)(A-B) > 2\beta T$,

which implies that the maximum tax rate is small relative to the productivity differential between the two sectors, and ensures that it is not worthwhile to go into farming only to receive future transfers.

First consider the case in which $n_0 \leq n^-$, so that $\tau_0 = 0$, hence $\theta_0 = \mu_0 = 0$. In this case, there are too few farmers at date $t = 0$ for them to have any power, so there is no redistribution. As a result, there exists a unique equilibrium in which all young agents go into manufacturing. Specifically, with $\tau_0 = 0$, $\mu_0 = 0$, so Assumption 1 ensures that

$$W^f(\mu_0 = 0) \leq (1+\beta)B + \beta T < (1+\beta)A - \beta T \leq W^m(\tau_0 = 0).$$

Therefore:

Proposition 13.1. *Suppose Assumption 1 holds and $n_0 < n^-$, then there exists a unique equilibrium with $n_1 = (1-\delta)n_0$, $\tau_0 = \tau_1 = \theta_0 = \theta_1 = \mu_0 = 0$, and $x(\mu_0 = 0) = 0$.*

Next, consider the case where $n_0 > \frac{n^+}{1-\delta}$. Farmers have large enough numbers so that even when $x = 0$, they retain maximal power. Therefore, they choose $\tau_0, \tau_1, \theta_0, \theta_1$, and μ_0 to maximize V^f, which gives $\tau_0 = \tau_1 = T$, $\mu_0 = 0$, $\theta_t = \frac{(1-n_t)T}{n_t}$, for $t = 0, 1$.

Notice that in this case

$$W^f = (1+\beta)\,B + \beta\left[\frac{(1-n_1)T}{n_1}\right], \text{ and}$$
$$W^m = (1+\beta)A - \beta T,$$

Assumption 1 implies that $W^m > W^f$, and $x(\mu = 0) = 0$.

Proposition 13.2. *Suppose Assumption 1 holds and $n_0 > \frac{n^+}{1-\delta}$, then there exists a unique equilibrium such that $\tau_0 = \tau_1 = T$, $\mu_0 = 0$, $\theta_0 = \frac{(1-n_0)T}{n_0}$, $x(\mu_0 = 0) = 0$, $n_1 = (1-\delta)n_0$ and $\theta_1 = \frac{(1-(1-\delta)n_0)T}{(1-\delta)n_0}$*

In both cases discussed so far, the equilibrium maximizes output and *the form of redistribution is efficient.*

Although there is redistribution, no production or occupational decisions are distorted. The reason for this efficient form of redistribution is that political power is not contested. When $n_0 < n^-$, manufacturers have total political power and this can never be transferred to farmers. Similarly, when $n_0 > \frac{n^+}{1-\delta}$, farmers have maximal political power and always retain it, even if all newcomers were to go into manufacturing. This highlights the main conclusion of this model that inefficient redistribution will arise in order to control political power.

Now consider the case where

$$n^- < n_0 < \frac{n^+}{1-\delta}.$$

Farmers have some political power in period 0, and the extent of their political power at date 1 depends on the actions of newcomers. It is straightforward from the analysis in the above proposition that if $\mu_0 = 0$, newcomers will prefer to go into manufacturing. Therefore, farmers may want to use $\mu_0 > 0$, i.e. inefficient redistribution, in order to attract newcomers into farming, and increase their political power.

The utility of old farmers can be written as

$$V^f = (1+\beta)B + \theta_0 + \mu_0 + \beta\phi(n_1) \tag{13.11}$$

where

$$\phi(n_1) \equiv \frac{\tau(n_1)\,(1-n_1)}{n_1} \equiv \frac{\tau((1-\delta)n_0+\delta x)\,(1-(1-\delta)n_0-\delta x)}{(1-\delta)n_0+\delta x} \tag{13.12}$$

is per capita redistribution at $t=1$.

For the farmers to attract newcomers, they need to provide them with at least as much utility in farming as in manufacturing, hence

$$W^f \geq W^m,$$

where W^m and W^f are given by (13.7) and (13.8).

Define

$$U^f(x) = (1+\beta)B + \beta\phi(n_1) = W^f - \mu_0$$

as the utility of a new agent entering farming when a fraction x of newcomers enter farming and there is no inefficient redistribution (i.e., $\mu_0 = 0$).

Also define $U^m(x) = W^m$ as the utility of a new agent entering manufacturing when a fraction x of newcomers enter farming. Now $x > 0$ requires that $\mu_0 \geq U^m(x) - U^f(x)$ so as to convince newcomers to enter farming. Moreover, existing farmers would never want to pay more than necessary to newcomers, so focus on the case where

$$\mu_0 = U^m(x) - U^f(x) = (1+\beta)\,(A-B) - \beta\,(\phi(n_1) + \tau(n_1))\,.$$

Solving (13.3) for $\theta_0 + \mu_0$, the return to old farmers when they ensure that a fraction x of new farmers enter farming, the utility of all farmers, $\widehat{V}^f(x)$, is

$$\widehat{V}^f(x) = (1+\beta)B + \beta\phi(n_1) + \frac{(1-\delta)(1-n_0)\tau_0 - \delta x\left[U^m(x) - U^f(x)\right]}{(1-\delta)n_0}. \tag{13.13}$$

Let $\overline{V}^f$ be their utility when $\mu_0 = 0$. Notice that $\widehat{V}^f(x=0) = \overline{V}^f$ because when $\mu_0 = 0$ and no new born agents are entering farming, $x = 0$, so the fact that $\mu_0 = U^m(x) - U^f(x)$ does not matter.

Whether farmers prefer to use inefficient methods of redistribution, and so attract newcomers, depends on

$$\frac{d\widehat{V}^f(x=0)}{dx} = \delta\left(\beta\phi'((1-\delta)\,n_0) - \frac{U^m(x=0) - U^f(x=0)}{(1-\delta)n_0}\right). \tag{13.14}$$

The first term in parenthesis is the benefit of attracting some of the newcomers, while the second term is the cost of doing so per existing farmer.

If this expression, (13.14), is positive, then the utility of old farmers can be increased by attracting some of the young agents into farming.

In this case farmers will design the redistribution system to be inefficient specifically to increase their numbers.

Notice that farmers will only want to use inefficient redistribution when increasing their numbers leads to greater per capita transfers, $\phi(n_1)$. This implies that taxes imposed on manufacturers should increase sufficiently in n_1 to ensure greater transfers to farmers.

PROPOSITION 13.3. *If*

$$\phi'((1-\delta)\,n_0) > \frac{1}{\beta\,(1-\delta)\,n_0}[U^m(x=0) - U^f(x=0)] \tag{13.15}$$

then there will be inefficient redistribution, i.e. $\mu_0 > 0$*. In equilibrium,*

$$\mu_0 = U^m(x^*) - U^f(x^*),$$

and a fraction x^* *of newcomers enter farming such that*

$$\beta\,((1-\delta)n_0 + \delta x^*)\,\phi'((1-\delta)\,n_0 + \delta x^*) - \mu_0 + \beta\delta x^*\tau'\,((1-\delta)n_0 + \delta x^*) = 0, \tag{13.16}$$

or $x^* = 1$ *if (13.16)>0 when evaluated at* $x^* = 1$*.*

The second part follows by noting that (μ_0, x^*) are chosen to maximize (13.13). Substituting for $\mu_0 = U^m(x^*) - U^f(x^*)$, $dU^m(x^*)/dx = -\beta\tau'\,(n_1)$, and $dU^f(x^*)/dx = \beta\phi'\,(n_1)$, and simplifying gives (13.16).

Inefficient redistribution arises because it keeps power in the hands of the farmers. It achieves this because it rewards potential farmers, not only those who are already locked into farming. Expressed differently, because θ_0 is a "lump-sum" transfer, it does not distort the decisions of marginal agents. Precisely for this reason, however, the political process may choose to redistribute via μ_0 not θ_0.

Finally, consider a comparative static implication of this model regarding specificity of factors.

To do this, consider the analogous situation where, rather than young agents that can potentially enter farming, there are existing farmers that can leave. The same logic as before suggests that inefficient redistribution may be useful to keep these farmers in farming, and hence maintain the power of the farming lobby.

Now consider a situation which these agents have very specific skills that are useful for farming. In this case, they are unlikely to leave, and there is little need for inefficient redistribution.

Alternatively, imagine the situation in which they have no specific skills. Now inefficient redistribution is necessary to keep them in farming.

Therefore, the prediction of this approach is that there should be more inefficient redistribution when factors are less specific.

This comparative static seems to be in line with the evidence, where sectors that receive a lot of subsidies do not appear to have very specific skills (e.g., textiles, farming etc.). Interestingly, it is the opposite of the predictions other approaches, which suggest that there should be more redistribution (and hence perhaps more inefficient redistribution), when there are more specific skills.

13.5. References

(1) Acemoglu, Daron and Robinson, James A. (2001) "Inefficient Redistribution," *American Political Science Review*, 95, 649-662.

(2) Bates, Robert H. (1981) *Markets and States in Tropical Africa*, University of California Press, Berkeley CA.

(3) Becker, Gary S. and Casey Mulligan (1998) "Deadweight Costs and the Size of Government," NBER Working Paper #6789.

(4) Coate, Stephen and Stephen Morris (1995) "On the Form of Transfers to Special Interests," *Journal of Political Economy*, 103, 1210-1235.

(5) Dixit, Avinash K., Gene Grossman and Elhanan Helpman (1997) "Common Agency and Coordination," *Journal of Political Economy*, 105, 752-769.

(6) Lizzeri, Alessandro and Nicola Persico (2001) "The Provision of Public Goods Under Alternative Electoral Incentives," *American Economic Review*, 91, 225-245.

(7) Persson, Torsten and Guido Tabellini (1999) "The Size and Scope of Government: Comparative Politics with Rational Politicians," *European Economic Review*, 43, 699-735.

(8) Robinson, James A. and Ragnar Torvik (2005) "White Elephants," *Journal of Public Economics*, 89, 197-210.

(9) Rodrik, Dani (1986) "Tariffs, Subsidies and Welfare with Endogenous Policy," *Journal of International Economics*, 21, 285-299.

(10) Rodrik, Dani (1995) "The Political Economy of Trade Policy," in G. Grossman and K. Rogoff (eds.), Handbook of International Economics, vol. 3, Amsterdam; North-Holland.

(11) Stiglitz, Joseph (1998) "The Private uses of Public Interests: Incentives and Institutions," *Journal of Economic Perspectives*, 12, 3-22.

(12) Wilson, John D. (1990) "Are Efficiency Improvements in Government Transfer Policies Self-Defeating in Political Equilibrium," *Economics and Politics*, 2, 241-258.

CHAPTER 14

Political Economy of States

An issue in the background of much of the discussion so far is the political economy role of the state. According to Max Weber's famous approach, the state is the agency and society distinguished by its "monopoly over legitimate violence". In other words, it is the agency with the ability to force individuals into various actions that may be ex post costly for them. This coercive ability can be used for many different purposes, for the benefit of some subgroups (for example the elites), or to carry out useful redistribution. These are the issues we briefly discuss now.

14.1. The Role of the State in Economic Development and in Economics

There is a large literature in political science and sociology which emphasizes the role of the state in development. In particular, this literature studies where the institutional structure of the state comes from and it emphasizes that this process of institution building takes different forms in different societies. This literature draws a theory of comparative development from this, arguing that underdevelopment stems from having an incompetent state that is unable to create social order, discipline politicians or provide public goods. In some sense this literature looks at the same set of facts we have been discussing, but analyzes them from a different angle. For example, let's go back to President Mobutu in the Congo. It is true that he ran the country for his personal benefit, but it is also true that he did not really have a strong state apparatus capable of controlling society. He bought people off, divided and ruled, extemporized, and could not trust his army. He could not raise income or wealth taxes and certainly did not have a bureaucracy capable of systematically doing so or, for instance, promoting industrialization (even if he had wished to do so). Scholars in the literature on state formation see all of these things, including the amazing personalization of political power in the Congo, as manifesting the absence of state formation. They point out that in the middle ages, Europe was not so different, but then a process of institutional creation took place.

The scholar most associated with these ideas is Charles Tilly (2000). Tilly emphasized that historically it was inter-state competition and warfare that led to the creation of 'strong states.' Strong states are those with the capacity to get things done, raise taxes, build armies, provide public goods. The origins of institutions able to do this is warfare, but a beneficial side effect is that these are the states that can also induce economic development.

One can turn Tilly's argument on its head and try to explain why the factors he emphasized as driving state institution building in Europe did not create similar institutions in the Third World. Influential books along these lines are Herbst (2000) on Africa or Centeno (2002) on Latin America. According to this view, African or Latin American countries are poor and disorderly because historically they did not develop the same state institutions as Europe.

This literature has been very influential in the literature on the success of East Asian countries. The first seminal book on this by Chalmers Johnson (1982) stressed that Japanese economic success in the 20th century was driven by interventions by its very strong and efficiency government bureaucracy. Other scholars soon began to argue that this was a key to the Asian miracle, but Evans (1989, 1995) was the first person to provide a historical account of why Asian countries had state capacity (see Kohli, 2004, for a recent attempt at the same issue).

This is an interesting perspective and a fascinating literature with a lot of historical information and ideas but it has yet to be formalized. One sees loose applications of it in many domains. For example, the amazing economic success of Botswana since independence seems to be closely related to a process of state formation and building institutions (see Acemoglu, Johnson and Robinson, 2003).

14.2. Weak Versus Strong States

We now try to make sense of a few of these ideas by writing down a simple model. In the above models, there was an issue of controlling the politicians or the state, but the productive roles of the government and society were not exposed to the model. Introducing these leads to an interesting trade-off, which is the focus now.

Consider the following infinite horizon economy. Time is discrete and indexed by t. There is a set of citizens, with mass normalized to 1, and a ruler. All agents discount the future with the discount factor β, and have the utility function

$$u_t = \sum_{j=0}^{\infty} \beta^j \left[c_{t+j} - e_{t+j} \right], \tag{14.1}$$

where c_{t+j} is consumption and e_{t+j} is investment (effort), and we assume that the ruler incurs no effort cost.

Each citizen i has access to the following Cobb-Douglas production technology to produce the unique final good in this economy:

$$y_t^i = \frac{1}{1-\alpha} A_t^{\alpha} \left(e_t^i\right)^{1-\alpha}, \tag{14.2}$$

where A_t denotes the level of public goods (e.g., the state of the infrastructure, or the degree of law and contract enforcement between private citizens), at time t. The level of A_t will be determined by the investment of the ruler as described below. The important point captured by the specification in (14.2) is that a certain degree of state investment in public goods, the infrastructure or law-enforcement is necessary for private citizens to be able to function productively, and in fact, investment by the state is complementary to the investments of the citizens.

The ruler sets a tax rate τ_t on income at time t. Also, each citizen can decide to hide a fraction z_t^i of his output, which is not taxable, but hiding output is costly, so a fraction δ of it is lost in the process. This formulation with an economic exit option for the citizens is a convenient, though reduced-form, starting point.

Given a tax rate τ_t, the consumption of agent i is:

$$c_t^i \leq \left[(1-\tau_t)\left(1-z_t^i\right) + (1-\delta) z_t^i\right] y_t^i, \tag{14.3}$$

where tax revenues are

$$T_t = \tau_t \int \left(1 - z_t^i\right) y_t^i di. \tag{14.4}$$

The ruler at time t decides how much to spend on A_{t+1}. I assume that

$$A_{t+1} = \left[\frac{(1-\alpha)\phi}{\alpha} G_t\right]^{1/\phi} \tag{14.5}$$

where G_t denotes government spending on public goods, and $\phi > 1$, so that there are decreasing returns in the investment technology of the ruler (a greater ϕ corresponds to greater decreasing returns). The term $\left[(1-\alpha)\phi/\alpha\right]^{1/\phi}$ is included as a convenient normalization. In addition, (14.5) implies full depreciation of A_t, which simplifies the analysis below. The consumption of the ruler is whatever is left over from tax revenues after his expenditure and transfers,

$$c_t^R = T_t - G_t.$$

The timing of events within every period is as follows:

- The economy inherits A_t from government spending at time $t-1$.
- Citizens choose their investments, $\left\{e_t^i\right\}$.
- The ruler decides how much to spend on next period's public goods, G_t, and sets the tax rate τ_t.

- Citizens decide how much of their output to hide, $\{z_t^i\}$.

14.2.1. The First-Best Allocation. The first best allocation maximizes net output:

$$NY_0 = \sum_{t=0}^{\infty} \beta^t \left[\int \left(\left(1 - z_t^i\right) + (1-\delta) z_t^i \right) \left(\frac{1}{1-\alpha} A_t^{\alpha} \left(e_t^i\right)^{1-\alpha} - e_t^i \right) di - \frac{\alpha}{(1-\alpha)\phi} A_{t+1}^{\phi} \right].$$

Total surplus subtracts from output both the investment costs of citizens and that of the ruler, and is invariant to the distribution of output and consumption; hence taxes do not feature in this expression.

It is straightforward to see that the first-best allocation involves $z_t^i = 0$ for all i and t (i.e., no output is hidden) and $e_t^i = e_t^{fb} = A_t$. Substituting this into (14.2) gives output as: $y_t^{fb} = \frac{1}{1-\alpha} A_t$. The optimal level of public goods is $A_t = \beta^{1/(\phi-1)}$. Consequently, the first-best allocation is characterized by $e_0^{fb} = A_0$, and for all $t > 0$:

$$e_t^{fb} = \beta^{1/(\phi-1)} \text{ and } y_t^{fb} = \frac{1}{1-\alpha} \beta^{1/(\phi-1)}.$$

14.2.2. Markov Perfect Equilibrium. We now characterize the Markov Perfect Equilibrium (MPE) of this game. An MPE is defined as a set of strategies at each date t, $\left(\{e_t^i\}, \tau_t, \{z_t^i\}, G_t\right)$, such that these strategies only depend on the current (payoff-relevant) state of the economy, A_t, and on actions taken before within the same date according to the timing of events above. Thus, an MPE is given by a set of strategies $\left(\{e^i(A_t)\}, \tau(A_t), \{z^i(A_t)\}, G(A_t)\right)$.

Let us start with the decisions to hide. Given the structure of the game and the focus on MPE, actions are taken simply to maximize current income, so

$$z_t^i \begin{cases} = 1 & \text{if } \tau_t > \delta \\ \in [0,1] & \text{if } \tau_t = \delta \\ = 0 & \text{if } \tau_t < \delta \end{cases}. \tag{14.6}$$

Given (14.6), the optimal tax rate for the ruler is

$$\tau_t = \delta. \tag{14.7}$$

Next, investment decisions will maximize the utility of citizens, (14.1) subject to (14.3). The Markov structure implies that this is equivalent to maximizing the current period returns, $(1-\tau_t) y_t^i - e_t^i$, thus: $e_t^i = (1-\tau_t)^{1/\alpha} A_t$. Individual investments are therefore decreasing in the tax rate τ_t because higher taxes reduce their net returns, and are increasing in the level of infrastructure, A_t, because a better infrastructure raises the marginal productivity of the producers.

Given the subgame perfect equilibrium tax rate implied by (14.7), we have

$$e_t^i = (1-\delta)^{1/\alpha} A_t. \tag{14.8}$$

Substituting (14.7) and (14.8) into (14.4), we obtain the equilibrium tax revenue as a function of the level of infrastructure as:

$$T(A_t) = \delta y_t = \frac{(1-\delta)^{(1-\alpha)/\alpha}\,\delta A_t}{1-\alpha}. \tag{14.9}$$

Finally, the ruler will choose public investment, G_t to maximize his consumption. To characterize this, it is useful to write the Bellman equation for the discounted net present value of the ruler, denoted by $V(A_t)$. This takes the standard form:

$$V(A_t) = \max_{A_{t+1}} \left\{ T(A_t) - \frac{\alpha}{(1-\alpha)\phi} A_{t+1}^{\phi} + \beta V(A_{t+1}) \right\}, \tag{14.10}$$

which simply follows from writing the discounted payoff of the ruler recursively, after substituting for his consumption, c_t^R, as equal to taxes given by (14.9) minus his spending on the infrastructure, (14.5).

Since, for $\phi > 1$, the instantaneous payoff of the ruler is bounded, continuously differentiable and strictly concave in A, by standard arguments, the value function $V(\cdot)$ is strictly concave and continuously differentiable. Hence, the first-order condition of the ruler in choosing A_{t+1} can be written as:

$$\frac{\alpha}{1-\alpha} A_{t+1}^{\phi-1} = \beta V'(A_{t+1}), \tag{14.11}$$

which links the marginal cost of greater investment in infrastructure for next period to the greater value that will follow from this. Now differentiating (14.10) with respect to A_t:

$$V'(A_t) = T'(A_t) = \frac{(1-\delta)^{(1-\alpha)/\alpha}\,\delta}{1-\alpha}. \tag{14.12}$$

The value of better infrastructure for the ruler is the additional tax revenue that this will generate, which is given by the expression in (14.12).

Combining these conditions, we obtain the unique Markov Perfect Equilibrium choice of the ruler as:

$$A_{t+1} = A[\delta] \equiv \left(\frac{\beta (1-\delta)^{\frac{1-\alpha}{\alpha}}\,\delta}{\alpha} \right)^{\frac{1}{\phi-1}} \text{ and } G_t = \frac{\alpha}{(1-\alpha)\phi} (A[\delta])^{\phi}, \tag{14.13}$$

which also defines $A[\delta]$, an expression that will feature prominently in what follows. Substituting (14.13) into (14.10) yields a simple form of the ruler's value function:

$$V^*(A_t) = \frac{(1-\delta)^{(1-\alpha)/\alpha}\,\delta A_t}{1-\alpha} + \frac{\beta(\phi-1)(1-\delta)^{(1-\alpha)/\alpha}\,\delta}{(1-\beta)(1-\alpha)\phi} A[\delta], \tag{14.14}$$

which is also useful for future reference. Note that the value of the ruler depends on the current state of infrastructure, A_t, which he inherits from the previous period, and from this point on, the equilibrium involves investment levels given by (14.8) and (14.13).

PROPOSITION 14.1. *There exists a unique MPE where, for all t,* $\tau_t(A_t) = \delta$, $G(A_t)$ *is given by (14.13), and, for all i and t,* $z^i(A_t) = 0$ *and* $e^i(A_t)$ *is given by (14.8). The equilibrium level of aggregate output is:*

$$Y_t = \frac{1}{1-\alpha}(1-\delta)^{(1-\alpha)/\alpha} A[\delta]$$

for all $t > 0$ *and*

$$Y_0(A_0) = \frac{1}{1-\alpha}(1-\delta)^{(1-\alpha)/\alpha} A_0.$$

What is the level of δ that maximizes output? This "second-best" level of δ is the solution to the simple maximization problem:

$$\max_{\delta} Y_t(\delta) = \frac{1}{1-\alpha}(1-\delta)^{(1-\alpha)/\alpha} A[\delta], \tag{14.15}$$

where $A[\delta]$ is given by (14.13). The output maximizing level of the economic power of the state, denoted δ^*, is

$$\delta^* = \frac{\alpha}{\phi(1-\alpha)+\alpha}. \tag{14.16}$$

If the economic power of the state is greater than δ^*, then the state is too powerful, and taxes are too high relative to the output-maximizing benchmark. This corresponds to the standard case that the political economy literature has focused on. In contrast, if the economic power of the state is less than δ^*, then the state is not powerful enough for there to be sufficient rents in the future to entice the ruler to invest in public goods (or in the infrastructure, law-enforcement etc.). This corresponds to the case that the political science literature has identified as "the problem of weak states". Notice an important difference, however. The standard arguments on the problem of weak states basically ignore the self-interested behavior of the rulers and agents controlling the state. In contrast, here, the problem arises because with only limited power of the state to raise taxes in the future, the ruler has no interest in increasing the future productive capacity of the economy.

The expression for δ^* is intuitive. For example, δ^* is an increasing function of α. This is because, from the production function (14.2), a greater α implies that the investment of the ruler is more important relative to the investments of the citizens. Thus the ruler should receive a greater fraction of the ex post rents to encourage him to invest further. δ^* is also decreasing in ϕ, which, recall, corresponds to the degree of decreasing returns in the public good technology. Greater decreasing returns imply that the investment of the ruler is less sensitive to his ex post share of the revenues, and thus the optimal division of economic strength in society will give more weight to the citizens.

It can also be verified that the value of δ that maximizes the ruler's utility is $\delta = \alpha$. When $\delta > \alpha$, taxes are so high that net tax revenues fall.

Proposition 14.2. *Let δ^* be the output-maximizing level of δ, then* $0 < \delta^* < \alpha < 1$.

14.2.3. Political Power. Let us now use this model to look into issues of balance of political power. We modify the baseline model as follows: now there is a large set of identical potential rulers, and citizens decide whether to replace the current ruler, denoted by $R_t \in \{0, 1\}$. After replacement, the existing ruler receives 0 utility, and citizens reclaim the tax revenue and redistribute it to themselves as a lump sum transfer, S_t. At the time of replacement, the public goods spending of the ruler, G_t, is already sunk, and citizens take back a fraction $\eta \in (0, 1]$ of the tax revenue, T_t, so the consumption of the ruler is $c_t^R = (1 - \eta) T_t - G_t$. Replacement is costly, however, and at time t citizens face a cost of replacing the current ruler with a new ruler equal to $\theta_t A_t$, where θ_t is a nonnegative random variable with a continuous distribution function $\tilde{F}_\lambda$, with (finite) density $\tilde{f}_\lambda$. The cost is multiplied by A_t to ensure that the level of public goods does not have mechanical effect on replacement. This formulation is as simple way of introducing stochastic replacements of the ruler along the equilibrium path.

Let us assume:

$$\text{(A1)} \qquad \frac{\tilde{f}_\lambda (x)}{1 - \tilde{F}_\lambda (x)} \text{ is nondecreasing in } x \text{ and } \tilde{F}_\lambda (0) < 1,$$

which is the standard monotone hazard (or log concavity) assumption. This type of assumption is necessary to get meaningful comparative statics in the majority of models that have realization of some uncertainty that matters for equilibrium behavior.

The timing of events in this endogenous replacement game can be summarized as:

- The economy inherits A_t from government spending at time $t - 1$.
- Citizens choose their investments, $\{e_t^i\}$.
- The ruler decides how much to spend on next period's public goods, G_t, and sets the tax rate τ_t.
- Citizens decide how much of their output to hide, $\{z_t^i\}$.
- θ_t is realized.

Citizens choose R_t. If $R_t = 1$, the current ruler is replaced and the tax revenue is redistributed to the citizens as a lump-sum subsidy $S_t = \eta T_t$.

An MPE is defined similarly to before, as a set of strategies at each date t, $(\{e_t^i\}, \tau_t, \{z_t^i\}, R_t, G_t)$, such that these strategies only depend on the current state of the economy, A_t, and on actions taken before within the same date according to the timing of events above. Thus, it can be summarized by a set of strategies $\left(\left\{e^i\left(A_t\right)\right\}, \tau\left(A_t\right), \left\{z^i\left(A_t\right)\right\}, R\left(A_t\right), G\left(A_t\right)\right)$. In addition, it is convenient to focus on a steady-state MPE where $A_t = A_{t+\ell}$ for all $\ell \geq 0$.

Also assume that

$$\delta \in \left(\delta^*, \alpha\right), \tag{A2}$$

where δ^* is given by (14.16). This assumption ensures that taxes are always less than the value α that maximizes ruler utility, and also allows the potential for excessively high taxes (i.e., $\tau > \delta^*$).

Citizens' hiding decisions are still given by the privately optimal rule, (14.6). Moreover, in the MPE, they will replace the ruler, i.e., $R_t = 1$, whenever

$$\theta_t < \frac{\eta T_t}{A_t}. \tag{14.17}$$

Intuitively, in the MPE replacing the ruler has no future costs or benefits (since all future rulers condition their strategies only on the pay off-relevant state variable, A_t), so it it is in the citizens' interest to replace the ruler when the immediate benefit, ηT_t, exceeds the cost, $\theta_t A_t$. The important substantive implication of (14.17) is that greater taxes will lead to a higher likelihood of ruler replacement.

Condition (14.17) immediately implies that the probability that the ruler will be replaced is $\tilde{F}_\lambda\left(\eta T_t / A_t\right)$. To simplify the notation, define

$$\mathcal{T}\left(\tau_t\right) = \frac{\left(1-\tau_t\right)^{(1-\alpha)/\alpha} \tau_t}{1-\alpha}, \tag{14.18}$$

so that $T_t = \mathcal{T}\left(\tau_t\right) A_t$. Let us also parameterize the distribution function as $\tilde{F}_\lambda\left(x/\eta\right) = \lambda F\left(x\right)$ for some continuous distribution function F with (finite) density f, which will be useful both to simplify notation and for comparative static exercises below. This assumption implies that the probability that the ruler will be replaced is $\lambda F\left(T_t/A_t\right)$. The monotone likelihood ratio or log concavity properties of $\tilde{F}_\lambda$ clearly carry over to F.

The relevant value function for the ruler can then be written as:

$$\begin{aligned} V\left(A_t\right) = \max_{\tau_t \in [0,\delta], A_{t+1}} \quad & \{ \left(1-\lambda F\left(\mathcal{T}\left(\tau_t\right)\right)\right) \left(\mathcal{T}\left(\tau_t\right) A_t - \frac{\alpha}{\phi(1-\alpha)} A_{t+1}^{\phi}\right) \\ & + \beta\left(1-\lambda F\left(\mathcal{T}\left(\tau_t\right)\right)\right) V\left(A_{t+1}\right) \}. \end{aligned} \tag{14.19}$$

Now the ruler's maximization problem involves two choices, τ_t and A_{t+1}, since taxes are no longer automatically equal to the maximum, δ. In this choice, the ruler takes into account that a higher tax rate will increase the probability of replacement. The first-order condition with respect to τ_t yields:

$$\frac{\partial \mathcal{T}(\tau_t)}{\partial \tau_t}\left[\left(1-\lambda F\left(\mathcal{T}(\tau_t)\right)\right)-\lambda f\left(\mathcal{T}(\tau_t)\right)\left(\mathcal{T}(\tau_t)-\frac{G_t}{A_t}+\beta\frac{V\left(A_{t+1}\right)}{A_t}\right)\right]\geq 0, \tag{14.20}$$

and $\tau_t \leq \delta$ with complementary slackness, where recall that
$G_t = \alpha A_{t+1}^{\phi}/\left(\phi(1-\alpha)\right)$. Assumption A2 implies that $\tau < \alpha$, so $\partial \mathcal{T}(\tau_t)/\partial\tau_t > 0$ and in an interior equilibrium the term in square brackets has to be equal to zero. In other words, the additional expected revenue brought by higher taxes (the first term in square brackets) must be balanced by higher probability of losing these taxes and the continuation value (the second term in square brackets).

Assuming that $V(A_{t+1})$ is differentiable in A_{t+1}, the first-order condition for A_{t+1} is still given by (14.11) above (since the term $\left(1-\lambda F\left(\mathcal{T}(\tau_t)\right)\right)$ cancels from both sides). The expression for $V'(A_{t+1})$ again follows from the envelope condition,

$$V'\left(A_{t+1}\right)=\left(1-\lambda F\left(\mathcal{T}(\tau_{t+1})\right)\right)\mathcal{T}(\tau_{t+1}). \tag{14.21}$$

It only differs from the corresponding condition above, (14.12), because with probability $\lambda F\left(\mathcal{T}(\tau_{t+1})\right)$, the ruler will be replaced and will not enjoy the increase in future tax revenues.

Using this, the first-order condition with respect to A_{t+1} implies that in an interior equilibrium:

$$A_{t+1}=A\left[\tau_{t+1}\right]\equiv\left(\alpha^{-1}\beta\left(1-\lambda F\left(\mathcal{T}(\tau_{t+1})\right)\right)\left(1-\tau_{t+1}\right)^{\frac{1-\alpha}{\alpha}}\tau_{t+1}\right)^{\frac{1}{\phi-1}}.$$

The optimal value of A_{t+1} for the ruler depends on τ_{t+1} since, from the envelope condition, (14.21), the benefits from a higher level of public good are related to future taxes.

To make further progress, let us focus on the steady state equilibrium where $\tau_{t+\ell}=\tau^*$ for all $\ell\geq 0$ (which follows since in steady state $A_t = A_{t+\ell}$). Hence

$$A\left[\tau^*\right]\equiv\left(\alpha^{-1}\beta\left(1-\lambda F\left(\mathcal{T}(\tau^*)\right)\right)\left(1-\tau^*\right)^{\frac{1-\alpha}{\alpha}}\tau^*\right)^{\frac{1}{\phi-1}}. \tag{14.22}$$

In addition, we have $G\left[\tau^*\right]=\beta\mathcal{T}(\tau_t)A\left[\tau^*\right]/\phi$. Therefore, the value function for the ruler in steady state can be written as $V\left(A\left[\tau^*\right]\right)=\left(1-\lambda F\left(\mathcal{T}(\tau^*)\right)\right)v\left[\tau^*\right]A\left[\tau^*\right]$, where

$$v\left[\tau^*\right]\equiv\mathcal{T}(\tau^*)+\frac{(\phi-1)}{\phi}\frac{\beta\left(1-\lambda F\left(\mathcal{T}(\tau^*)\right)\right)\mathcal{T}(\tau^*)}{\left(1-\beta\left(1-\lambda F\left(\mathcal{T}(\tau^*)\right)\right)\right)}. \tag{14.23}$$

Now using (14.19), (14.20), (14.23), and the fact that in steady state $\tau_{t+\ell}=\tau^*$ for all $\ell\geq 0$, we obtain the following equation for an interior steady-state equilibrium tax rate, τ^*:

$$\lambda f\left(\mathcal{T}(\tau^*)\right)v\left[\tau^*\right]-\left(1-\lambda F\left(\mathcal{T}(\tau^*)\right)\right)=0. \tag{14.24}$$

This equation is intuitive. The first term is the cost of a unit increase in $T(\tau)$. This increase reduces the probability of staying in power by an amount equal to $\lambda f\left(\mathcal{T}(\tau)\right)$. This is multiplied by the (normalized) value of staying power, $v[\tau^*]$, (since $v[\tau^*] = \mathcal{T}(\tau^*) - G(A[\tau^*])/A[\tau^*] + \beta V(A[\tau^*])/A[\tau^*]$). The second term is the benefit of a unit increase in tax revenue, which the ruler receives with probability $1 - \lambda F(\mathcal{T}(\tau))$. Note that $\tau^* = 0$ can never be a solution to this equation, since $\beta v[0] = \mathcal{T}(0) = 0$. Therefore, there will be an interior solution as long as

$$\lambda f(\mathcal{T}(\delta))\, v[\delta] > 1 - \lambda F(\mathcal{T}(\delta)). \tag{14.25}$$

If, on the other hand, (14.25) does not hold, then the equilibrium will be a corner solution with $\tau = \delta$ and $A[\tau = \delta]$ This establishes the existence of an MPE. To establish uniqueness, we need to impose an additional condition:

$$\left(1 - \frac{\beta}{\phi}(1 - \lambda F(0))\right)^2 - (\phi - 1)\frac{\beta}{\phi}(1 - \lambda F(0)) > 0. \tag{A3}$$

This assumption requires $\beta(1 - \lambda F(0))$ not to be too large, and can be satisfied either if β is not too close to 1 or if $\lambda F(0)$ is not equal to zero. Intuitively, if $\beta(1 - \lambda F(\tau^*))$ is close to 1, $v[\tau^*]$ can be very large, creating a non-monotonicity in (14.24). Assumption (A3) is sufficient to ensure that this is not the case, so the left-hand side of (14.24) is everywhere increasing and there is a unique equilibrium (see the proof of Proposition **??**). This observation, combined with the mop on hazard ratio assumption, Assumption A1, also leads to unambiguous comparative static results:

Proposition 14.3. *Suppose (A1), (A2) and (A3) hold. Then, in the endogenous replacement game of this section, there exists a unique steady-state MPE.*

- *In this equilibrium, if (14.25) does not hold all rulers set $\tau = \delta$ and $A[\tau = \delta]$. If (14.25) holds, then they set $\tau^* < \delta$ given by (14.24), and $A[\tau^*]$ given by (14.22). Citizens replace a ruler whenever $\theta_t < \eta T(\tau^*)$ where $T(\cdot)$ is given by (14.18).*
- *In this equilibrium, $\partial\tau^*/\partial\lambda \le 0$, $\partial\tilde{A}/\partial\lambda \le 0$, $\partial\tau^*/\partial\delta \ge 0$ and $\partial\tilde{A}/\partial\delta \ge 0$.*
- *There also exists $\lambda^* \in (0, \infty)$ such that output is maximized when $\lambda = \lambda^*$.*

The most important result in this proposition is that, similar to the case of the economic power of the state, there is an optimal level of the political power of the state. Intuitively, when $\lambda < \lambda^*$, the state is too powerful and taxes are too high and citizens' investments are too low. When $\lambda > \lambda^*$, the state is too weak and taxes and public investments are too low. The intuition is also related to the earlier result. When the state is excessively powerful, i.e., $\lambda < \lambda^*$, citizens expect high taxes and choose very low levels of investment (effort). In contrast, when $\lambda > \lambda^*$, the state is excessively weak and there is the reverse holdup problem; high taxes will encourage citizens to replace the ruler, and anticipating this, the ruler has little incentive to invest in public goods, because he will not be able to recoup the costs of current investment in public goods with future revenues. Therefore, this proposition reiterates the main insight from the analysis so far: there needs to be a balanced distribution of (both economic and political) power between the state and the citizens to encourage both parties to make investments in the productive resources of the society.

The proposition also establishes the equilibrium tax rate and public good spending are decreasing in λ (and increasing in δ). These are intuitive. A lower value of λ corresponds to a situation in which politicians are more entrenched and more costly to replace, thus politically more powerful. Since taxes are constrained by the political power of the citizens (i.e., their power to replace the ruler when taxes are high), a lower λ implies that the ruler will impose higher taxes, and consequently, will be willing to invest more in public goods.

14.2.4. Consensually-Strong States. The analysis so far focused on Markov Perfect Equilibria (MPEs), where the repeated nature of the game between the ruler and the citizens is not exploited. In this framework, weak states are costly because rulers are unable to impose high taxes and do not have sufficient incentives to invest in public goods. However, when the state is politically weak, in the sense that the politician in power can be replaced easily, a consensus between state and society can develop whereby citizens will tolerate high taxes (and will not replace the government because of these high taxes) as long as a sufficient fraction of the proceeds are invested in public goods. We refer to this as a "consensually-strong state," and now investigate how a consensually-strong state can arise as a subgame perfect equilibrium in the game with endogenous replacement of politicians.

An analysis of consensually-strong states is interesting not only to relax the restriction to MPE (which may not be warranted given the repeated interaction between the ruler and the citizens), but also because the concept of a consensually-strong state might be useful in providing us with a simple framework to think about state-society relations in many developed countries where governments can be replaced relatively easily, and nonetheless impose relatively high taxes, but then spend a high fraction of the proceeds on public goods. Such an outcome appears difficult in the models focusing on MPE; if δ is high or λ is low, the government imposes high taxes, but consumes a high fraction of the proceeds. We will see that in the "consensually-strong state" equilibrium being studied here, the pattern with high taxes and relatively high investments in public goods will emerge as the equilibrium when both δ and λ are high (also when the discount factor β is high).

A subgame perfect equilibrium is defined as a set of strategies that are best responses to each other given all histories, and the on-the-equilibrium-path behavior in this equilibrium

can be described as a set of strategies
$\left(\left\{e^i\left(A\right)\right\},\tau\left(A\right),\left\{z^i\left(A\right)\right\},R\left(A\right),G\left(A\right)\right)$. In addition, we will look at the steady state given the equilibrium strategies, which will be the case when $A_{t+\ell}=A_t$ for all $\ell>0$. We refer to the resulting equilibrium as a "steady state" subgame perfect equilibrium.

The focus here will be to characterize the best stationary subgame perfect equilibrium from the point of view of the citizens. Since all rulers are ex ante identical, the best such equilibrium should keep the ruler in power as long as he follows the implicitly-agreed strategy. Let us think of this equilibrium as a policy vector $(\tau,\tilde{A})$ such that as long as the ruler follows this policy vector, he will never be replaced, and his continuation value when he deviates is given by a credible punishment strategy of the citizens.

To simplify the analysis further, we first discuss the case where $\tilde{F}_\lambda=\tilde{F}_\lambda^*$ with $\tilde{F}_\lambda^*$ taking the following simple form: with probability $1-\lambda$, $\theta=\infty$, and with probability λ, $\theta=0$. This implies that with probability $1-\lambda$, citizens cannot replace the ruler, and with probability λ, they can do so without any costs.

Let $V^c(\tilde{\tau},\tilde{A}\mid A)$ be the value of the ruler in such an equilibrium where the current state is A, and all future taxes and public good investments are given by $(\tau,\tilde{A})$. We then have:

$$V^c(\tilde{\tau},\tilde{A}\mid A)=\frac{(1-\tilde{\tau})^{\frac{1-\alpha}{\alpha}}\tilde{\tau}A}{1-\alpha}-\frac{1}{1-\beta}\frac{\alpha\tilde{A}^{\phi}}{(1-\alpha)\phi}+\frac{\beta}{1-\beta}\frac{(1-\tilde{\tau})^{\frac{1-\alpha}{\alpha}}\tilde{\tau}\tilde{A}}{1-\alpha}. \tag{14.26}$$

This expression follows since the ruler will incur the cost of investing for $\tilde{A}$ in every period, while the benefits start with one period delay. Here the superscript c denotes "cooperation," and the form of this expression immediately follows from (14.14), incorporating the fact that future policies will be $(\tau,\tilde{A})$ and there is no replacement of the ruler.

In contrast, if the ruler decides to deviate from the implicitly-agreed policy $(\tau,\tilde{A})$, his continuation value will depend on the punishment strategies he expects. Recall that with probability $1-\lambda$ citizens are unable to replace the ruler ($\theta=\infty$), whereas with probability λ, they can replace the ruler without any cost. Since citizens cannot coordinate in their economic decisions, replacing the ruler with probability λ and then playing the MPE strategies is the worst (credible) punishment for the ruler. Anticipating replacement with this probability, the problem of the ruler is similar to that analyzed above. In particular, the ruler will always tax at the maximum rate, δ, and choose the level of public investment consistent with his own objectives (since following a deviation, the ruler is replaced with probability λ irrespective of the tax rate, he sets the highest possible tax rate, δ). Thus his deviation value as a function of the current state A and the tax expectation of the citizens, $\tilde{\tau}$, is given by

$$(14.27)\quad V^d(A \mid \tilde{\tau}) = \max_{A^d}\left\{\frac{(1-\lambda)(1-\tilde{\tau})^{\frac{1-\alpha}{\alpha}}\delta A}{1-\alpha} - \frac{\alpha}{(1-\alpha)\phi}\left(A^d\right)^{\phi} + \beta(1-\lambda)\tilde{V}^d(A^d)\right\}.$$

This expression takes into account that when the ruler deviates, he takes advantage of the fact that citizens invested expecting a tax rate of $\tilde{\tau} < \delta$, and then taxes them at the rate δ. Subsequently, he invests an amount A^d in the public good, consistent with his own maximization problem, and receives the MPE continuation value. An analysis similar to that before shows that this value is

$$\tilde{V}^d(A) = \frac{(1-\lambda)(1-\delta)^{\frac{1-\alpha}{\alpha}}\delta A}{1-\alpha} + \frac{\beta(1-\lambda)^2(\phi-1)(1-\delta)^{\frac{1-\alpha}{\alpha}}\delta A[\delta \mid \lambda]}{(1-\beta(1-\lambda))(1-\alpha)\phi},$$

with $A[\delta \mid \lambda]$ defined by:

$$(14.28)\qquad A[\delta \mid \lambda] \equiv \left(\frac{\beta(1-\lambda)^2(1-\delta)^{\frac{1-\alpha}{\alpha}}\delta}{\alpha}\right)^{\frac{1}{\phi-1}}.$$

Therefore, the deviation value of the ruler is:

$$(14.29)\qquad V^d(A \mid \tilde{\tau}) = \frac{(1-\lambda)(1-\tilde{\tau})^{\frac{1-\alpha}{\alpha}}\delta A}{1-\alpha} + \frac{\beta(1-\lambda)^2(\phi-1)(1-\delta)^{\frac{1-\alpha}{\alpha}}\delta A[\delta \mid \lambda]}{(1-\beta(1-\lambda))(1-\alpha)\phi}.$$

The ruler will follow the agreed policy, $(\tau, \tilde{A})$, if his incentive compatibility constraint,

$$(14.30)\qquad V^c(\tilde{\tau}, \tilde{A} \mid A) \geq V^d(A \mid \tilde{\tau}),$$

is satisfied. This incentive compatibility constraint requires that the ruler prefers the equilibrium strategy to deviating and taxing at the highest possible rate for his own consumption. It must also be in the interest of the citizens not to replace the ruler pursuing the implicitly-agreed policy. This also introduces an additional incentive compatibility constraint which we will ignore here for simplicity.

Then, the (stationary) equilibrium from the point of view of the citizens can be characterized as a policy combination $\left(\tilde{\tau}, \tilde{A}\right)$ that can be supported in steady state starting at a level of public good A is the solution to the following maximization problem:

$$(14.31)\qquad \max_{\tilde{\tau}, \tilde{A}} U^c\left(\tilde{\tau}, \tilde{A} \mid A\right)$$

subject to (14.30).

Since both the objective function (14.31) and the boundary of (14.30) are twice continuously differentiable in $(\tilde{\tau}, \tilde{A})$, we can analyzed the equilibrium by looking at the first-order

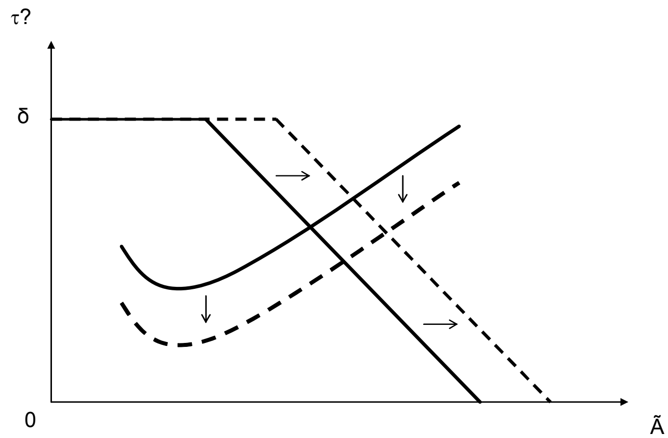


Figure 14.1

conditions:

$$\frac{(1-\tilde{\tau})^{\frac{1-\alpha}{\alpha}}\tilde{\tau}A}{1-\alpha}-\frac{1}{1-\beta}\frac{\alpha\tilde{A}^{\phi}}{(1-\alpha)\phi}+\frac{\beta}{1-\beta}\frac{(1-\tilde{\tau})^{\frac{1-\alpha}{\alpha}}\tilde{\tau}\tilde{A}}{1-\alpha} - \frac{(1-\lambda)(1-\tilde{\tau})^{\frac{1-\alpha}{\alpha}}\delta A}{1-\alpha}-\frac{\beta(1-\lambda)^2(\phi-1)(1-\delta)^{\frac{1-\alpha}{\alpha}}\delta A[\delta\mid\lambda]}{(1-\beta(1-\lambda))\phi(1-\alpha)}=0, \tag{14.32}$$

which represents the incentive compatibility constraint of the ruler, and the condition

$$\beta(1-\tilde{\tau})^{\frac{1-\alpha}{\alpha}}\left(\alpha+(1-\beta)\left(1-\lambda\right)\left(1-\alpha\right)\delta\right)=\alpha\tilde{A}^{\phi-1}, \tag{14.33}$$

if $\tilde{\tau}<\delta$ and $\beta(1-\tilde{\tau})^{\frac{1-\alpha}{\alpha}}\left(\alpha+(1-\beta)\left(1-\lambda\right)\left(1-\alpha\right)\delta\right)\geq\alpha\tilde{A}^{\phi-1}$ if $\tilde{\tau}=\delta$, which captures the trade-off between taxes and public good investments.

Equation (14.33) defines the locus of combinations of $(\tilde{\tau},\tilde{A})$ consistent with equilibrium when $\tilde{\tau}<\delta$. Since $\phi>1$, this locus is downward sloping. Therefore, in the accompanying figure, the thick broken line gives the combinations of $(\tilde{\tau},\tilde{A})$ consistent with equilibrium. These combinations also have to satisfy (14.32), which is drawn as an upward-sloping curve. Intuitively, if the ruler is required to invest more in public goods, taxes also need to increase to ensure incentive compatibility.

Consequently, an increase in λ, which corresponds to the state becoming politically weaker, leads to lower taxes (i.e., $\partial\tilde{\tau}/\partial\lambda<0$) and higher investments in public good (i.e., $\partial\tilde{A}/\partial\lambda>0$).

Proposition 14.4. *Consider the endogenous replacement game studied above, and suppose that* $\tilde{F}_\lambda = \tilde{F}^*_\lambda$ *(i.e.,* $\theta = \infty$ *with probability* $1-\lambda$*, and* $\theta = 0$ *with probability* λ*), Assumptions A1 and A2 hold, assume that* $\phi - 1 \geq 1 - \alpha$*, and let* $\beta^* \equiv \left(\sqrt{1+4\lambda} - 1\right)/2\lambda$*. Then for all* $\beta \geq \beta^*$*, a consensually-strong state equilibrium exists. In this equilibrium, the ruler always follows the policy* $(\tilde{\tau}, \tilde{A})$*, and is never replaced, and taxes are lower than in the MPE, i.e.,* $\tilde{\tau} \leq \delta$*. Moreover, as long as* $\tilde{\tau} < \delta$*, we have that when economic or political power of the state increases, investments in public goods decrease, i.e.,* $\partial\tilde{A}/\partial\lambda > 0$ *and* $\partial\tilde{A}/\partial\delta < 0$*.*

Therefore, the results with the consensually-strong state are very different from those in the previous two sections; in particular, as the economic or political power of the state decreases, investments in public goods increase, while the implications for the equilibrium tax rate are ambiguous. For example, when the state becomes politically less powerful (i.e., λ increases), the incentive compatibility constraint of the ruler, (14.32) shifts down as shown by the shift to the dashed curve in the above figure. Simultaneously, the curve for (14.33) shifts out (again to the dashed curve). Consequently, while $\tilde{A}$ increases, the implications for $\tilde{\tau}$ are ambiguous. Intuitively, when it becomes easier to control the ruler (because deviating from the agreed policy becomes less profitable for him), citizens demand greater investments in public goods, which may necessitate greater taxes to cover the public expenditures and the rents that the ruler needs to be paid to satisfy his incentive compatibility constraint. Similar results obtain in response to changes in the economic power of the state, δ. Interestingly, however, the comparative static with respect to δ need not hold when $\tilde{\tau} = \delta$; in this case, a decline in δ forces a lower tax rate, and investments in public goods may also need to decrease to satisfy the incentive compatibility constraint of the ruler.

These results enable us to envisage a situation similar to those in OECD countries, where the government imposes high taxes but also invests a high fraction of the proceeds in public goods. This would correspond to a high value of δ (otherwise, $\tilde{\tau} = \delta$ and taxes would be constrained to be low) and also a high value of λ (otherwise, the incentive compatibility constraint of the ruler would be excessively tight, and only low levels of investment in public goods can be supported). Naturally, for all of these outcomes the society also needs to coordinate on the consensually-strong state equilibrium, and the discount factor, β, needs to be sufficiently high.

This proposition can also be generalized to the case where the distribution of costs does not take the simple two-point form:

PROPOSITION 14.5. *Consider the endogenous replacement game, and suppose that Assumptions A1, A2 and A4 hold. Then there exists* $\beta^{**} < 1$ *such that for* $\beta \geq \beta^{**}$*, a consensually-strong state equilibrium exists. In this equilibrium, the ruler always follows the policy* $(\tilde{\tau}, \tilde{A})$*, and is never replaced. Moreover, as long as* $\tilde{\tau} < \delta$*, we have that when economic or political power of the state increases, investments in public goods decrease, i.e.,* $\partial \tilde{A}/\partial \lambda > 0$ *and* $\partial \tilde{A}/\partial \delta < 0$*.*

The reason why this generalization is important is that, contrary to Proposition 14.4, this proposition no longer states that the consensually-strong state tax rate is below the MPE tax rate. In fact, a simple example shows that this is no longer true. Take the case where $\lambda \to \infty$; the analysis above shows that in the MPE $\tau^* \to 0$ and $A[\tau^*] \to 0$, a very undesirable outcome from the point of view of the citizens. In contrast, with the consensually-strong state, the equilibrium tax rate always satisfies $\tilde{\tau} > 0$ (as long as $\alpha > 0$).

This result is of interest for the interpretation of an otherwise puzzling feature; OECD governments typically tax at higher rates than the governments of many less-developed countries. This analysis shows that this need not be because governments are "politically stronger" in these more developed polities. Instead, it might be the outcome of a consensually-strong state equilibrium where politically weak governments are allowed to impose high taxes as long as a sufficient fraction of the proceeds are invested in public goods. Interestingly, the analysis also highlights that even in the consensually-strong state equilibrium, the delivery of public goods comes with significant rents for the ruler; the incentive compatibility constraint necessitates that, despite its political weakness, the ruler receive sufficient rents so that he is not tempted to use the tax revenues for his own benefit. Therefore, the image of OECD-type governments that emerges from this model is one of politically weak, but economically strong states that are allowed to impose high taxes with the (credible) promise of delivering public services. Naturally, what makes this whole equilibrium possible is sufficient rents for the politicians.

14.3. The Formation of the State

Many scholars have argued that differences in state capacity are a key factor in comparative economic and political development. Although state capacity is multi-faceted, it inevitably relies on Weber's famous notion of the state as

> "a human community that (successfully) claims the *monopoly of the legitimate use of physical force* within a given territory".

States vary greatly in their capacities and whether or not they have such a monopoly of violence, and there is little evidence that this variation has decreased over the recent past. For example, in the 1990s the state in Somalia, Sierra Leone, Liberia, the Congo and Rwanda, completely collapsed and gave up any pretence of undertaking the tasks that we associate with states. In Latin America, Colombia, Peru, Guatemala, El Salvador, and Nicaragua have all recently experienced or are now experiencing prolonged civil wars, with the writ of the state being absent from large parts of the country. In Pakistan the central state in Islamabad has little control of the 'tribal areas' such as Waziristan. Similarly, the Iraqi state in Baghdad exercises little authority in Kurdistan.

Why do some states fail to establish this monopoly? The social science literature emphasizes several key ideas, for instance, the inability of states to establish such monopoly because of 'difficult geography' (Herbst, 2000), 'rough terrain' (Fearon and Laitin, 2003), or simply poverty (Fearon and Laitin, 2003). It has also suggested that inter-state competition and warfare and domestic political competition influence the incentives of politicians to build state capacity. Common to all of these explanations is a type of 'modernization' view, suggesting that as society modernizes and grows richer, state capacity will simultaneously develop. In particular 'state formation' involves eliminating armed actors and establishing a monopoly of violence, in the same way that after the Wars of the Roses the victorious Tudors disarmed the English aristocracy.

Yet several of the examples above are quite puzzling from this point of view. In the case of Pakistan, the tribal areas have existed since the formation of the country in 1947, and even though they have been largely out of the control of the central state, they have also been represented within it. Under the 1973 Constitution the tribal areas had 8 representatives in the National Assembly elected by the tribal elders, or the Maliks. Under General Musharraf's regime this was increased to 12. In Iraq, while the peshmerga militia control the streets of Mosul, a coalition of Kurdish political parties keeps the government in power in Baghdad. In Colombia, as much as one third of the legislature may have been elected in elections heavily influenced by armed paramilitary groups. After many of these were arrested by the Supreme Court, the Colombian President did little to stop their alternates from voting in their absence.

These examples point to a different path of state formation than the one taken by England under the Tudors and subsequently enshrined in the social science literature. Instead, they suggest that state formation can take place without a monopoly of violence being established. In this paper we develop a new perspective on state formation, emphasizing the idea that aspects of state weakness, particularly the lack of monopoly of violence in peripheral areas, can be an equilibrium outcome. Moreover, in contrast to the implicit notion common in the previous literature, 'modernization' need not automatically eradicate non-state armed actors. Here, let us consider how the monopoly emerges in the context of democratic politics, and then apply some of these ideas to Colombia.

Let us begin with the observation that in a democracy non-state armed actors (in our context, paramilitaries) can control citizens' voting behavior. Since paramilitaries naturally have preferences over policies, when they choose to become involved in politics, this reduces the incentives of the politicians they favor to eliminate them. The model predicts that in non-paramilitary areas policies are targeted at citizens while in paramilitary areas they cater to the preferences of paramilitaries. This implies that in paramilitary areas citizens obtain fewer public goods (and other policies they value). The model further implies that paramilitaries will tend to persist to the extent that they deliver votes to politicians they prefer—in the Colombian case, to President Álvaro Uribe—and that this effect is stronger in areas where these politicians would have otherwise not done as well. Thus non-state armed actors can persist because they can be in a symbiotic relationship with specific politicians holding power: paramilitaries deliver votes to politicians with preferences relatively close to theirs, while politicians they helped elect implicitly or explicitly support laws and policies that they prefer.

We empirically investigate the implications of our model using the recent Colombian experience, where two main non-state armed actors, the 'left-wing' guerrillas Fuerzas Armadas Revolucionarias de Colombia (FARC—The Revolutionary Armed Forces of Colombia) and 'right-wing' paramilitary forces, which in 1997 coalesced into the Autodefensas Unidas de Colombia (AUC—United Self-Defense Organization of Colombia), have shaped the recent political landscape. We first provide evidence that paramilitaries, though interestingly not the FARC, have systematically influenced electoral outcomes. In particular, after the AUC got involved in politics in 2001, the presence of paramilitaries in a municipality is correlated with the rise of non-traditional 'third parties' (that is, parties other than the Liberals, the Conservatives, and the Socialists), which are widely recognized to be often directly or indirectly associated with the paramilitaries. We also find that paramilitary presence is also associated with a greater concentration of votes within a municipality in legislative elections and with greater support for President Álvaro Uribe, who has enacted several key policies in line with the preferences of the paramilitaries, in the presidential elections.

The effect of paramilitaries on the elections is further substantiated by the fact that when a senator's list receives a greater proportion of its votes in areas with high paramilitary presence, the senator is more likely to be subsequently arrested for illegal connections with paramilitaries and to support the two clauses of the Justice and Peace Law that were highly lenient towards the paramilitaries.

The evidence mentioned so far is consistent with the assumptions of our model, that paramilitaries were actively involved in influencing elections. The main prediction of our model is that paramilitaries should persist more where they deliver votes to the executive that they prefer, particularly in areas where this politician would otherwise not do well. This is because eliminating paramilitaries would implicitly cost valuable votes in the election. We also show that the correlations in the data are broadly consistent with this prediction.

14.3.1. Model. Let us now present a simple model to formalize the possible channels of interaction between central government and paramilitaries. Motivated by the Colombian experience, our focus will be on democratic politics, where an incumbent is facing reelection and decides whether to reconquer some of the areas under paramilitary control. The model will highlight how paramilitary preferences influence electoral outcomes because paramilitaries can coerce voters to support one candidate over another. It will then show how the effect of paramilitaries on electoral outcomes influences the willingness of the democratic central government to reconquer and remove the paramilitaries from different areas—the conditions of the formation of the modern Weberian state with a monopoly of violence over the entire country. Finally, we also investigate how the presence of paramilitaries affects the policy choices of the party in power. Our purpose is to communicate the main ideas in the simplest possible way. Our empirical work will then provide evidence showing how paramilitaries influence electoral outcomes and how this effect on elections interacts with the persistence of paramilitaries in certain areas.

We consider a two-period model of political competition between two parties. Party A is initially (at $t = 0$) in power and at $t = 1$, it competes in an election against party B. The country consists of a large equal-sized number, N, of regions, with each region inhabited by a large number of individuals. We denote the collection of these regions by $\mathcal{N}$. The party that wins the majority of the votes over all regions wins the election at $t = 1$. Regions differ in terms of their policy and ideological preferences and, in addition, some regions are under paramilitary control. We assume as in standard Downsian models that parties can make commitments to their policies, but their ideological stance is fixed (and may capture dimensions of policies to which they cannot make commitments).

We first introduce the details of electoral competition at date $t = 1$ and then return to the decisions at $t = 0$, in particular, to those concerning whether the government in power will expend the resources to reconquer some of the territories under paramilitary control. To start with, let us ignore the regions that are under paramilitary control (these will be introduced below).

The utility of individual i in region $j \in \mathcal{N}$ (i.e. $j = 1, ..., N$) when party $g \in \{A, B\}$ is in power is given by

$$U_{ij}\left(q, \tilde{\theta}^g\right) = u_j(q) - Y\left(\tilde{\theta}_j - \tilde{\theta}^g\right) + \tilde{\varepsilon}_{ij}^g,$$

where $q \in Q \subset \mathbb{R}^K$ is a vector of policies, u_j denotes the utility of all individuals in region j over this policy vector, $\tilde{\theta}_j$ is the ideological bliss point of the individuals in region $j \in \mathcal{N}$, so that $Y\left(\tilde{\theta}_j - \tilde{\theta}^g\right)$ is a penalty term for the ideological distance of the party in power and the individual (i.e., Y is a function that's increasing in $\left|\tilde{\theta}_j - \tilde{\theta}^g\right|$). This ideological distance captures policy choices not included in q (and to which the party cannot make a commitment at the election stage). We also assume that each u_j is strictly concave and differentiable. Finally, $\tilde{\varepsilon}_{ij}^g$ is an individual-specific utility term that will play the role of smoothing regional preferences over the two parties as in standard probabilistic voting models. We assume that

$$\tilde{\varepsilon}_{ij}^A - \tilde{\varepsilon}_{ij}^B = \xi + \varepsilon_{ij},$$

where ξ is a common "valance" term determining the relative popularity of one party versus another and ε_{ij} is an iid term. To simplify the discussion, we assume that ξ and each ε_{ij} have uniform distributions over $\left[-\frac{1}{2\phi}, \frac{1}{2\phi}\right]$. Therefore, conditional on the realization of ξ, the fraction of individuals in region $j \in \mathcal{N}$ who vote for party A will be

$$\frac{1}{2} + \phi\left[u_j\left(q^A\right) - u_j\left(q^B\right) + \theta_j + \xi\right],$$

where q^A and q^B are the policy vectors of the two parties, and

$$\theta_j \equiv Y\left(\tilde{\theta}_j - \tilde{\theta}^B\right) - Y\left(\tilde{\theta}_j - \tilde{\theta}^A\right)$$

is the ideological advantage of party A relative to party B in region $j \in \mathcal{N}$. Now using the fact that ξ is also uniformly distributed, the probability that party A gets elected as a function of its policies, the policies of the rival party, and its ideological advantage is

$$P^A\left(q^A, q^B \mid \boldsymbol{\theta}\right) = \frac{1}{2} + \frac{\phi}{N}\sum_{j=1}^{N}\left[u_j\left(q^A\right) - u_j\left(q^B\right) + \theta_j\right],$$

where $\boldsymbol{\theta}$ is the vector of ideological biases in favor of party A. In the election at time $t = 1$, Party A's problem is

$$\max_{q \in Q} P^A\left(q, q^B \mid \boldsymbol{\theta}\right) R^A, \tag{14.34}$$

where R^A is party A's rent from holding office. Conversely, the problem of party B is

$$\max_{q \in Q} \left[1 - P^A\left(q^A, q \mid \boldsymbol{\theta}\right)\right] R^B, \tag{14.35}$$

where R^B is party B's rent from holding office and we have used the fact that the probability of party B coming to power is the complement of that for party A. An *electoral equilibrium* at time $t = 1$ is a tuple $\left(q^A, q^B\right)$ that solves problems (14.34) and (14.35) simultaneously (given the ideological biases $\boldsymbol{\theta}$). Given the concavity and differentiability assumptions, an equilibrium is uniquely defined; moreover, as long as it is interior, it satisfies the following equations

$$\sum_{j=1}^{N} \nabla u_j\left(q^A\right) = 0 \text{ and } \sum_{j=1}^{N} \nabla u_j\left(q^B\right) = 0, \tag{14.36}$$

where ∇u_j denotes the gradient of function u_j with respect to the vector q. Clearly, (14.36) may not be satisfied if the solution is not in the feasible set of policies, Q, and in this case, an obvious complementary slackness generalization of (14.36) holds. Strict concavity of each u_j immediately implies that $q^A = q^B = q^*$. It is also straightforward to see that strict concavity implies $q^A = q^B = q^*$ for some q^*, even if the equilibrium is not interior. Therefore, party A will win the election at time $t = 1$ with probability

$$P^A\left(q^*, q^* \mid \boldsymbol{\theta}\right) = \frac{1}{2} + \frac{\phi}{N}\sum_{j=1}^{N} \theta_j = \frac{1}{2} + \phi \mathbb{E}\theta_j, \tag{14.37}$$

where $\mathbb{E}\theta_j$ denotes the expectation or the mean of θ_j across all regions. This then leads to the following proposition, characterizing the equilibrium (proof in the text).

Proposition 14.6. *Without paramilitaries, there exists a unique electoral equilibrium (at* $t = 1$*) where* $q^A = q^B = q^*$*, and* q^**, if interior, satisfies (14.36). Party A wins the election with probability given by (14.37).*

Two important points to note are as follows. First, without paramilitary presence, national policies are chosen to cater to the preferences of all voters in all regions. This feature is fairly general, though as is well known the fact that both parties choose the same policy vector (policy convergence) is special and relies on the fact that the two parties do not themselves have preferences over policies. Whether they do or not is not important for the results here, and we therefore opted for the simpler specification. Second, average ideological bias across all regions determines the probability of reelection for party A (which is currently in power). We will next see how this result changes under various different assumptions about paramilitary behavior.

Next, let us suppose that a subset of the regions, denoted by $\mathcal{Z} \subset \mathcal{N}$ are under paramilitary control. Denote the total number of these regions by Z and their fraction (their ratio to the number of total regions) by z. The key feature of paramilitary-controlled areas for our purposes is that, as we will document in detail below, voting is not free but influenced by the implicit or explicit pressure of the paramilitaries. Throughout the rest of this section we impose this feature.

We start with paramilitaries with "exogenous preferences," meaning that how the paramilitaries influence the voting behavior of the citizens in the regions they control is exogenous. This will contrast with the case in which the support of the paramilitaries is endogenous to the policy choices, studied next. In particular, we take the behavior of the paramilitaries (and the voting behavior of the citizens in paramilitary-controlled areas) as given. In particular, suppose that in each paramilitary-controlled region $j \in \mathcal{Z}$, a fraction $\tilde{m}_j$ of the voters will vote for party A regardless of policies (so the voting behavior of these individuals in these paramilitary-controlled regions is *insensitive* to policies). Let us denote the complement of the set $\mathcal{Z}$ by $\mathcal{J} = \mathcal{N} \backslash \mathcal{Z}$ and the total number of regions in this (non-paramilitary-controlled) set by $J = N - Z$. Let us also define $m_j \equiv \tilde{m}_j - 1/2$. Then with an identical reasoning to that above, the probability that party A will win the election at time $t = 1$ is

$$P^A\left(q^A, q^B \mid \boldsymbol{\theta}, \mathbf{m}\right) = \frac{1}{2} + \frac{\phi\left(1-z\right)}{J} \sum_{j \in \mathcal{J}} \left[u_j\left(q^A\right) - u_j\left(q^B\right) + \theta_j\right] + \frac{z}{Z} \sum_{j \in \mathcal{Z}} m_j,$$

where $\mathbf{m}$ denotes the vector of m_j's (together with information on which j's are in the set $\mathcal{Z}$).

We again assume that both parties maximize the probability of coming to power and define an electoral equilibrium in the same way. With an identical argument to that before, we obtain the following proposition.

Proposition 14.7. *Under paramilitaries with endogenous preferences, there exists a unique electoral equilibrium (at $t = 1$) where $q^A = q^B = q^*$. If q^* is interior, it satisfies $\sum_{j \in \mathcal{J}} \nabla u_j\left(q^*\right) = 0$. Party A wins the election with probability*

$$P^A\left(q^*, q^* \mid \boldsymbol{\theta}, \mathbf{m}\right) = \frac{1}{2} + \phi\left(1-z\right) \mathbb{E}\left[\theta_j \mid j \in \mathcal{J}\right] + z\mathbb{E}\left[m_j \mid j \in \mathcal{Z}\right].$$

Two features that are noteworthy relative to Proposition 14.6 are as follows. First, policies no longer cater to the preferences of all regions. Since citizens in paramilitary-controlled areas cannot reward or punish a government according to the policy proposals that it makes, both parties only target their policies to the voters in the non-paramilitary-controlled areas. This implies that, endogenously, public goods and other amenities will be reduced in the paramilitary-controlled areas beyond the direct effect of paramilitary presence. Thus, all else equal, we may expect paramilitary presence to increase inequality across regions. Second, electoral outcomes will now be dependent on the influence of the paramilitaries on voting behavior, which is captured by the last term in $P^A(q^*, q^* \mid \boldsymbol{\theta}, \mathbf{m})$. If paramilitaries prefer party A, meaning that $\mathbb{E}[m_j \mid j \in \mathcal{Z}] > 0$, then the probability that party A will win the election (and stay in power) is greater, other things equal. The more areas are controlled by the paramilitaries, the stronger is this effect. In the empirical work below, we will provide indirect evidence consistent with Proposition 14.7 by showing the influence of paramilitaries on electoral outcomes.

This last feature already highlights how paramilitaries can have a major influence on democratic politics. Nevertheless, this effect was minimized by the model in this subsection by assuming exogenous preferences for the paramilitaries. We will relax this assumption below. But first we discuss how the electoral role of paramilitaries affects the decision of the central government to extend ("broadcast") its power to peripheral areas controlled by the paramilitaries.

Taking the electoral equilibrium at time $t = 1$ as given, let us now consider the decisions of the government (party A) at time $t = 0$. In particular, as discussed in the Introduction, a key dimension of the process of the formation of the state is the ability and willingness of the central government to establish its monopoly of violence and thus remove the power of other groups with access to guns and means of exercising (local) violence. Let us model this in the simplest possible way and suppose that at time $t = 0$, the objective of the governing party is

$$\sum_{j \in \mathcal{R}} \gamma_j + P^A\left(q^A, q^B \mid \boldsymbol{\theta}\right) R^A, \tag{14.38}$$

where $\mathcal{R} \subset \mathcal{Z}$ is a subset of the areas previously controlled by the paramilitary that are "reconquered" by the central government, and γ_j is the net benefit of reconquering area $j \in \mathcal{R}$, which accrues to the government at time $t = 0$. This net benefit includes the additional tax revenues or security gains that the central government will drive and subtracts the potential "real" cost of the reconquest (spending on the military, potentially stability and loss of life). However, the objective of the governing party, party A, also includes the probability that it will remain in power, thus enjoying rents from power at time $t = 1$. In particular, if some area $j \in \mathcal{Z}$ is reconquered, then in the subsequent electoral equilibrium at time $t = 1$, party A will obtain a fraction $1/2 + \phi\theta_j$ of the votes from this region as opposed to receiving $\tilde{m}_j = m_j + 1/2$ of the votes had this place remained under paramilitary control.

A *subgame perfect equilibrium* of this game is defined as an electoral equilibrium at date $t = 1$ together with decisions by party A at date $t = 0$ that maximizes its utility taking the date $t = 1$ equilibrium as given.

This analysis in the preceding paragraph then establishes the following proposition.

Proposition 14.8. *A subgame perfect equilibrium involves the electoral equilibrium characterized in Proposition 14.7 at time $t = 1$, and at time $t = 0$, Party A reconquers*

$$\text{all } j \in \mathcal{Z} \text{ such that } \gamma_j + (\phi\theta_j - m_j) R^A > 0$$

and does not reconquer

$$\text{any } j \in \mathcal{Z} \text{ such that } \gamma_j + (\phi\theta_j - m_j) R^A < 0.$$

This proposition is an important result of our analysis and will be investigated in our empirical work. It implies that the willingness of the state to reconquer areas controlled by the paramilitaries, and thus establish the monopoly of violence envisaged as an essential characteristic of the modern state by Max Weber, is affected not only by the real costs and benefits of doing so, but also by the implications of this expansion of the authority of the state on electoral outcomes. In particular, if many of these paramilitary-controlled areas have $m_j > \phi\theta_j$, then the state, currently controlled by party A, will be reluctant to reconquer these areas, because doing so will make it more difficult for this party to succeed in the upcoming elections (and moreover, this effect will be stronger when rents from power at $t = 1$, R^A, are higher). Naturally, the areas that are most valuable in the hands of the paramilitaries are those that have both low θ_j and high m_j; that is, *areas that would have otherwise voted for party B, but paramilitaries are forcing citizens to vote in favor of party A*. A government that does not require electoral support (e.g., a "purely non-democratic" government) would have decided to reconquer all areas with $\gamma_j > 0$. Therefore, to the extent that $\phi\mathbb{E}[\theta_j \mid j \in \mathcal{J}] < \mathbb{E}[m_j \mid j \in \mathcal{Z}]$, i.e., to the extent that paramilitaries are ideologically closer to the government in power than the opposition party, a democratic government may be less willing to broadcast its power and reconquer areas under paramilitary control than such a non-democratic government (or a government that is secure in its position).

Note an important implication of the functional form assumptions we have imposed so far, in particular the uniform distributions of idiosyncratic preference and valance terms: the value of additional votes to the party in power is constant and independent of its "expected winning probability". As a consequence, Proposition 14.8 takes a simple form, where the value of paramilitary votes to the party in power is independent of this probability. With other functional forms, as in reality, this value, and thus the behavior of this party towards the paramilitary groups, may depend on its expected winning probability, for example, making it less responsive to the votes delivered by these paramilitary forces when it is ex ante more likely to win the election.

The discussion so far was for paramilitaries with endogenous preferences and thus took the vector $\mathbf{m}$ as given. Naturally, the willingness of the paramilitaries to coerce citizens to vote for one candidate or another is also endogenous and depends on their policy and ideological preferences. We now investigate these issues. Suppose that, as with the citizens, the preferences of the paramilitaries controlling region $j \in \mathcal{Z}$ are given by

$$W_j\left(q, \theta^g\right) = w_j\left(q\right) - \hat{Y}\left(\tilde{\theta}_j - \tilde{\theta}^g\right) + \tilde{\varepsilon}_j^g,$$

where $\hat{Y}$ is another function (possibly the same as Y) that is also increasing in $\left|\tilde{\theta}_j - \tilde{\theta}^g\right|$ and now $\tilde{\theta}_j$ is the policy preference of the group of paramilitaries controlling region j. With a similar reasoning to that above, let us define

$$\hat{\theta}_j \equiv \hat{Y}\left(\tilde{\theta}_j - \tilde{\theta}^B\right) - \hat{Y}\left(\tilde{\theta}_j - \tilde{\theta}^A\right)$$

as the ideological leanings of the paramilitaries in region j in favor of party A (we use $\hat{\theta}_j$ instead of θ_j to highlight that this refers to the paramilitaries). And in addition, suppose that $\tilde{\varepsilon}_j^A - \tilde{\varepsilon}_j^B$ has a uniform distribution over $\left[-\frac{1}{2\hat{\phi}}, \frac{1}{2\hat{\phi}}\right]$. Then the probability that paramilitaries in region $j \in \mathcal{Z}$ will prefer party A to party B is given by

$$\frac{1}{2} + \hat{\phi}\left[w_j\left(q^A\right) - w_j\left(q^B\right) + \hat{\theta}_j\right].$$

Let us also assume that paramilitaries can force all voters in their sphere of influence to vote for whichever party they prefer. Then the probability that party A will win the election becomes

$$\begin{aligned} P^A\left(q^A, q^B \mid \hat{\boldsymbol{\theta}}\right) &= \frac{1}{2} + \frac{\phi\left(1-z\right)}{J}\sum_{j\in\mathcal{J}}\left[u_j\left(q^A\right) - u_j\left(q^B\right) + \theta_j\right] \\ &\quad + \frac{\hat{\phi}z}{Z}\sum_{j\in\mathcal{Z}}\left[w_j\left(q^A\right) - w_j\left(q^B\right) + \hat{\theta}_j\right], \end{aligned}$$

where now $\hat{\boldsymbol{\theta}}$ denotes the vector of all ideological preferences, including those of the paramilitaries. Naturally, the model with paramilitaries with exogenous preferences studied previously is a special case of this model where $w_j\left(q\right) \equiv 0$ for all $q \in Q$, so that paramilitaries do not care about policy (though they may still care about the ideological stance of the party in power).

With a similar reasoning to our analysis above, electoral competition will lead to the same policy choice for both parties, and when it is interior, this vector will be given by the solution to the following set of equations:

$$\phi\left(1-z\right)\nabla u_j\left(\hat{q}^*\right)+\hat{\phi}z\nabla w_j\left(\hat{q}^*\right)=0. \tag{14.39}$$

Naturally, these equations hold in the complementary-slackness form when $\hat{q}^*$ may be at the boundary of the feasible policy set Q.

Therefore, we obtain the following characterization of electoral equilibrium and efforts by the state to reconquer paramilitary-controlled areas under the control of the paramilitaries (proof in the text).

Proposition 14.9. *Under paramilitaries with endogenous preferences, there exists a unique electoral equilibrium at $t=1$ where $q^A=q^B=q^*$. If q^* is interior, it satisfies (14.39). Party A wins the election with probability*

$$P^A\left(q^*,q^*\mid\boldsymbol{\theta},\mathbf{m}\right)=\frac{1}{2}+\phi\left(1-z\right)\mathbb{E}\left[\theta_j\mid j\in\mathcal{J}\right]+z\hat{\phi}\mathbb{E}\left[\hat{\theta}_j\mid j\in\mathcal{Z}\right].$$

Moreover, the subgame perfect equilibrium involves Party A reconquering (at time $t=0$)

$$\text{all } j\in\mathcal{Z}\text{ such that }\gamma_j+\left(\phi\theta_j-\hat{\phi}\hat{\theta}_j\right)R^A>0,$$

and not reconquering

$$\text{any } j\in\mathcal{Z}\text{ such that }\gamma_j+\left(\phi\theta_j-\hat{\phi}\hat{\theta}_j\right)R^A<0.$$

There are several new features in this proposition. First, when paramilitaries adjust their support depending on the policies and ideological stance of the two parties, the parties then change their policies in order to be more attractive to the paramilitaries' policy preferences. That is, rather than catering to the preferences of the citizens in the areas that are controlled by the paramilitaries (which they would have done without the paramilitaries), parties appease the paramilitaries themselves. This result is the basis of the potential *symbiotic* relationship between paramilitaries and the executive mentioned in the Introduction. Moreover, it can further increase the inequality among the regions, with the policies chosen specifically to support, or refrain from threatening, the paramilitaries and the areas where the paramilitaries are strongest. Two features determine how slanted towards the paramilitaries equilibrium policies are. These are: the size of the paramilitary-controlled areas (the greater is z, the more influential are the paramilitaries in shaping equilibrium policy) and the relative responsiveness of the paramilitaries to policy concessions (the greater is $\hat{\phi}$ relative to ϕ, the more responsive are policies to paramilitary preferences relative to citizen preferences). In addition, because electoral competition makes both parties cater to the wishes of the paramilitaries, at the end the paramilitaries ideological preferences play a central role in whether they force the population to vote for party A or party B.

Finally, we can also allow both parties or one of the parties to modify its ideological stance (in a credible fashion). The same analysis as here will then imply that in order to attract votes from paramilitary-controlled areas, one or both parties may decide to pander to the ideological preferences of the paramilitaries.

An important question in the context of Colombian politics is why right-wing paramilitary groups have become more involved in influencing elections than left-wing guerrillas, in particular, more so than the relatively well-organized FARC. One possible answer is that in contrast to the guerrillas, the paramilitaries do not have national ambitions, making a coalition between them and the executive controlling the central state more feasible. The model presented so far has implicitly made this assumption, since we did not introduce the risk that the non-state armed actors may take over the central state.

A simple way of introducing this possibility would be to have a probability $\Phi(z)$ that the non-state armed actors would become strong enough to challenge the central state, perhaps overthrow it. Such an overthrow of the central government by non-state armed actors is not uncommon in weak African states, such as Somalia, Sierra Leone or Liberia, and has certainly been the objective of the FARC. Naturally, we would expect $\Phi(z)$ to be increasing in z, so that when these groups control more areas, they are more likely to pose such a national challenge. In that case, we would need to change the objective function of party A to incorporate this possibility. For example, equation (14.38) could be modified to

$$\sum_{j\in\mathcal{R}}\gamma_j+[1-\Phi(z)]\,P^A\left(q^A,q^B\mid\boldsymbol{\theta}\right)R^A.$$

This specification makes it clear that when $\Phi(z)>0$, there will be stronger incentives for party A to reconquer territories controlled by these non-state armed groups (thus reducing z). When $\Phi(z)$ is sufficiently high and sufficiently decreasing in z, this effect can more than compensate for the electoral advantage that local control by these groups creates for the party in power. Thus factoring in the national ambitions of non-state armed actors reduces the room for a coalition or a symbiotic relationship between these groups and the executive. Expressed differently, this reasoning suggests that when non-state armed actors have national ambitions, it will be advantageous for the central state to eliminate them (sooner or later), thus any implicit or explicit policy promises that it makes to such groups would be non-credible, making a coalition between them impossible. This perspective suggests a natural reason for why, in Colombia, such a coalition may have been much more likely to arise between the executive and the paramilitaries rather than with the FARC.

14.3.2. Empirical Predictions and Evidence from Columbia. What makes the above simple model interesting is that the empirical predictions are borne out with recent data from Colombia. The following broad patterns hold in recent Colombian data:

(1) Consistent with Proposition 14.7, paramilitaries, once they became sufficiently powerful, started influencing electoral outcomes in the areas of Colombia they controlled.
(2) Consistent with Proposition 14.8, we will show that paramilitaries located in areas that voted for the current President in great numbers, but in past elections tended to vote for more liberal politicians, are more likely to persist.
(3) Consistent with Proposition 14.9, we will show that the President has proposed legislation in line with the preferences of the paramilitaries, and the Senators elected from high paramilitary areas have supported this legislation.

14.4. References

(1) Acemoglu, Daron (2005) "Politics and Economics in Weak and Strong States" *Journal of Monetary Economics*, 52, 1199-1226.
(2) Acemoglu, Daron, Mike Golosov and Aleh Tsyvinski (2006) "Markets Versus Governments: Political Economy of Mechanisms" unpublished.
(3) Acemoglu, Daron, Simon Johnson and James A. Robinson (2003) "An African Success Story: Botswana," in Dani Rodrik ed. In Search of Prosperity: Analytic Narratives on Economic Growth, Princeton; Princeton University Press.
(4) Acemoglu, Daron, James A. Robinson and Rafael Santos (2009) "Monopoly of Violence: The Case of Colombia," mimeo.
(5) Centeno, Miguel (2002) Blood and debt: war and the nation-state in Latin America, University Park, Pa.; Pennsylvania State University Press.
(6) Evans, Peter (1989) "Predatory, Developmental and Other Apparatuses: A Comparative Political Economy Perspective on the Third World State," *Sociological Forum*. 4(4):561-587 (December, 1989).
(7) Evans, Peter (1995) Embedded Autonomy, Princeton; Princeton University Press.
(8) Hayek, Freidrich (1945) "The Use of Knowledge in Society," *American Economic Review*, 35, 519-530.
(9) Herbst, Jeffery I. (2000) *States and Power in Africa: Comparative Lessons in Authority and Control,* Princeton University Press, Princeton NJ.

(10) Johnson, Chalmers A. (1982) MITI and the Japanese miracle : the growth of industrial policy, 1925-1975, Stanford; Stanford University Press.

(11) Kohli, Atul (2004) State-directed development: political power and industrialization in the global periphery, New York; Cambridge University Press.

(12) Tilly, Charles (1990) *Coercion, Capital and European States, AD 990-1990*, Blackwell, Cambridge MA.

CHAPTER 15

Oligarchy Versus Democracy

We now present a dynamic model of institutions focusing on different economic incentives under two different sets of political institutions, loosely called oligarchy/nondemocracy and democracy. In many ways, this is a natural extension of the model of "politics under elite control" we studied earlier. But it will involve richer interactions, and as such, it will allow us to start the discussion of regime changes and institutional origins, which will be our main focus in the following lectures.

In the models so far, political power was either in the hands of a specific set of politicians or a government, or alternatively, political institutions were already democratic. In the first instance we examined the incentives of politicians to choose inefficient economic policies but where their power was constrained only by a very reduced form model of replacement. In the second instance, we again looked at some simple ideas about the incentives of democratic politicians to choose different public policies and we analyzed how these incentives depeded on the detailed institutional structure (presidentialism versus parliamentarianism). But we did not pay much attention to whether democracy or oligarchy were relatively better or worse nor did we present a framework for thinking about why some societies but not others might be democratic. Now we delve into these questions and also discuss new mechanisms which may influence the efficiency of resource allocation. We do so in a model where there is a strong dichotomy betweem "elite control" and "popular democracy", which will be respectively referred to as oligarchy/nondemocracy and democracy. In line with our previous discussions, will see that both oligarchy and democracy will cause economic distortions. In the case of democracy this will be a rather familiar distortion stemming from costly redistributive taxation. In the case of oligarchy, this force will be absent, but policies that favor the economically powerful may do so at the expense of the rest of society, and may be equally or more inefficient. Now we present a model capturing this trade-off.

15.1. Basic Model

Let us consider an infinite horizon economy populated by a continuum 1 of risk neutral agents, with discount factor equal to $\beta < 1$. There is a unique non-storable final good denoted by y. The expected utility of agent j at time 0 is given by:

$$U_0^j = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t c_t^j, \tag{15.1}$$

where $c_t^j \in \mathbb{R}$ denotes the consumption of agent j at time t and $\mathbb{E}_t$ is the expectations operator conditional on information available at time t.

We assume that each individual dies with a small probability ε in every period, and a mass ε of new individuals are born (with the convention that after death there is zero utility and β is the discount factor inclusive of the probability of death). We will consider the limit of this economy with $\varepsilon \to 0$. The reason for introducing the possibility of death is to avoid the case where the supply of labor is exactly equal to the demand for labor for a range of wage rates, which can otherwise arise in the oligarchic equilibrium. In other words, in the economy with $\varepsilon = 0$, there may also exist other equilibria, and in this case, the limit $\varepsilon \to 0$ picks a specific one from the set of equilibria.

The key distinction in this economy is between production workers on the one hand and capitalists/entrepreneurs on the other. Each agent can either be employed as a worker or set up a firm to become an entrepreneur. While all agents have the same productivity as workers, their productivity in entrepreneurship differs. In particular, agent j at time t has entrepreneurial talent/skills $a_t^j \in \{A^L, A^H\}$ with $A^L < A^H$. To become an entrepreneur, an agent needs to set up a firm, if he does not have an active firm already. Setting up a new firm may be costly because of entry barriers created by existing entrepreneurs.

Each agent therefore starts period t with skill level $a_t^j \in \{A^H, A^L\}$ and $s_t^j \in \{0, 1\}$ which denotes whether the individual has an active firm. We refer to an agent with $s_t^j = 1$ as a member of the "elite", since he will have an advantage in becoming an entrepreneur (when there are entry barriers), and in an oligarchic society, he may be politically more influential than non-elite agents.

Within each period, each agent makes the following decisions: an occupation choice $e_t^j \in \{0, 1\}$, and in addition if $e_t^j = 1$, i.e., if he becomes an entrepreneur, he also makes investment, employment, and hiding decisions, k_t^j, l_t^j and h_t^j, where h_t^j denotes whether he decides to hide his output in order to avoid taxation (since the final good is not storable, the consumption decision is simply given by the budget constraint).

Agents also make the policy choices in this society. How the preferences of various agents map into policies differs depending on the political regime, which is discussed in detail below. For now we note that there are three policy choices: a tax rate $\tau_t \in [0, 1]$ on output, lump-sum transfers to all agents denoted by $T_t \in [0, \infty)$, and a cost $B_t \in [0, \infty)$ to set up a new firm. I assume that the entry barrier B_t is pure waste, for example corresponding to the bureaucratic procedures that individuals have to go through to open a new business. As a result, lump-sum transfers are financed only from taxes.

An entrepreneur with skill level a_t^j can produce

$$y_t^j = \frac{1}{1-\alpha}(a_t^j)^\alpha (k_t^j)^{1-\alpha}(l_t^j)^\alpha \tag{15.2}$$

units of the final good, where l_t^j is the amount of labor hired by the entrepreneur and $k_t^j \geq 0$ is the capital stock of the entrepreneur. We assume that there is full depreciation of the capital at the end of the period, so k_t^j is also the level of investment of entrepreneur j at time t. To simplify the analysis we also assume that all firms have to operate at the same size, λ, so $l_t^j = \lambda$. Finally, suppose that the entrepreneur himself can work in his firm as one of the workers, which implies that the opportunity cost of becoming an entrepreneur is 0.

To simplify the expressions below, we define $b_t \equiv B_t/\lambda$. Profits are then $\pi_t^j = (1-\tau_t)\, y_t^j - w_t l_t^j - k_t^j$, as the return to entrepreneur j gross of the cost of entry barriers. Intuitively, the entrepreneur produces y_t^j, pays a fraction τ_t of this in taxes, pays a total wage bill of $w_t l_t^j$, and incurs an investment cost of k_t^j. Given a tax rate τ_t and a wage rate $w_t \geq 0$ and using the fact that $l_t^j = \lambda$, the net profits of an entrepreneur with talent a_t^j at time t are:

$$\pi\left(k_t^j \mid a_t^j, w_t, \tau_t\right) = \frac{1-\tau_t}{1-\alpha}(a_t^j)^\alpha (k_t^j)^{1-\alpha}\lambda^\alpha - w_t\lambda - k_t^j, \tag{15.3}$$

as long as the entrepreneur chooses $h_t^j = 0$. If he instead hides his output, i.e., $h_t^j = 1$, he avoids the tax, but loses a fraction $\delta < 1$ of his revenues, so his profits are:

$$\tilde{\pi}\left(k_t^j \mid a_t^j, w_t, \tau_t\right) = \frac{1-\delta}{1-\alpha}(a_t^j)^\alpha (k_t^j)^{1-\alpha}\lambda^\alpha - w_t\lambda - k_t^j.$$

The comparison of these two expressions immediately implies that if $\tau_t > \delta$, all entrepreneurs will hide their output, and there will be no tax revenue. Therefore, the relevant range of taxes will be

$$0 \leq \tau_t \leq \delta.$$

The (instantaneous) gain from entrepreneurship for an agent of talent $z \in \{L, H\}$ as a function of the tax rate τ_t, and the wage rate, w_t, is:

$$\Pi^z(\tau_t, w_t) = \max_{k_t^j} \pi\left(k_t^j \mid a_t^j = A^z, w_t, \tau_t\right). \tag{15.4}$$

Note that this is the *net gain* to entrepreneurship since the agent receives the wage rate w_t irrespective (either working for another entrepreneur when he is a worker, or working for himself—thus having to hire one less worker—when he is an entrepreneur). More importantly, the gain to becoming an entrepreneur for an agent with $s_t^j = 0$ and ability $a_t^j = A^z$ is $\Pi^z(\tau_t, w_t) - B_t = \Pi^z(\tau_t, w_t) - \lambda b_t$, since this agent will have to pay the additional cost imposed by the entry barriers. With this notation we can also define the budget constraint of workers as $c_t^j \leq w_t + T_t$ and that for an entrepreneur of ability A^z as $c_t^j \leq w_t + T_t + \Pi^z(\tau_t, w_t)$, where T_t is the level of lump-sum transfer.

Labor market clearing requires the total demand for labor not to exceed the supply. Since entrepreneurs also work as production workers, the supply is equal to 1, so:

$$\int_0^1 e_t^j l_t^j dj = \int_{j \in \mathbf{S}_t^E} \lambda dj \leq 1, \tag{15.5}$$

where $\mathbf{S}_t^E$ is the set of entrepreneurs at time t.

It is also useful at this point to specify the law of motion of the vector $\left(s_t^j, a_t^j\right)$ which determines the "type" of agent j at time t. The transition rule for s_t^j is straightforward: if agent j at time t sets up a firm, then his offspring inherits a firm at time $t+1$, so

$$s_{t+1}^j = e_t^j, \tag{15.6}$$

with $s_0^j = 0$ for all j, and also $s_t^j = 0$ if an individual j is born at time t. The important assumption here is that if an individual does not operate his firm, he loses "the license", so next time he wants to set up a firm, he needs to incur the entry cost (and the assumption that $l_t^j = \lambda$ rules out the possibility of operating the firm at a much smaller scale).

Finally, we assume that there is imperfect correlation between the entrepreneurial skill over time with the following Markov structure:

$$a_{t+1}^j = \begin{cases} A^H & \text{with probability} \quad \sigma^H & \text{if } a_t^j = A^H \\ A^H & \text{with probability} \quad \sigma^L & \text{if } a_t^j = A^L \\ A^L & \text{with probability } 1 - \sigma^H & \text{if } a_t^j = A^H \\ A^L & \text{with probability } 1 - \sigma^L & \text{if } a_t^j = A^L \end{cases}, \tag{15.7}$$

where σ^H, $\sigma^L \in (0,1)$. Here σ^H is the probability that an agent has high skill in entrepreneurship conditional on being high skill in the previous period, and σ^L is the probability when he was low skill. It is natural to suppose that $\sigma^H \geq \sigma^L > 0$, so that skills are persistent and low skill is not an absorbing state. What is important for the results is imperfect correlation of entrepreneurial talent over time, i.e., $\sigma^H < 1$, so that the identities of the entrepreneurs necessary to achieve productive efficiency change over time.

It can be verified easily that

$$M \equiv \frac{\sigma^L}{1 - \sigma^H + \sigma^L} \in (0,1)$$

is the fraction of agents with high skill in the stationary distribution (i.e., $M\left(1 - \sigma^H\right) = (1 - M)\,\sigma^L$). Since there is a large number (continuum) of agents, the fraction of agents with high skill at any point is M. Throughout we assume that

$$M\lambda > 1,$$

so that, without entry barriers, high-skill entrepreneurs generate more than sufficient demand to employ the entire labor supply. Moreover, think of M as small and λ as large; in particular, we assume $\lambda > 2$, which ensures that the workers are always in the majority and simplifies the political economy discussion below.

Finally, the timing of events within every period is:

- Entrepreneurial talents/skills, $\left[a_t^j\right]$, are realized.
- The entry barrier for new entrepreneurs b_t is set.
- Agents make occupational choices, $\left[e_t^j\right]$, and entrepreneurs make investment decisions, $\left[k_t^j\right]$.
- The labor market clearing wage rate, w_t, is determined.
- The tax rate on entrepreneurs, τ_t, is set.
- Entrepreneurs make hiding decisions, $\left[h_t^j\right]$.

Note that we used the notation $\left[a_t^j\right]$ to describe the whole set $\left[a_t^j\right]_{j\in[0,1]}$, or more formally, the mapping $\mathbf{a}_t : [0,1] \to \left\{A^L, A^H\right\}$, which assigns a productivity level to each individual j, and similarly for $\left[e_t^j\right]$, etc.

Entry barriers and taxes will be set by different agents in different political regimes as will be specified below. Notice that taxes are set after the investment decisions, which can be motivated by potential commitment problems whereby entrepreneurs can be "held up" after they make their investments decision. Once these investments are sunk, it is in the interest of the workers to tax and redistribute entrepreneurial income.

Throughout the analysis we focus on the Markov Perfect Equilibrium (MPE), where strategies are only a function of the payoff relevant states. For individual j the payoff relevant state at time t includes his own state $\left(s_t^j, a_t^j\right)$, and potentially the fraction of entrepreneurs that are high skill, denoted by μ_t, and defined as

$$\mu_t = \Pr\left(a_t^j = A^H \mid e_t^j = 1\right) = \Pr\left(a_t^j = A^H \mid j \in \mathbf{S}_t^E\right).$$

The MPE can be characterized by considering the appropriate Bellman equations, and characterizing the optimal strategies within each time period by backward induction. I start with the "economic equilibrium," which is the equilibrium of the economy described above given a policy sequence $\{b_t, \tau_t\}_{t=0,1,...}$. Let $x_t^j = \left(e_t^j, k_t^j, , h_t^j\right)$ be the vector of choices of agent j at time t, $x_t = \left[x_t^j\right]_{j\in[0,1]}$ denote the choices for all agents, and $p_t = (b_t, \tau_t)$ denote the vector of policies at time t. Moreover, let $\mathbf{p}^t = \{p_n\}_{n=t}^{\infty}$ denote the infinite sequence of policies from time t onwards, and similarly $\mathbf{w}^t$ and $\mathbf{x}^t$ denote the sequences of wages and choices from t onwards. Then $\hat{\mathbf{x}}^t$ and a sequence of wage rates $\hat{\mathbf{w}}^t$ constitute an economic equilibrium given a policy sequence $\mathbf{p}^t$ if, given $\hat{\mathbf{w}}^t$ and $\mathbf{p}^t$ and his state $\left(s_t^j, a_t^j\right)$, $\hat{x}_t^j$ maximizes the utility of agent j, (15.1), and $\hat{w}_t$ clears the labor market at time t, i.e., equation (15.5) holds. Each agent's type in the next period, $\left(s_{t+1}^j, a_{t+1}^j\right)$, then follows from equations (15.6) and (15.7) given $\mathbf{x}^t$.

We now characterize this equilibrium. Recall that $s_0^j = 0$ for all j, and suppose $b_0 = 0$, so that in the initial period there are no entry barriers (since $s_0^j = 0$ for all j, any positive entry barrier would create waste, but would not affect who enters entrepreneurship).

Since $l_t^j = \lambda$ for all $j \in \mathbf{S}_t^E$ (where, recall that, $\mathbf{S}_t^E$ is the set of entrepreneurs at time t), profit-maximizing investments are given by:

$$k_t^j = (1-\tau_t)^{1/\alpha} a_t^j \lambda, \tag{15.8}$$

so that the level of investment is increasing in the skill level of the entrepreneur, a_t^j, and the level of employment, λ, and decreasing in the tax rate, τ_t. (Alternatively, (15.8) can be written as $k_t^j = (1-\hat{\tau}_t)^{1/\alpha} a_t^j \lambda$ where $\hat{\tau}_t$ is the tax rate expected at the time of investment; in equilibrium, $\hat{\tau}_t = \tau_t$).

Now using (15.8), the net current gain to entrepreneurship, as a function of entry barriers, taxes, equilibrium wages, for an agent of type $z \in \{L, H\}$ (i.e., of skill level A^L or A^H) can be obtained as:

$$\Pi^z(\tau_t, w_t) = \frac{\alpha}{1-\alpha}(1-\tau_t)^{1/\alpha} A^z \lambda - w_t \lambda. \tag{15.9}$$

Moreover, the labor market clearing condition (15.5) implies that the total mass of entrepreneurs at any time is $\mathbf{e}_t \equiv \int_{j \in \mathbf{S}_t^E} dj = 1/\lambda$. Tax revenues at time t and the per capita lump-sum transfers are given as:

$$T_t = \sum_{j \in \mathbf{S}_t^E} \tau_t y_t^j = \frac{1}{1-\alpha} \tau_t (1-\tau_t)^{\frac{1-\alpha}{\alpha}} \lambda \sum_{j \in \mathbf{S}_t^E} a_t^j. \tag{15.10}$$

Let us now denote the value of an entrepreneur with skill level $z \in \{L, H\}$ as a function of the sequence of future policies and equilibrium wages, $\left(\mathbf{p}^t, \mathbf{w}^t\right)$, by $V^z\left(\mathbf{p}^t, \mathbf{w}^t\right)$, and the value of a worker of type z in the same situation by $W^z\left(\mathbf{p}^t, \mathbf{w}^t\right)$. We have

$$W^z\left(\mathbf{p}^t, \mathbf{w}^t\right) = w_t + T_t + \beta C W^z\left(\mathbf{p}^{t+1}, \mathbf{w}^{t+1}\right), \tag{15.11}$$

where $CW^z\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right)$ is the continuation value for a worker of type z from time $t+1$ onwards, given by

$$\begin{aligned}(15.12)\qquad CW^z\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right) &= \sigma^z \max\left\{W^H\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right), V^H\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right)-\lambda b_{t+1}\right\}\\ &+\left(1-\sigma^z\right)\max\left\{W^L\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right), V^L\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right)-\lambda b_{t+1}\right\}.\end{aligned}$$

The expressions for both (15.11) and (15.12) are intuitive. A worker of type z receives a wage income of w_t (independent of his skill), a transfer of T_t, and the continuation value $CW^z\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right)$. To understand this continuation value, note that the worker stays high skill with probability σ^z, and in this case, he can either choose to remain a worker, receiving value $W^H\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right)$, or decide to become an entrepreneur by incurring the entry cost λb_{t+1}, receiving the value of a high-skill entrepreneur, $V^H\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right)$. The max operator makes sure that he chooses whichever option gives higher value. With probability $1-\sigma^z$, he transitions from high skill to low skill, and receives the corresponding values.

Similarly, the value functions for entrepreneurs are given by:

$$(15.13)\qquad V^z\left(\mathbf{p}^t,\mathbf{w}^t\right)=w_t+T_t+\Pi^z\left(\tau_t,w_t\right)+\beta CV^z\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right),$$

where Π^z is given by (15.9) and now crucially depends on the skill level of the agent, and $CV^z\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right)$ is the continuation value for an entrepreneur of type z:

$$\begin{aligned}(15.14)\quad CV^z\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right) &= \sigma^z \max\left\{W^H\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right), V^H\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right)\right\}\\ &+\left(1-\sigma^z\right)\max\left\{W^L\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right), V^L\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right)\right\},\end{aligned}$$

An entrepreneur of ability A^z also receives the wage w_t (working for his own firm) and the transfer T_t, and in addition makes profits equal to $\Pi^z\left(\tau_t,w_t\right)$. The following period, this entrepreneur has high skill with probability σ^z and low skill with probability $1-\sigma^z$, and conditional on the realization of this event, he decides whether to remain an entrepreneur or become a worker. Two points are noteworthy here. First, in contrast to the expressions in (15.11), there is no additional cost of becoming an entrepreneur, λb_{t+1}, since this individual already has a firm. Second, if he decides to become a worker, he obtains the value as given by the expressions in (15.11) so that the next time the agent wishes to operate a firm, he has to incur the cost of doing so.

Finally, let us define the *net value* of entrepreneurship as a function of an individual's skill a and ownership status, s,

$$NV\left(\mathbf{p}^t,\mathbf{w}^t\mid a_t^j=A^z,s_t^j=s\right)=V^z\left(\mathbf{p}^t,\mathbf{w}^t\right)-W^z\left(\mathbf{p}^t,\mathbf{w}^t\right)-\left(1-s\right)\lambda b_t,$$

where the last term is the entry cost incurred by agents with $s = 0$. The max operators in (15.12) and (15.14) imply that if $NV > 0$ for an agent, then he prefers to become an entrepreneur.

Who will become an entrepreneur in this economy? Standard arguments (combined with the fact that instantaneous payoffs are strictly monotonic) immediately imply that $V^z\left(\mathbf{p}^t,\mathbf{w}^t\right)$ is strictly monotonic in w_t, T_t and $\Pi^z(\tau_t, w_t)$, so that $V^H\left(\mathbf{p}^t,\mathbf{w}^t\right) > V^L\left(\mathbf{p}^t,\mathbf{w}^t\right)$. By the same arguments,
$NV\left(\mathbf{p}^t, \mathbf{w}^t \mid a_t^j = A^z, s_t^j = s\right)$ is also increasing in $\Pi^z(\tau_t, w_t)$. This in turn implies that for all a and s,

$$NV\left(\mathbf{p}^t, \mathbf{w}^t \mid a_t^j = A^H, s_t^j = 0\right) \geq NV\left(\mathbf{p}^t, \mathbf{w}^t \mid a_t^j, s_t^j = s\right) \geq NV\left(\mathbf{p}^t, \mathbf{w}^t \mid a_t^j = A^L, s_t^j = 1\right).$$

In other words, the net value of entrepreneurship is highest for high-skill existing entrepreneurs, and lowest for low-skill workers. However, it is unclear ex ante whether
$NV\left(\mathbf{p}^t, \mathbf{w}^t \mid a_t^j = A^H, s_t^j = 0\right)$ or $NV\left(\mathbf{p}^t, \mathbf{w}^t \mid a_t^j = A^L, s_t^j = 0\right)$ is greater, that is, whether entrepreneurship is more profitable for incumbents with low skill or for current outsiders with high skill, who will have to pay the entry cost.

We can then define two different types of equilibria:

- *Entry equilibrium* where all entrepreneurs have $a_t^j = A^H$.
- *Sclerotic equilibrium* where agents with $s_t^j = 1$ become entrepreneurs irrespective of their productivity.

An entry equilibrium requires the net value of entrepreneurship to be greater for a non-elite high skill agent than for a low-skill elite, i.e., $NV\left(\mathbf{p}^t, \mathbf{w}^t \mid a_t^j = A^H, s_t^j = 0\right) \geq NV\left(\mathbf{p}^t, \mathbf{w}^t \mid a_t^j = A^L, s_t^j = 1\right)$. To facilitate the analysis let us define w_t^H such that at this wage rate, $NV\left(\mathbf{p}^t, \left[w_t^H, \mathbf{w}^{t+1}\right] \mid a_t^j = A^H, s_t^j = 0\right) = 0$, where we have introduced the notation $\mathbf{w}^t \equiv \left[w_t, \mathbf{w}^{t+1}\right]$. Now using (15.11) and (15.13), we have:

(15.15)
$$w_t^H \equiv \max\left\{\frac{\alpha}{1-\alpha}(1-\tau_t)^{1/\alpha}A^H - b_t + \frac{\beta\left(CV^H\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right) - CW^H\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right)\right)}{\lambda}; 0\right\},$$

and similarly, let w_t^L be such that $NV\left(\mathbf{p}^t, \left[w_t^L, \mathbf{w}^{t+1}\right] \mid a_t^j = A^L, s_t^j = 1\right) = 0$,

(15.16)
$$w_t^L \equiv \max\left\{\frac{\alpha}{1-\alpha}(1-\tau_t)^{1/\alpha}A^L + \frac{\beta\left(CV^L\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right) - CW^L\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right)\right)}{\lambda}; 0\right\}.$$

Both expressions are intuitive. For example, in (15.15), the term $\alpha(1-\tau_t)^{1/\alpha}A^H/(1-\alpha)$ is the per worker profits that a high-skill entrepreneur will make before labor costs. b_t is the per worker entry cost (λb_t divided by λ). Finally, the term $\beta\left(CV^H\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right)-CW^H\left(\mathbf{p}^{t+1},\mathbf{w}^{t+1}\right)\right)$ is the indirect (dynamic) benefit, the additional gain from changing status from a worker to a member of the elite for a high-skill agent. Naturally, this benefit will depend on the sequence of policies, for example, it will be larger when there are greater entry barriers in the future. Consequently, if $w_t < w_t^H$, the total benefit of becoming an entrepreneur for a non-elite high-skill agent exceeds the cost. Equation (15.16) is explained similarly. Given these definitions, the condition for an entry equilibrium to exist at time t can simply be written as

$$w_t^H \geq w_t^L. \tag{15.17}$$

A sclerotic equilibrium emerges, on the other hand, only if the converse of (15.17) holds.

Moreover, in an entry equilibrium, i.e., when (15.17) holds, we must have that $NV\left(\mathbf{p}^t,\mathbf{w}^t \mid a_t^j = A^z, s_t^j = 0\right) = 0$. If it were strictly positive, or in other words, if the wage were less than w_t^H, all agents with high skill would enter entrepreneurship, and since by assumption $M\lambda > 1$ there would be "excess demand" for labor. This argument also shows that $e_t = 1/\lambda$. From (15.9), (15.11) and (15.13), this implies that the equilibrium wage must be

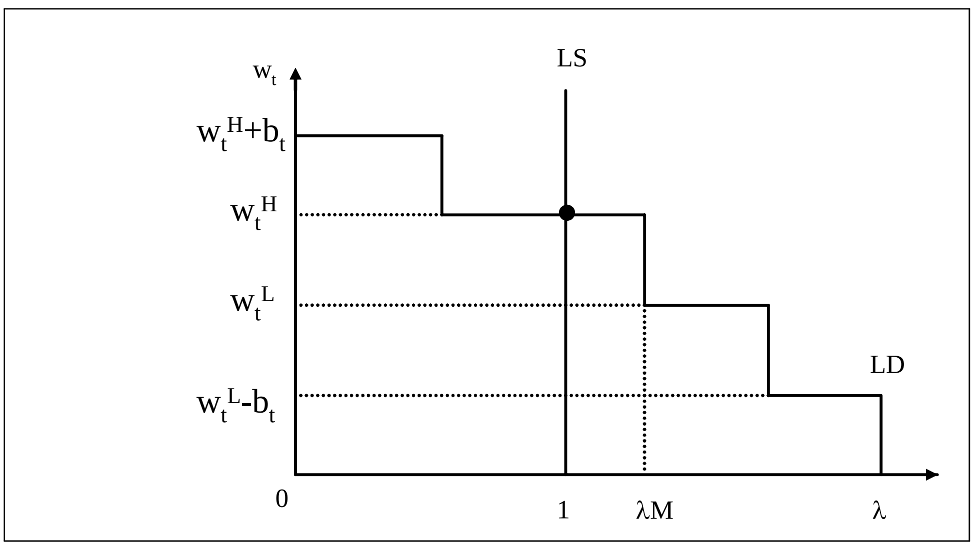
$$w_t^e = w_t^H. \tag{15.18}$$

We can also note that when (15.17) holds, $NV\left(\mathbf{p}^t,\left[w_t^H,\mathbf{w}^{t+1}\right] \mid a_t^j = A^L, s_t^j = 1\right) < 0$, so low-skill incumbents would be worse off if they remained as entrepreneurs at the wage rate w_t^H.

The next figure illustrates the entry equilibrium diagrammatically by plotting labor demand and supply in this economy. Labor supply is constant at 1, while labor demand is decreasing as a function of the wage rate. This figure is drawn for the case where condition (15.17) holds, so that there exists an entry equilibrium. The first portion of the curve shows the willingness to pay of high-skill elites, i.e., agents with $a_t^j = A^H$ and $s_t^j = 1$, which is $w_t^H + b_t$ (since entrepreneurship is as profitable for them as for high-skill entrants and they do not have pay the entry cost). The second portion is for high-skill non-elites, i.e., those with $a_t^j = A^H$ and $s_t^j = 0$, which is by definition w_t^H. These two groups together demand $M\lambda > 1$ workers, ensuring that labor demand intersects labor supply at the wage given in (15.18).

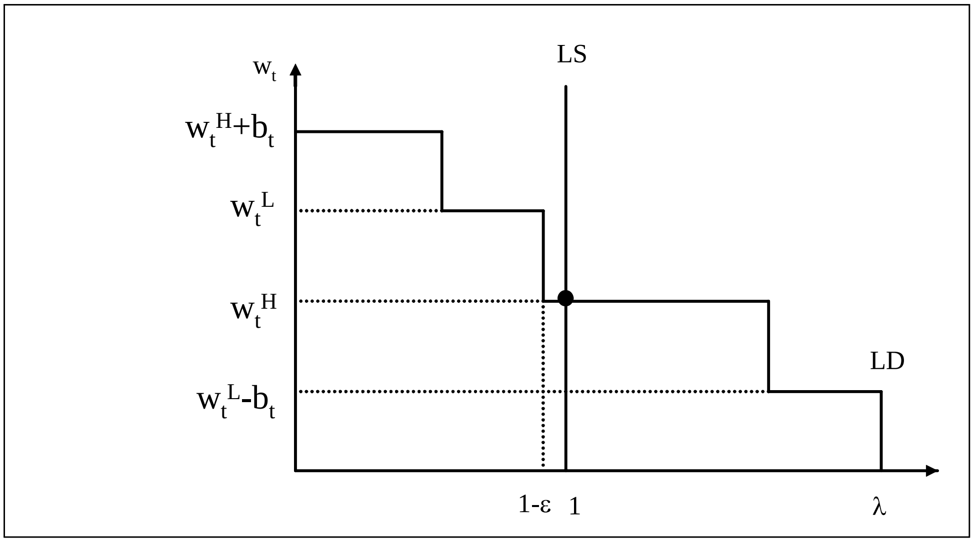
In a sclerotic equilibrium, on the other hand, $w_t^H < w_t^L$, and low-skill elites remain in entrepreneurship, i.e., $s_t^j = s_{t-1}^j$. If there were no deaths so that $\varepsilon = 0$, we would have $\mathbf{e}_t = 1/\lambda$ and for any $w_t \in \left[w_t^H, w_t^L\right]$, labor demand would exactly equal labor supply—$1/\lambda$

agents demanding exactly λ workers each, and a total supply of 1. Hence, there would be multiple equilibrium wages. In contrast, when $\varepsilon > 0$, the measure of entrepreneurs who could pay a wage of w_t^L is $\mathbf{e}_t = (1-\varepsilon)\,\mathbf{e}_{t-1} < 1/\lambda$ for all $t > 0$, thus there would be excess supply of labor at this wage, or at any wage above the lower support of the above range. This implies that the equilibrium wage would be equal to this lower support, w_t^H, which is identical to (15.18). Since at this wage agents with $a_t^j = A^H$ and $s_t^j = 0$ are indifferent between entrepreneurship and working, in equilibrium a sufficient number of them enter entrepreneurship, and $\mathbf{e}_t = 1/\lambda$. In the remainder, we focus on the limiting case of this economy where $\varepsilon \to 0$, which picks w_t^H as the equilibrium wage even when labor supply coincides with labor demand for a range of wages.



Labor supply and labor demand when (15.17) holds and there exists an entry equilibrium.

The following figure illustrates this case diagrammatically. Because (15.17) does not hold in this case, the second flat portion of the labor demand curve is for low-skill elites, i.e., agents with $a_t^j = A^L$ and $s_t^j = 1$, who, given the entry barriers, have a higher marginal product of labor than high-skill non-elites.



Labor supply and labor demand when (15.17) does not hold and there exists a sclerotic equilibrium.

Finally, since at time $t = 0$ we have initially $b_0 = 0$, the initial period equilibrium will feature all high-ability entrepreneurs entering. Since $\lambda M > 1$, the equilibrium wage w_0 must be such that $NV\left(\left[b_t = 0, \tau_t, \mathbf{p}^{t+1}\right], \left[w_0, \mathbf{w}^{t+1}\right] \mid a_t^j = H, s_t^j = 0\right) = 0$, where we used the notation $\mathbf{p}^t \equiv \left[b_t, \tau_t, \mathbf{p}^{t+1}\right]$.

In addition, note that at $t = 0$, all entrepreneurs have high skill, so the economy starts with $\mu_0 = 1$, and the law of motion of the fraction of high-skill entrepreneurs, μ_t, is:

$$\mu_t = \begin{cases} \sigma^H \mu_{t-1} + \sigma^L(1 - \mu_{t-1}) & \text{if (15.17) does not hold} \\ 1 & \text{if (15.17) holds} \end{cases}. \tag{15.19}$$

15.2. Political Equilibrium: Democracy

To obtain a full political equilibrium, we need to determine the policy sequence $\mathbf{p}^t$. Let us consider two extreme cases: (1) Democracy: the policies b_t and τ_t are determined by majoritarian voting, with each agent having one vote. (2) Oligarchy (elite control): the policies b_t and τ_t are determined by majoritarian voting among the elite at time t.

A democratic equilibrium is an MPE where b_t and τ_t are determined by majoritarian voting at time t. The timing of events implies that the tax rate at time t, τ_t, is decided after investment decisions, whereas the entry barriers are decided before. The assumption $\lambda > 2$ above ensures that non-elite agents (workers) are always in the majority.

At the time taxes are set, investments are sunk, agents have already made their occupation choices, and workers are in the majority. Therefore, taxes will be chosen to maximize per capita transfers. We can use equation (15.10) to write tax revenues as:

$$T_t\left(b_t, \tau_t \mid \hat{\tau}_t\right) = \begin{cases} \frac{1}{1-\alpha}\tau_t(1-\hat{\tau}_t)^{\frac{1-\alpha}{\alpha}}\lambda \sum_{j \in \mathbf{S}_t^E} a_t^j & \text{if } \tau_t \leq \delta \\ 0 & \text{if } \tau_t > \delta \end{cases}, \tag{15.20}$$

where $\hat{\tau}_t$ is the tax rate expected by entrepreneurs and τ_t is the actual tax rate set by voters. This expression takes into account that if $\tau_t > \delta$, entrepreneurs will hide their output, and tax revenue will be 0. T_t is a function of the entry barrier, b_t, since this can affect the selection of entrepreneurs, and thus the $\sum_{j \in \mathbf{S}_t^E} a_t^j$ term.

The entry barrier, b_t, is set before occupational choices. Low-productivity non-elite agents, i.e., those with $s_t^j = 0$ and $a_t^j = A^L$, know that they will be workers at time t, and in MPE, the policy choice at time t has no influence on strategies in the future except through its impact on payoff relevant variables. Therefore, the utility of agent j with $s_t^j = 0$ and $a_t^j = A^L$ depends on b_t and τ_t only through the equilibrium wage, $w_t^H\left(b_t \mid \hat{\tau}_t\right)$, and the transfer, $T_t\left(b_t, \tau_t \mid \hat{\tau}_t\right)$, where we have written the equilibrium wage explicitly as a function of the current entry barrier, b_t, and anticipated taxes, $\hat{\tau}_t$. The equilibrium wage depends on $\hat{\tau}_t$ because the labor market clears before tax decisions, in equilibrium, naturally, $\tau_t = \hat{\tau}_t$. Thus $w_t^H\left(b_t \mid \hat{\tau}_t\right)$ is given by (15.18) with the anticipated tax, $\hat{\tau}_t$, replacing τ_t.

High-productivity non-elite agents, i.e., those with $s_t^j = 0$ and $a_t^j = A^H$, may become entrepreneurs, but as the above analysis shows, in this case, $NV\left(\mathbf{p}^t, \mathbf{w}^t \mid a_t^j = A^H, s_t^j = 0\right) = 0$, we have $W^H = W^L$, so their utility is also identical to those with low skill. Consequently, all non-elite agents will choose b_t to maximize $w_t^H\left(b_t \mid \hat{\tau}_t\right) + T_t\left(b_t, \tau_t \mid \hat{\tau}_t\right)$. Since the preferences of all non-elite agents are the same and they are in the majority, the democratic equilibrium will maximize these preferences.

A democratic equilibrium is therefore policy, wage and economic decision sequences $\hat{\mathbf{p}}^t$, $\hat{\mathbf{w}}^t$, and $\hat{\mathbf{x}}^t$ such that $\hat{\mathbf{w}}^t$ and $\hat{\mathbf{x}}^t$ constitute an economic equilibrium given $\hat{\mathbf{p}}^t$, and $\hat{\mathbf{p}}^t$ is such that:

$$\left(\hat{b}_t, \hat{\tau}_t\right) \in \arg\max_{b_t, \tau_t} \left\{w_t^H\left(b_t \mid \hat{\tau}_t\right) + T_t\left(b_t, \tau_t \mid \hat{\tau}_t\right)\right\}.$$

Since $T_t\left(b_t, \tau_t \mid \hat{\tau}_t\right)$ is maximized at $\tau_t = \delta$ and $w_t^H\left(b_t \mid \hat{\tau}_t\right)$ does not depend on τ_t, workers will choose $\tau_t = \delta$. Inspection of (15.18) and (15.20) also shows that wages and tax revenue are both maximized when $b_t = 0$, so the democratic equilibrium will not impose any entry barriers. This is intuitive; workers have nothing to gain by protecting incumbents, and a lot to lose, since such protection reduces labor demand and wages. Since there are no entry barriers, only high-skill agents will become entrepreneurs, or in other words $e_t^j = 1$ only if $a_t^j = A^H$. Given this stationary sequence of MPE policies, we can use the value functions

(15.11) and (15.13) to obtain

$$V^H = W^H = W^L = W = \frac{w^D + T^D}{1-\beta}, \tag{15.21}$$

where w^D is the equilibrium wage in democracy, and T^D is the level of transfers, given by δY^D. The following proposition therefore follows immediately (proof in the text):

PROPOSITION 15.1. *A democratic equilibrium always features $\tau_t = \delta$ and $b_t = 0$. Moreover, we have $e_t^j = 1$ if and only if $a_t^j = A^H$, so $\mu_t = 1$. The equilibrium wage rate is given by*

$$w_t^D = w^D \equiv \frac{\alpha}{1-\alpha}(1-\delta)^{1/\alpha} A^H, \tag{15.22}$$

and the aggregate output is

$$Y_t^D = Y^D \equiv \frac{1}{1-\alpha}(1-\delta)^{\frac{1-\alpha}{\alpha}} A^H. \tag{15.23}$$

An important feature of this equilibrium is that aggregate output is constant over time, which will contrast with the oligarchic equilibrium. Another noteworthy feature is that there is perfect equality because the excess supply of high-skill entrepreneurs ensures that they receive no rents.

15.3. Political Equilibrium: Oligarchy

In oligarchy, policies are determined by majoritarian voting among the elite. At the time of voting over the entry barriers, b_t, the elite are those with $s_t = 1$, and at the time of voting over the taxes, τ_t, the elite are those with $e_t = 1$.

Let us start with the taxation decision among those with $e_t = 1$. It can be shown that as long as

$$\lambda \geq \frac{1}{2}\frac{A^H}{A^L} + \frac{1}{2}, \tag{15.24}$$

then both high-skill and low-skill entrepreneurs prefer zero taxes, i.e., $\tau_t = 0$. Intuitively, condition (15.24) requires the productivity gap between low and high-skill elites not to be so large that low-skill elites wish to tax profits in order to indirectly transfer resources from high-skill entrepreneurs to themselves.

When condition (15.24) holds, the oligarchy will always choose $\tau_t = 0$. Then anticipating this tax choice, at the stage of deciding the entry barriers, high-skill entrepreneurs would like to maximize $V^H\left(\left[b_t, 0, \mathbf{p}^{t+1}\right], \left[w_t, \mathbf{w}^{t+1}\right]\right)$, while low-skill entrepreneurs would like to maximize $V^L\left(\left[b_t, 0, \mathbf{p}^{t+1}\right], \left[w_t, \mathbf{w}^{t+1}\right]\right)$. Both of these are maximized by setting a level of the entry barrier that ensures the minimum level of equilibrium wages. Equilibrium wage, given in (15.18), will be minimized at $w_t^H = 0$, by choosing any

$$b_t \geq b_t^E \equiv \frac{\alpha}{1-\alpha} A^H + \beta \left(\frac{CV^H\left(\mathbf{p}^{t+1}, \mathbf{w}^{t+1}\right) - CW^H\left(\mathbf{p}^{t+1}, \mathbf{w}^{t+1}\right)}{\lambda} \right). \tag{15.25}$$

Without loss of any generality, set $b_t = b_t^E$.

An oligarchic equilibrium then can be defined as a policy sequence $\hat{\mathbf{p}}^t$, wage sequence $\hat{\mathbf{w}}^t$, and economic decisions $\hat{\mathbf{x}}^t$ such that $\hat{\mathbf{w}}^t$ and $\hat{\mathbf{x}}^t$ constitute an economic equilibrium given $\hat{\mathbf{p}}^t$, and $\hat{\mathbf{p}}^t$ is such $\tau_{t+n} = 0$ and $b_{t+n} = b_{t+n}^E$ for all $n \geq 0$. In the oligarchic equilibrium, there is no redistributive taxation and entry barriers are sufficiently high to ensure a sclerotic equilibrium with zero wages.

Imposing $w_{t+n}^e = 0$ for all $n \geq 0$, we can solve for the values of high- and low-skill entrepreneurs from the value functions (15.13). Let $\tilde{V}^z = V^z\left(\left[b_t^E, 0, \mathbf{p}^{t+1}\right], \left[0, \mathbf{w}^{t+1}\right]\right) = V_{t+1}^z\left(\mathbf{p}^{t+1}, \mathbf{w}^{t+1}\right)$, then

$$\tilde{V}^L = \frac{1}{1-\beta}\left[\frac{\alpha\lambda}{1-\alpha}\frac{\left(1-\beta\sigma^H\right)A^L + \beta\sigma^L A^H}{\left(1-\beta\left(\sigma^H - \sigma^L\right)\right)}\right], \tag{15.26}$$

and

$$\tilde{V}^H = \frac{1}{1-\beta}\left[\frac{\alpha\lambda}{1-\alpha}\frac{\left(1-\beta\left(1-\sigma^L\right)\right)A^H + \beta\left(1-\sigma^H\right)A^L}{\left(1-\beta\left(\sigma^H - \sigma^L\right)\right)}\right]. \tag{15.27}$$

These expressions are intuitive. First, consider $\tilde{V}^L$ and the case where $\beta \to 1$; then, starting in the state L, an entrepreneur will spend a fraction $\sigma^L / \left(1 - \sigma^H + \sigma^L\right)$ of his future with low skill A^L and a fraction $\left(1-\sigma^H\right) / \left(1 - \sigma^H + \sigma^L\right)$ with high skill A^H. $\beta < 1$ implies discounting and the low-skill states which occur sooner are weighed more heavily (since the agent starts out as low skill). The intuition for $\tilde{V}^H$ is identical.

Since there will be zero equilibrium wages and no transfers, it is clear that $W = 0$ for all workers. Therefore, for a high-skill worker, $NV = \tilde{V}^H - b$, implying that the lower bound of the set $\mathbf{B}_t$ is

$$b_t = b^E \equiv \frac{1}{1-\beta}\left[\frac{\alpha\lambda}{1-\alpha}\frac{\left(1-\beta\left(1-\sigma^L\right)\right)A^H + \beta\left(1-\sigma^H\right)A^L}{\left(1-\beta\left(\sigma^H - \sigma^L\right)\right)}\right]. \tag{15.28}$$

In this oligarchic equilibrium, aggregate output is:

$$Y_t^E = \mu_t \frac{1}{1-\alpha} A^H + (1-\mu_t)\frac{1}{1-\alpha} A^L, \tag{15.29}$$

where $\mu_t = \sigma^H \mu_{t-1} + \sigma^L (1 - \mu_{t-1})$ as given by (15.19), with $\mu_0 = 1$. Since μ_t is a decreasing sequence converging to M, aggregate output Y_t^E is also decreasing over time with:

$$\lim_{t\to\infty} Y_t^E = Y_\infty^E \equiv \frac{1}{1-\alpha}\left(A^L + M(A^H - A^L)\right). \tag{15.30}$$

The reason for this is that as time goes by, the comparative advantage of the members of the elite in entrepreneurship gradually disappears because of the imperfect correlation between ability over time.

Another important feature of this equilibrium is that there is a high degree of (earnings) inequality. Wages are equal to 0, while entrepreneurs earn positive profits—in fact, each entrepreneur earns λY_t^E (gross of investment expenses), and their total earnings equal aggregate output. This contrasts with relative equality in democracy.

Proposition 15.2. *Suppose that condition (15.24) holds. Then an oligarchic equilibrium features $\tau_t = 0$ and $b_t = b^E$ as given by (15.28), and the equilibrium is sclerotic, with equilibrium wages $w_t^e = 0$, and fraction of high-skill entrepreneurs $\mu_t = \sigma^H \mu_{t-1} + \sigma^L (1 - \mu_{t-1})$ starting with $\mu_0 = 1$. Aggregate output is given by (14.15) and decreases over time starting at $Y_0^E = \frac{1}{1-\alpha} A^H$ with $\lim_{t\to\infty} Y_t^E = Y_\infty^E$ as given by (15.30).*

15.4. Comparison Between Democracy and Oligarchy

The first important result in the comparison between democracy and oligarchy is that aggregate output in the initial period of the oligarchic equilibrium, Y_0^E, is greater than the constant level of output in the democratic equilibrium, Y^D. In other words, as long as $\delta > 0$, then

$$Y^D = \frac{1}{1-\alpha}(1-\delta)^{\frac{1-\alpha}{\alpha}} A^H < Y_0^E = \frac{1}{1-\alpha} A^H.$$

Therefore, for all $\delta > 0$, oligarchy initially generates greater output than democracy, because it is protecting the property rights of entrepreneurs. However, the analysis also shows that Y_t^E declines over time, while Y^D is constant. Consequently, the oligarchic economy may subsequently fall behind the democratic society. Whether it does so or not depends on whether Y^D is greater than Y_∞^E as given by (15.30). This will be the case if $(1-\delta)^{\frac{1-\alpha}{\alpha}} A^H / (1-\alpha) > \left(A^L + M(A^H - A^L)\right) / (1-\alpha)$, or if

$$(1-\delta)^{\frac{1-\alpha}{\alpha}} > \frac{A^L}{A^H} + M\left(1 - \frac{A^L}{A^H}\right). \tag{15.31}$$

If condition (15.31) holds, then at some point the democratic society will overtake ("leapfrog") the oligarchic society. This discussion and inspection of (15.31) immediately establish the following result:

Proposition 15.3. *Suppose that condition (15.24) holds. Then at $t = 0$, aggregate output is higher in an oligarchic society than in a democratic society, i.e., $Y_0^E > Y^D$. If (15.31) does not hold, then aggregate output in oligarchy is always higher than in democracy, i.e., $Y_t^E > Y^D$ for all t. If (15.31) holds, then there exists $t' \in N$ such that for $t \leq t'$, $Y_t^E \geq Y^D$*

and for $t > t'$, $Y_t^E < Y^D$, so that the democratic society leapfrogs the oligarchic society. Leapfrogging is more likely when δ, A^L/A^H and M are low.

There are three important conclusions that follow from this proposition. Oligarchies are more likely to be relatively inefficient in the long run:

when δ is low, meaning that democracy is unable to pursue highly populist policies with a high degree of redistribution away from entrepreneurs/capitalists. The parameter δ may correspond both to certain institutional impediments limiting redistribution, or more interestingly, to the specificity of assets in the economy; with greater specificity, taxes will be limited, and redistributive distortions may be less important.

when A^H is high relative to A^L, so that comparative advantage and thus selecting the high-skill agents for entrepreneurship are important for the efficient allocation of resources.

M is low, so that a random selection of agents contains a small fraction of high-skill agents, making oligarchic sclerosis highly distortionary. Alternatively, M is low when σ^H is low, so oligarchies are more likely to lead to low output in the long run when the efficient allocation of resources requires a high degree of churning in the ranks of entrepreneurs.

On the other hand, if the extent of taxation in democracy is high and the failure to allocate the right agents to entrepreneurship only has limited costs, then an oligarchic society may always achieve greater output than a democracy.

These comparative static results may be useful in interpreting why the Northeastern United States so conclusively outperformed the Caribbean plantation economies during the 19th century. First, the American democracy was not highly redistributive, corresponding to low δ in terms of the model here. More important, the 19th century was the age of industry and commerce, where the allocation of high-skill agents to entrepreneurship was probably quite important, and potentially only a small fraction of the population were really talented as inventors and entrepreneurs. This can be thought of as a low value of A^L/A^H and M.

It is important to note that Proposition 15.3 compares the income and consumption levels, and not the welfare levels in the two regimes. Since in oligarchy there are high levels of consumption early on, the average expected discounted utility at time $t = 0$ could be higher than in democracy, even when (15.31) holds.

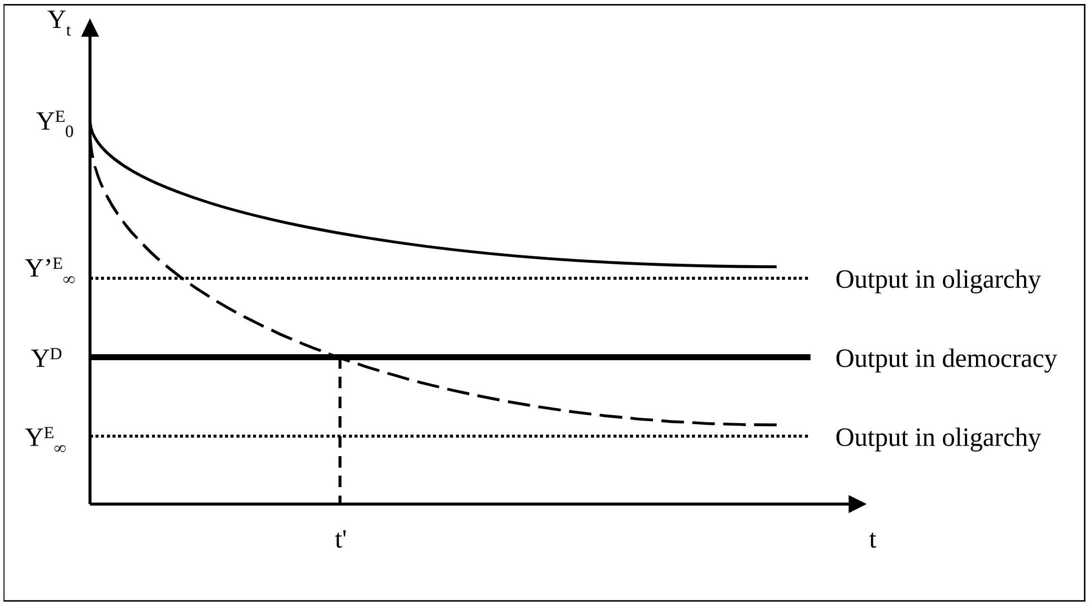


Figure 3: Comparison of aggregate output in democracy and oligarchy. The dashed curve depicts output in oligarchy when (15.31) holds, and the solid line when it does not.

Figure 3 illustrates both the case in which (15.31) holds and the converse case diagrammatically. The thick flat line shows the level of aggregate output in democracy, Y^D. The other two curves depict the level of output in oligarchy, Y_t^E, as a function of time for the case where (15.31) holds and for the case where it does not. Both of these curves asymptote to some limit, either Y_∞^E or $Y_\infty'^E$, which may lie below or above Y^D. The dashed curve shows the case where (15.31) holds, so after date t', oligarchy generates less aggregate output than democracy. When (15.31) does not hold, the solid curve applies, and aggregate output in oligarchy asymptotes to a level higher than Y^D.

It is also useful to point out that some alternative arrangements would dominate both democracy and oligarchy in terms of aggregate output performance. For example, a society may restrict the amount of redistribution by placing a constitutional limit on taxation, and let the decisions on entry barriers be made democratically. Alternately, it may prevent entry barriers constitutionally, and place the taxation decisions in the hands of the oligarchy. The perspective here is that these arrangements are not possible in practice because of the inherent commitment problem in politics: those with the political power in their hands make the policy decisions, and previous promises are not necessarily credible. Consequently, it is not possible to give political power to incumbent producers, and then expect them not to use their political power to erect entry barriers, or to vest political power with the poorer agents and expect them not to favor redistribution.

What about inequality and the preferences of different groups over regimes? First, it is straightforward to see that oligarchy always generates more (consumption) inequality relative to democracy, since the latter has perfect equality—the net incomes and consumption of all agents are equalized in democracy because of the excess supply of high-skill entrepreneurs.

Moreover, non-elites are always better off in democracy than in oligarchy, where they receive zero income. In contrast, and more interestingly, it is possible for low-skill elites to be better off in democracy than in oligarchy (though high-skill elites are always better off in oligarchy). This point will play a role below in leading to smooth transitions to democracy from oligarchy, so it is useful to understand the intuition. Recall that the utility of low-skill elites in oligarchy is given by (15.26) above, whereas combining (15.21), (15.22) and (15.23), these low-skill agents in democracy would receive

$$W^L = \frac{1}{1-\beta}\left[\left(\frac{\alpha + (1-\delta)\,\delta}{1-\alpha}(1-\delta)^{(1-\alpha)/\alpha}\right)A^H\right].$$

Comparing this expression to (15.26) makes it clear that if δ, A^L/A^H, σ^L and/or λ are sufficiently low, these low-skill elites would be better off in democracy than in oligarchy. More specifically, we have:

Proposition 15.4. *If*

$$\alpha\lambda\frac{\left(1-\beta\sigma^H\right)A^L/A^H + \beta\sigma^L}{\left(1-\beta\left(\sigma^H-\sigma^L\right)\right)} < \left(\left(\alpha + (1-\delta)\,\delta\right)(1-\delta)^{(1-\alpha)/\alpha}\right), \tag{15.32}$$

then low-skill elites would be better off in democracy.

Despite this result low-skill elites, even when (15.32) holds, prefer to remain in entrepreneurship. This is because, given the structure of the political game, if the low-skill incumbent elites give up entrepreneurship, the new entrepreneurs will make the political choices, and they will naturally choose high entry barriers and no redistribution. Therefore, by quitting entrepreneurship, low-skill elites would be giving up their political power. Consequently, they are choosing between being elites and being workers in oligarchy, and clearly, the former is preferred.

15.5. New Technologies

Another important idea is that democracies may be more flexible institutions than oligarchies. The model here provides a potential way of thinking about this claim, based on the idea that established groups may be opposed (or less open) to new technologies because of their existing rents (see also Krusell and Rios-Rull, 1996, Olson, 1982).

Suppose that at some date $t' > 0$, there is an unanticipated and exogenous arrival of a new technology, enabling entrepreneur j to produce:

$$y_t^j = \frac{1}{1-\alpha}(\psi \hat{a}_t^j)^\alpha (k_t^j)^{1-\alpha}(l_t^j)^\alpha,$$

where $\psi > 1$ and $\hat{a}_t^j$ is the talent of this entrepreneur with the new technology. Assuming that $l_t^j = \lambda$ for the new technology as well, entrepreneur j's output can be written as

$$\max\left\{\frac{1}{1-\alpha}(\psi \hat{a}_t^j)^\alpha (k_t^j)^{1-\alpha}\lambda^\alpha, \frac{1}{1-\alpha}(a_t^j)^\alpha (k_t^j)^{1-\alpha}\lambda^\alpha\right\}.$$

Also to simplify the discussion, assume that the law of motion of $\hat{a}_t^j$ is similar to that of a_t^j, given by

$$\hat{a}_{t+1}^j = \begin{cases} A^H & \text{with probability} \quad \sigma^H & \text{if } \hat{a}_t^j = A^H \\ A^H & \text{with probability} \quad \sigma^L & \text{if } \hat{a}_t^j = A^L \\ A^L & \text{with probability } 1-\sigma^H & \text{if } \hat{a}_t^j = A^H \\ A^L & \text{with probability } 1-\sigma^L & \text{if } \hat{a}_t^j = A^L \end{cases}, \tag{15.33}$$

for all $t > t'$ and $Pr\left(\hat{a}_t^j = A^H \mid a_{\tilde{t}}^j\right) = M$ for any $t, \tilde{t}$ and $a_{\tilde{t}}^j$. In other words, $\hat{a}_t^j$, and in particular $\hat{a}_{t'}^j$, is independent of past and future a_t^j's. This implies that $\hat{a}_{t'}^j = A^H$ with probability M and $\hat{a}_{t'}^j = A^L$ with probability $1-M$ irrespective of the talent of the individual with the old technology. This is reasonable since new technologies exploit different skills and create different comparative advantages than the old ones.

It is straightforward to see that the structure of the democratic equilibrium is not affected, and at the time t', agents with comparative advantage for the new technology become the entrepreneurs, so aggregate output jumps from Y^D as given by (15.23) to

$$\hat{Y}^D \equiv \frac{\psi}{1-\alpha}(1-\delta)^{\frac{1-\alpha}{\alpha}} A^H.$$

In contrast, in oligarchy, the elites are in power at time t', and they would like to remain the entrepreneurs even if they do not have comparative advantage for working with the new technology. How aggregate output in the oligarchic equilibrium changes after date t' depends on whether $\psi A^L > A^H$ or not. If it is, then all incumbents switch to the new technology and aggregate output in the oligarchic equilibrium at date t' jumps up to

$$\hat{Y}_\infty^E \equiv \frac{\psi}{1-\alpha}\left(A^L + M(A^H - A^L)\right),$$

and remains at this level thereafter. This is because $\hat{a}_t^j$ and a_t^j are independent, so applying the weak law of large numbers, exactly a fraction M of the elite have high skill with the new technology, and the remainder have low skill.

If, on the other hand, $\psi A^L < A^H$, then those elites who have high skill with the old technology but turn out to have low skill with the new technology prefer to use the old technology, and aggregate output after date t' follows the law of motion

$$\tilde{Y}_t^E = \frac{1}{1-\alpha}\left[M\psi A^H + \mu_t (1-M) A^H + (1-\mu_t)(1-M)\psi A^L\right],$$

with μ_t given by (15.19) as before. Intuitively, now the members of the elite who have high skill with the new technology and those who have low skill with the old technology switch to the new technology, while those with high skill with the old and low skill with the new remain with the old technology (they switch to new technology only when they lose their high-skill status with the old technology). As a result, we have that $\tilde{Y}_t^E$, just like Y_t^E before, is decreasing over time, with

$$\lim_{t\to\infty} \tilde{Y}_t^E = \frac{1}{1-\alpha}\left[M\psi A^H + M(1-M) A^H + (1-M)^2 \psi A^L\right].$$

It is also straightforward to verify that, as long as $Y_\infty^E \leq Y^D$, the gap $\hat{Y}^D - \hat{Y}^E$ or $\hat{Y}^D - \tilde{Y}_t^E$ (or whichever is relevant) is always greater than the output gap before the arrival of the new technology, $Y^D - Y_t^E$ (for $t > t'$). In other words, the arrival of a new technology creates a further advantage for the democratic society. In fact, it may have been the case that $Y^D - Y_t^E < 0$, i.e., before the arrival of the new technology, the oligarchic society was richer than the democratic society, but the ranking is reversed after the arrival of the new technology at date t'. Intuitively, this is because the democratic society immediately makes full use of the new technology by allowing those who have a comparative advantage to enter entrepreneurship, while the oligarchic society typically fails to do so, and therefore has greater difficulty adapting to technological change.

15.6. Regime Dynamics: Smooth Transitions

Let us first discuss how oligarchy may "voluntarily" transform itself into a democracy in a modified version of the model. We change one assumption from the baseline model. We allow the current elite to also vote to disband oligarchy, upon which a permanent democracy is established. We denote this choice by $d_t \in \{0,1\}$, with 0 standing for continuation with the oligarchic regime. To describe the law of motion of the political regime, let us denote oligarchy by $D_t = 0$ and democracy by $D_t = 1$. Since transition to democracy is permanent, we have

$$D_t = \begin{cases} 0 & \text{if } d_{t-n} = 0 \text{ for all } n \geq 0 \\ 1 & \text{if } d_{t-n} = 1 \text{ for some } n \geq 0 \end{cases}.$$

Voting over d_t in oligarchy is at the same time as voting over b_t (there are no votes over d_t in democracy, since a transition to democracy is permanent), so agents with $s_t = 1$ vote over these choices (recall the timing of events above). We assume that after the vote for $d_t = 1$, there is immediate democratization and all agents participate in the vote over taxes starting in period t.

First, imagine a situation where condition (15.32) does not hold so that even low-skill elites are better off in oligarchy. Then all elites will always vote for $d_t = 0$, and as in Proposition 15.2, they will also choose $b_t = b^E$ and $\tau_t = 0$. Hence, in this case, the equilibrium remains oligarchic throughout.

What happens when (15.32) holds? Current low-skill elites, i.e., those with $s_t = 1$ and $a_t = A^L$, would be better off in democracy (recall Proposition 15.4). If they vote for $d_t = 0$, they stay in oligarchy, which gives them a lower payoff. If instead they vote for $d_t = 1$ and $b_t = 0$, then this will immediately take us to a democratic equilibrium; following this vote, high-skill agents enter entrepreneurship and there are redistributive taxes at the rate $\tau_t = \delta$ as in Proposition 15.1.

Consequently, when they are in the majority, low-skill elites prefer to transition to a permanent democracy by voting for $d_t = 1$. Since $\mu_0 = 1$, they are initially in the minority, and the oligarchic equilibrium applies. However, because in the oligarchic equilibrium the fraction of high-skill elites decreases over time, the low-skill agents eventually become the majority and choose to disband the oligarchy. This discussion immediately leads to the following proposition:

Proposition 15.5. *Suppose that (15.24) holds, the society starts as oligarchic and agents are sufficiently patient (i.e., $\beta \geq \bar{\beta}$ for some $\bar{\beta} < 1$). If (15.32) does not hold, then for all t the society remains oligarchic with $d_t = 0$; the equilibrium involves no redistribution, $\tau_t = 0$ and high entry barriers, $b_t = b^E$ as given by (15.28), and the fraction of high-skill entrepreneurs is given by $\mu_t = \sigma^H \mu_{t-1} + \sigma^L (1 - \mu_{t-1})$ starting with $\mu_0 = 1$. If (15.32) holds, then the society remains oligarchic, $d_t = 0$, with no redistribution, $\tau_t = 0$ and high entry barriers, $b_t = b_t^E$ as given by (15.25) until date $t = \tilde{t}$ where $\tilde{t} = \min t' \in N$ such that $\mu_{t'} \leq 1/2$ (whereby $\mu_t = \sigma^H \mu_{t-1} + \sigma^L (1 - \mu_{t-1})$ for $t < \tilde{t}$ starting with $\mu_0 = 1$). At $\tilde{t}$, the society transitions to democracy with $d_{\tilde{t}} = 1$, and for $t \geq \tilde{t}$, we have $\tau_t = \delta$, $b_t = 0$ and $\mu_t = 1$.*

Intuitively, when (15.32) holds, low-skill entrepreneurs are better off transitioning to democracy than remaining in the oligarchic society, while high-skill entrepreneurs are always better off in oligarchy. As a result, the society remains oligarchic as long as high-skill entrepreneurs are in the majority, i.e., as long as $t < \tilde{t}$, and the first period in which low-skill entrepreneurs become majority within the oligarchy, i.e., at $\tilde{t}$ such that $\mu_t < 1/2$ for the first time, the oligarchy disbands itself transitioning to a democratic regime (and at that point μ_t jumps up to 1).

This configuration is especially interesting when (15.31) holds such that oligarchy ultimately leads to lower output than democracy. In this case, if (15.32) holds, oligarchy transitions into democracy avoiding the long-run adverse efficiency consequences of oligarchy—but when this condition does not hold, oligarchy survives forever with negative consequences for efficiency and output. This extension therefore provide a simple model for thinking about how a society can transform itself from oligarchy to a more democratic system, before the oligarchic regime becomes excessively costly. Interestingly, however, the reason for the transition from oligarchy to democracy is *not* increased inefficiency in the oligarchy, but conflict between high and low-skill agents *within* the oligarchy; the transition takes place when the low-skill elites become the majority.

15.7. Regime Dynamics: Conflict Over Regimes

Finally, let us consider perhaps even a more realistic case in which while oligarchy is economically costly, the oligarchs still prefer this regime to democracy. Therefore, any change will result from political conflict, or exercise of political power. To capture these issues, consider the following extension where the distribution of income affects the conflict over political regime. In particular, suppose that (15.32) does not hold, so that while non-elites would like to switch from oligarchy to democracy, both high-skill and low-skill elites would like to preserve the oligarchic system. How will these conflicting interests between elites and non-elites be mediated? A plausible answer is that there is no easy compromise, and whichever group is politically or militarily more powerful will prevail. This is the perspective adopted here, and the political or military power of a group is linked to its economic power. In other words, in the conflict between the elite and the non-elites, the likelihood that the elite will prevail is increasing in their relative economic strength. This assumption is plausible: a nondemocratic regime often transforms itself into a more democratic one in the face of threats or unrest, and the degree to which the regime will be able to protect itself depends on the resources available to it.

We model the effect of economic power on political power in a reduced-form way, and assume that the probability that an oligarchy switches to democracy is $q_t^D = q^D(\Delta\mathcal{W}_{t-1})$, where $\Delta\mathcal{W}_{t-1} = \mathcal{W}_{t-1}^E - \mathcal{W}_{t-1}^W$ is the wealth gap, the difference between the levels of wealth of the elite and the citizens at time $t-1$. The assumption that economic power buys political power is equivalent to $q^D(\cdot)$ being decreasing. We also assume that when democratic, a society becomes oligarchic with probability

$$q_t^O = q^O\left(\Delta\mathcal{W}_{t-1}\right)$$

where now $q^O(\cdot)$ is a non-decreasing function, with $q^O(0) = 0$, so that with perfect equality, there is no danger of switching back to oligarchy. Here $\Delta\mathcal{W}_t$ refers to the income gap between the initial elite (those with with $s_1^j = 1$) and the citizens. This discussion immediately leads to the following law of motion for D_t:

$$D_t = \begin{cases} 0 & \text{with probability } 1 - q^D\left(\Delta\mathcal{W}_{t-1}\right) & \text{if } D_{t-1} = 0 \\ 1 & \text{with probability } q^D\left(\Delta\mathcal{W}_{t-1}\right) & \text{if } D_{t-1} = 0 \\ 0 & \text{with probability } q^O\left(\Delta\mathcal{W}_{t-1}\right) & \text{if } D_{t-1} = 1 \\ 1 & \text{with probability } 1 - q^O\left(\Delta\mathcal{W}_{t-1}\right) & \text{if } D_{t-1} = 1 \end{cases}, \tag{15.34}$$

We also assume that each agent saves out of current income at a constant (exogenous) rate $\nu < 1$, which considerably simplifies the analysis. Let us start with the oligarchy, $D_{t-1} = 0$. Since citizens earn zero income in oligarchy, when we have $\mathcal{W}_t^W = 0$ and $\Delta\mathcal{W}_t = \mathcal{W}_t^E$ for all t. Therefore,

$$\Delta\mathcal{W}_t = \nu\left(\Delta\mathcal{W}_{t-1} + \lambda Y_{t-1}^E\right). \tag{15.35}$$

This implies

$$\Delta\mathcal{W}_t = \lambda\sum_{n=1}^{t}\nu^n Y_{t-n}^E, \text{ and } \lim_{t\to\infty}\Delta\mathcal{W}_t = \Delta\mathcal{W}_\infty \equiv \frac{\lambda Y_\infty^E}{1-\nu} \text{ if } D_{t-1} = 0, \tag{15.36}$$

where Y_∞^E is given by (15.30). We can prove that Y_t^E is still given by (15.29) above. In addition, let us assume that $\mathcal{W}_0^E$ is small, in particular, less than $\Delta\mathcal{W}_\infty$, which amounts to assuming that the wealth of the elite, and thus the wealth gap, will be increasing over time

Now two interesting cases can be distinguished: (1) There exists $\overline{\Delta\mathcal{W}} < \Delta\mathcal{W}_\infty$ such that $q^D\left(\overline{\Delta\mathcal{W}}\right) = 0$. (2) $q^D\left(\Delta\mathcal{W}_\infty\right) > 0$. [A third possibility is $\lim_{t\to\infty} q^D\left(\Delta\mathcal{W}_t\right) = 0$, in which case the nature of the limiting equilibrium depends on the rate at which $q^D\left(\Delta\mathcal{W}_t\right)$ converges to 0.]

In the former case, there also exists $\bar{t}$ such that for all $t \geq \bar{t}$, we have $\Delta\mathcal{W}_t \geq \overline{\Delta\mathcal{W}}$, so if the economy does not switch to democracy before $\bar{t}$, it will be permanently stuck in oligarchy. In the second case, as time passes, the economy will switch out of oligarchy into democracy with probability 1.

In contrast to oligarchy, in democracy, all agents earn the same amount, so when $D_{t+k} = 1$ for all $k \geq 0$,

$$\Delta\mathcal{W}_{t+1} = \nu\Delta\mathcal{W}_t \text{ and } \lim_{t\to\infty}\Delta\mathcal{W}_t = 0. \tag{15.37}$$

Consequently, an equilibrium with regime changes is a policy sequence $\hat{\mathbf{p}}^t$, a wage sequence $\hat{\mathbf{w}}^t$ and economic decisions $\hat{\mathbf{x}}^t$ such that $\hat{\mathbf{w}}^t$ and $\hat{\mathbf{x}}^t$ constitute an economic equilibrium given $\hat{\mathbf{p}}^t$, and if $D_t = 0$, then $\hat{\mathbf{p}}^t$ is the oligarchic equilibrium policy sequence and $\Delta\mathcal{W}_{t+1}$ is given by (15.35), and if $D_t = 1$, then $\hat{\mathbf{p}}^t$ is the democratic equilibrium policy sequence and $\Delta\mathcal{W}_{t+1}$ is given by (15.37). D_t is in turn given by (15.34) with $D_0 = 0$.

In this case, we can prove the following result:

Proposition 15.6. *Suppose that (15.24) holds, (15.32) does not hold, and there exists* $\overline{\Delta\mathcal{W}} < \Delta W_\infty$ *such that* $q^O\left(\overline{\Delta\mathcal{W}}\right) = 0$ *where* ΔW_∞ *is given by (15.36), and let* $\bar{t} = 1 + \min t \in N : \Delta W_t \geq \overline{\Delta\mathcal{W}}$*. Then:*

- *If $D_0 = 0$ and $D_t = 0$ for all $t \leq \bar{t}$, then $D_t = 0$ for all t; i.e., if a society starts oligarchic and remains oligarchic until $\bar{t}$, it will always remain oligarchic.*
- *If $D_0 = 0$ and $D_{t'} = 1$ for the first time in $t' \geq 0$ then $D_t = 1$ for all $t \geq t'$; i.e., if a society becomes democratic at t', it will remain democratic thereafter, and if it starts as democratic, it will always remain democratic.*
- *If $D_0 = 0$ and $D_{t'} = 1$, then the probability of switching back to oligarchy for the first time at time $t > t'$ after the switch to democracy at t', $Q_{t|t'}$ is non-increasing in t and non-decreasing in t', with $\lim_{t\to\infty} Q_{t|t'} = 0$—i.e., a society faces the highest probability of switching back to oligarchy immediately after the switch from oligarchy to democracy, and this probability is higher if it has spent a longer time in oligarchy.*

The most interesting result contained in this proposition is that of path dependence. Of two identical societies, if one starts oligarchic and the other as democratic, they can follow very different political and economic trajectories. The initial democracy will always remain democratic, generate an income level Y^D and an equal distribution of income, ensuring that $\Delta Y_t = 0$ and therefore $q^O = 0$. On the other hand, if it starts oligarchic, it will follow the oligarchic equilibrium, with an unequal distribution of income. The greater income of the elites will enable them to have the power to sustain the oligarchic equilibrium, and if there is no transition to democracy until some date $\bar{t}$ (which may be $t = 0$), they will be sufficiently richer than the workers to be able to sustain the oligarchic regime forever.

This type of path dependence provides a potential explanation for the different development experiences in the Americas and in the former European colonies, discussed by Engerman and Sokoloff (1997) and Acemoglu, done some and Robinson (2002). Similar path dependence also results when we compare two societies that start out as oligarchies, but one of them switches to democracy early on, while the other remains oligarchic until income inequality is wide enough to prevent a transition to democracy. Finally, the analysis also shows that newly-created democracies will have the greatest instability and danger of switching back to oligarchy, because wealth inequality between the previous elite and citizens is highest. This inequality disappears over time, and democracy becomes more likely to be consolidated.

15.8. References

(1) Acemoglu, Daron (2003) "The Form of Property Rights: Oligarchic versus Democratic Societies," NBER Working Paper #10037. Updated version: http://econ-www.mit.edu/faculty/download_pdf.php?id=832.

(2) Acemoglu, Daron, Simon Johnson, and James A. Robinson (2002) "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution," *Quarterly Journal of Economics* vol 117, November, 1231-1294.

(3) Engerman, Stanley L. and Kenneth L. Sokoloff (1997) "Factor Endowments, Institutions, and Differential Paths of Growth among New World Economies," in S.H. Haber ed. *How Latin America Fell Behind*, Stanford University Press, Stanford CA.

(4) Olson, Mancur (1982) *The Rise and Decline of Nations: Economic Growth, Stagflation, and Economic Rigidities,* Yale University Press, New Haven and London.

(5) Krusell, Per and Jose-Victor Rios-Rull (1996). "Vested Interests in a Positive Theory of Stagnation and Growth," *Review of Economic Studies* 63, 301-329.

CHAPTER 16

Power

In the conceptual approaches we have been discussing so far, economic development arises when a society invests, saves, innovates etc. Whether or not this happens depends on the incentives and constraints that agents and groups face. This incentive environment is determined by the institutions and policies. But institutions and policies don't arise at random, but are rather the outcome of a process of collective choice. Different people and different groups have different induced preferences over institutions and policies and the nature of these preferences and how they are aggregated is going to determine what institutions and policies get chosen. Whose preferences win? One way of saying this is that it is the preferences of people or groups that have most power that will determine collective choices over institutions and policies and ultimately whether a society prospers or stagnates.

What is power and where does it come from? Most definitions of power are not very revealing and go along the lines of "A has power over B to the extent that he can get B to do something he would otherwise not have done." This at least reveals that power is a relational concept. In our context, an individual or group has power in some dimension to the extent that their preferences determine the outcome. If you thought of allocating the resources of society by maximizing a weighted sum of the utilities of individuals in society then you could think of the welfare weight on the utility of a particular individual as being his 'power' since this parameter will determine how much he gets in the allocation.

Power is often present in economic models but usually implicit. For example, in current political economy models (e.g the probabilistic voting model) politicians target policies at different voters to win elections. Who they target depends on various types of characteristics. For example redistribution in democracies is targeted at groups,

- Which are relatively numerous.
- Which are able to solve the collective action problem when others are not (Bates, 1981, Persson and Tabellini, 2000, Section 3.5, Grossman and Helpman, 1996).
- Who manage to form political parties while others are not (Wittman, 1983) or who capture the platform of political parties (Acemoglu and Robinson, 2006).

- Who are un-ideological 'swing' or 'floating' voters (Lindbeck and Weibull, 1987, Dixit and Londregan, 1996).
- Who are relatively well informed (Strömberg, 2001, Besley and Burgess, 2002).
- Who turn out to vote in high numbers.
- Who are relatively poor (Dixit and Londregan, 1996).
- Who are in the same social network as politicians (Dixit and Londregan, 1996, Robinson and Verdier, 2002).

Depending on the specification of the model, possessing any of these characteristics makes an agent or group an attractive target for politicians to target and thus gives such agents 'power' to influence policy.

In Acemoglu and Robinson (2006) the emphasis is on two sources of power: de jure, which is allocated by the formal political institutions, and de facto. In the analysis dictatorship and democracy allocate power in different ways and the main consequence is that different agents are empowered to determine the tax rate under different political institutions. The framework in our book also emphasizes de facto power however. There we conceptualize this in terms of collective action, but take as exogenous the circumstances that lead a group to be able to solve its' collective action problem.

But what is de facto power and where does it come from? More generally, what other sources of power are there? You might imagine that the most basic source of power is physical force or the control over weapons. Ideas like this were first formalized by the use of a 'power function'.

16.1. The Power Function

16.1.1. Basics. A central approach to the study of power conceptualizes the balance of power in terms of a production function, the 'power function' (we follow the recent very useful survey by Garfinkle and Skaperdas, 20006). One can think of this as a technology of conflict. Imagine that two agents, 1 and 2, are fighting, the probability that 1 wins is $p(G_1, G_2)$ while the probability that 2 wins is $1-p(G_1, G_2)$. G_1 and G_2 are the inputs which might be effort, income, guns, or maybe people. Here $p : \mathbb{R}_+^2 \to [0, 1]$ and differentiable with some intuitive properties, i.e.

$$\frac{\partial p(G_1, G_2)}{\partial G_1} > 0, \frac{\partial p(G_1, G_2)}{\partial G_2} < 0$$

and you might think that homogeneity of degree zero was also plausible. The power function as typically used is a particular parametrization of this

$$p(G_1, G_2) = \frac{f(G_1)}{f(G_1) + f(G_2)} \tag{16.1}$$

when there are n agents fighting for a prize we have the probability that i wins the prize

$$p_i(G_i, G_{-i}) = \frac{f(G_i)}{\sum_{j=1}^{n} f(G_j)}$$

The most common form of (16.1) assumes that

$$p(G_1, G_2) = \frac{G_1^m}{G_1^m + G_2^m} \tag{16.2}$$

where $m > 0$. You should note here that this specification builds in that not fighting is never going to be an equilibrium. To see this note that for player 1 the marginal product of guns is

$$\frac{\partial p(G_1, G_2)}{\partial G_1} = \frac{mG_1^{m-1}G_2^m}{(G_1^m + G_2^m)^2}$$

Taking G_2^m as given you can see that as $G_1 \to 0$, $\frac{\partial p(G_1,G_2)}{\partial G_1} \to \infty$. Hence the marginal benefit of investing in guns gets very large as the amount of guns you have goes to zero.

A simple application of this model is as follows. There are two agents $i = 1, 2$ who have an endowment R_i of some resource. The resource can be allocated to producing guns, G_i or butter X_i. Assume

$$R_i = G_i + X_i/\beta_i$$

where β_i measures the marginal rate of transformation between guns and butter. We'll interpret this shortly in terms of the different agents having different comparative advantages.

Utility is linear and increasing in butter. The two agents fight and whoever wins gets all of the butter produced. The probability that an agent wins is increasing in the amount of guns he has and decreasing in the amount of guns the other agent has. The expected payoff of agent i is

$$V_i = p_i(G_1, G_2)\left(\beta_1 (R_1 - G_1) + \beta_2 (R_2 - G_2)\right)$$

with first-order conditions

$$\frac{\partial p_1(G_1, G_2)}{\partial G_1}\left(\beta_1 (R_1 - G_1) + \beta_2 (R_2 - G_2)\right) - p_1(G_1, G_2)\beta_1 = 0, \tag{16.3}$$

and

$$\frac{\partial p_2(G_1, G_2)}{\partial G_2}\left(\beta_1 (R_1 - G_1) + \beta_2 (R_2 - G_2)\right) - p_2(G_1, G_2)\beta_2 = 0. \tag{16.4}$$

These FOCs have intuitive interpretations. Consider (16.3). The first term is the marginal benefit of investing in guns, the change in the probability of winning the prize (butter) times the total quantity of butter (the prize). The second term is the marginal cost, the amount of output sacrificed by investing a little bit more in guns, β_1, times the initial probability that agent 1 wins the prize.

Consider first the symmetric case where $\beta_1 = \beta_2 = 1$ and $R_1 = R_2 = R$. Using (16.2) there is a unique Nash Equilibrium which we can compute by simultaneously solving (16.3) and (16.4) with

$$G_i = G^* = \frac{m}{m+1} R \text{ and } X_i = X^* = \frac{1}{m+1} R$$

this allocation gives the expected utilities $V_i = \frac{1}{m+1} R$.

Now consider $\beta_1 = \beta_2 = 1$ and $R_1 \neq R_2$ and let $\bar{R} = \frac{1}{2}[R_1 + R_2]$. Then the unique Nash Equilibrium is characterized by

$$G_i = G^* = \frac{m}{m+1} \bar{R} \text{ and } V_i = \frac{1}{m+1} \bar{R}.$$

Here even though endowments differ, the unqiue interior Nash Equilibrium is symmetric in the sense that both agents create the same amount of guns. The agent with the larger endowment then produces more butter. What is interesting about this equilibrium, first noticed by Hirsleifer, is that the expected utility of the players is independent of the distribution of the endowment. Thus the poorer agent gets the same expected utility as the rich agent. In a sense, the introduction of conflict levels any initial disparity in endowments. Indeed, the poorer agent allocates a larger proportion of his endowment to creating guns.

Extending these insights further, consider now the case where $\beta_1 \neq \beta_2$ and $R_1 \neq R_2$. Now the equilibrium implies that the Nash equilibrium choices (G_1^*, G_2^*) satisfy the equation

$$\frac{G_2^*}{G_1^*} = \left(\frac{\beta_1}{\beta_2} \right)^{\frac{1}{m+1}}$$

so that $G_1^* > G_2^*$ if and only if $\beta_1 < \beta_2$. If $\beta_1 < \beta_2$ then $p_1(G_1^*, G_2^*) > p_2(G_1^*, G_2^*)$. Moreover, one can show that the expected utility of player 1 is actually greater than that of player 2 in this model.

In this model we have the *paradox of power*, namely the agent with the comparative disadvantage in production invests more in guns and actually has a higher expected utility than the agent who has the greater economic productivity.

It is again true in this model that the distribution of endowments plays no role in determining the relative power of the players. Nevertheless, as you might conjecture, this result turns on assumptions about functional forms. For instance, with strictly concave utility an agent's probability of winning is strictly increasing in the size of his endowment relative to his opponent.

16.2. Economic Power

Apart from brute force and guns, it is often thought that there are economic sources of power. Simply the fact that I own an asset which you need to use to earn a living gives me power over you because I can get you to do things by denying you access to this asset. An old Marxist idea, for instance, is that there is a 'structural dependence of the state on capitalism.' What they meant by this is that a democratic state in a capitalist society is intrinsically limited in what it can do because of the economic power of capitalists. For instance capitalists can refuse to invest, in which case the economy collapses and tax revenues fall denying the state of resources. Another important source of power is undoubtedly social networks. These many be intimately related to the collective action problem, but this is an important under-research area (see the papers by Musacchio and Khwaja and Mian on the reafing list). Let's look at one specific example of how the control of assets may give power drawn from Baland and Robinson (2005).

16.2.1. The Baland-Robinson Model.

> "It is the most cruel mockery to tell a man he may vote for A or B, when you know that he is so much under the influence of A, or the friends of A, that his voting for B would be attended with the destruction of him. It is not he who has the vote, really and substantially, but his landlord, for it is for his benefit and interest that it is exercised in the present system."
> David Ricardo ([1824], 1951-1973, p. 506).

16.2.2. Introduction. Study the implications of the absence of a secret ballot for voting behavior and factor market equilibrium. When voting is not secret, it becomes feasible to buy and sell votes. While there are recorded instances of an individualized market for votes, the main stylized fact which emerges from the case study literature is that rather than individuals freely selling their votes to politicians, others, usually employers, control and supply the votes of their employees in exchange for money, favors or policies. More specifically, as discussed by Ricardo (1824), employers are usually landlords.

We argue that employment and political control can be connected. Many employment relationships concede rents to workers. For example, when worker effort is crucial for production, but only imperfectly observed. We show that the fact that landlords already concede rents to workers allows them to use the threat of withdrawing these rents to control their voting behavior. We thus demonstrate that employment does not simply generate income, it also gives *power* to control the behavior of others. This control makes it cheaper for political parties to secure the workers' votes indirectly, by purchasing them from the landlords, than to purchase them directly from the individual voters.

In Britain, before the introduction of the secret ballot, this factor was critical in determining the outcome of rural elections. As observed by Lord Stanley in 1841,

> "when any man attempted to estimate the probable result of a county election in England, it was ascertained by calculating the number of the great landed proprietors in the county and weighing the number of occupiers under them."

In Chile the control of voting by landowners was very frankly discussed in the debate leading up to the introduction of the secret ballot in 1958 in language strikingly similar to that used by Lord Stanley quoted above. For example, Socialist senator Martones argued in favor of introducing the secret ballot because,

> "if that law [the old electoral law without a secret ballot] did not exist, instead of there being 9 Socialist senators there would be 18, and you [the Conservatives] would be reduced to 2 or 3 ... [laughter] you laugh, but the truth is that there would be not 2 Conservative senators from O'Higgins and Colchagua, which corresponds exactly to the number of *inquilinos* in the fundos which belong to the Conservative hacendados in that region. Conservatives would have only one or perhaps none."

16.2.3. The Model. We consider a discrete time infinite horizon model of the rural sector. There is a unit mass of agents and a proportion x of rural agents have access to the capital market and can therefore purchase land and hire workers. All rural agents have the option to be self-employed and earn an income of $\underline{w}$. We let m denote the proportion of rural agents who become agricultural workers, and $1-m-x$ who remain self-employed.

There are L units of land which are owned by landowners with each owning $L/x = l$ units of land. There is a single numeraire consumption good which is produced from land and labor. The technology is characterized by a standard constant returns to scale neoclassical production function. On any farm, the expected output of a worker in any period is equal to $\tilde{\theta} g\left(\frac{l}{n}\right)$ where n is employment, g is the intensive form of the production function so that $g' > 0$ and $g'' < 0$, and $\tilde{\theta}$ is a plot-specific stochastic shock to output which is distributed independently across plots and time and can take two values, θ and 0 (by normalization) (since we focus on stationary equilibria we do not introduce time subscripts). The probability that θ occurs in period t depends on the effort exerted by a worker in that period. Effort, ε, takes two values, $\varepsilon \in \{0, e\}$. If $\varepsilon = e$, θ occurs with probability γ^h, while if $\varepsilon = 0$, θ occurs with probability $\gamma^l < \gamma^h$. Expected output on a farm of size l with n workers is therefore $\tilde{\theta} g\left(\frac{l}{n}\right) n$.

While output is perfectly observable by the landlord, the level of effort exerted by the worker is not. This induces a moral hazard problem. We assume that effort can never be observed so that the only possible wage contract depends on the realization of $\tilde{\theta}$.

There are also two political parties, 'Left' (denoted L) and 'Right' (denoted R) competing for votes to win an election and all individuals have exogenous preferences for one of these parties which means that they get utility from voting for the party they prefer (as in a standard probabilistic voting model).

All agents in the rural sector have utility functions which are linear in consumption, c, effort, ε, and voting decision σ^j for $j = L, R$ which depends on the ideological orientation of the agent. Thus, $U(c, \varepsilon, \sigma^j) = c - \varepsilon + \sigma^j$ is the utility of an agent of type j if they vote for the party they prefer, otherwise it is $U(c, \varepsilon, \sigma^j) = c - \varepsilon$. All agents maximize the expected present discounted value of utility and discount the future at rate $\beta \in (0, 1)$.

The political parties have per-period utility functions,

$$U_j = \varphi_j W^j - M_j, \; j = R, L$$

where φ_j stands for the probability that party j wins the election, and W^j, the gain in utility for party j if it wins the election. M_j represents the amount of rents (income) transferred by party j to other agents in the economy so that neither party is liquidity constrained. The price that a party offers for the vote of an agent will in general depend on the occupation of the agent: let p^j_ℓ be the price paid by party j to a landlord, p^j_w be the price paid by party j for the vote of a worker, and p^j_s be the price paid for the vote of a self-employed agent. Let μ be the impact of one vote in party j's favor on party j's chances of winning the election. From this we can deduce that the maximal price that party j would be prepared to pay for a vote is μW^j.

16.2.4. Timing of the stage game. The stage game has the following timing:

- The political parties non-cooperatively announce a price at which they will purchase votes from each type of rural agent.
- The land market opens with each landlord deciding how much land to buy.
- Landowners hire workers by proposing a contract.
- Agents sell votes to the political parties.
- Workers vote and choose their effort level.
- Production take place and the output shock $\tilde{\theta}$ is observed.

- Landlords and the political parties observe voting behavior and the state of nature.
- Rents are distributed by the political parties, wages are paid and workers may be fired and consumption takes place.

We now characterize the stationary subgame perfect equilibrium of this game. We begin with the competition for vote buying between the political parties.

16.3. Electoral Corruption and Resource Allocation

For ease of exposition we first characterize the outcome of political competition for votes. We do so by assuming, as will be the interesting case, that landlords control and sell the votes of their workers. In the next section we analyze the circumstances under which this will happen in equilibrium. To keep the discussion focused we assume that all landowners are right-wing while all other agents are left-wing. In addition we assume that the right-wing party values winning more than the left-wing party. This will have the implication that the right-wing party will be prepared to pay more for votes than the left-wing party.

The political parties engage in Bertrand competition. We first consider the situation in which the right-wing party will always wish to outbid the left-wing party for votes. This implies that $\mu W^R \geq \mu W^L + \sigma^L$ and the following prices are offered by the parties in equilibrium,

$$\text{Party } R \text{ offers } \begin{cases} p_\ell^R = \mu W^L - \sigma^R \\ p_w^R = \mu W^L \\ p_s^R = \mu W^L + \sigma^L \end{cases}$$

and,

$$\text{Party } L \text{ offers } \; p_\ell^L = p_w^L = p_s^L = \mu W^L$$

In this case, for any price that the left-wing party proposes for votes, the right-wing party is always willing to outbid that offer for the three categories of rural agents. As a result, in equilibrium, the left-wing party announces the maximal price it is ready to pay for one vote, μW^L. Given this price, landlords will be willing to sell their own votes to the right-wing party provided they can achieve the same utility level that they could by selling their votes to the left-wing party. This implies that the right-wing party must offer them a price at least equal to $p_\ell^R = \mu W^L - \sigma^R$. Landlords will also sell the votes of their workers if they are given the same price that is offered by the left-wing party, which is then the price the right-wing party announces. Lastly, for the self-employed agents, the right-wing party must compensate them for not voting for their own preferred party, which implies that he has to pay a price $p_s^R = \mu W^L + \sigma^L$ to those agents.

Given these prices, all rural agents sell their votes to the right-wing party, with right-wing landlords stipulating that their left-wing workers vote right-wing in their employment voting contracts.

In the case where $\mu W^L + \sigma^L \geq \mu W^R > \mu W^L$, then

$$\text{Party } R \text{ offers } \begin{cases} p_\ell^R = \mu W^L - \sigma^R, \\ p_w^R = \mu W^L, \\ p_s^R = \mu W^R, \end{cases}$$

and,

$$\text{Party } L \text{ offers } \begin{cases} p_\ell^L = \mu W^L, \\ p_w^L = \mu W^L, \\ p_s^L = \mu W^R - \sigma^L. \end{cases}$$

In this case, it is no longer optimal for the right-wing party to outbid the left-wing party for the votes of the self-employed agents. Now, rather than buying the votes of all rural agents, the right-wing party buys the votes of the landlords and their workers, but the self-employed sell their votes to the left-wing party. Here, moving from being self-employed to becoming a worker leads to a switch in voting behavior.

Under either scenario we have the following result.

Proposition 16.1. *It is cheaper for the right-wing party to buy votes from a landlord than to buy votes directly from the self-employed.*

This result follows immediately from the fact that in equilibrium $p_w^R < p_s^R$. This proposition has the implication that it will never be profitable for a rural agent to become a political entrepreneur, buying votes from individuals and then selling them to parties. For the rest of the paper we focus on the situation where $\mu W^R \geq \mu W^L + \sigma^L$, the analysis of the other parts of the parameter space follow directly.

16.4. Employment and Power

We solve for the stationary subgame perfect equilibrium of this game which is best from the point of view of landlords. In general a strategy for a landlord is a contract offer at date t which specifies wages as a function of $\tilde{\theta}$ and a voting decision as a function of the history of play up to t. For a worker a strategy determines an effort and voting decision as a function of the history and the contract offered at t.

We start by describing the optimal labor-voting contract. As is standard, we endow the landlord with all the bargaining power with respect to workers and he can therefore make take-it-or-leave-it contract offers to his worker(s) specifying his expected voting behavior and effort level. As there are two dimensions to the worker's behavior, there are four possible wages, corresponding to whether output is high or low, and whether the worker is observed voting for the specified party or not. We assume that liability is limited so that wages must be non-negative. To ensure maximal incentives, a landlord will optimally propose a wage, w, and continued employment if output is high and the worker is not observed voting for the wrong party. If output is low or the worker votes for the left-wing party, the landlord will pay zero and fire the worker. We assume that if a worker is ever fired by a landlord he is never employed again by a landlord and is thus perpetually self-employed.

We focus here on the situation under which a worker is required by his landlord to vote for the right-wing party. Given his voting behavior, the worker will exert the optimal amount of effort if the following incentive compatibility condition is satisfied. Let $V_w(\varepsilon = e)$ be the value to the worker if he exerts effort, while $V_w(\varepsilon = 0)$ is the value if the worker shirks. The worker will exert effort if

$$V_w(\varepsilon = e) \geq V_w(\varepsilon = 0)$$

First consider the value from exerting effort which is,

$$V_w(\varepsilon = e) = \gamma^h \left(w + \beta V_w(\varepsilon = e)\right) + (1 - \gamma^h) \left(\frac{\beta \left(\underline{w} + \mu W^L + \sigma^L\right)}{1 - \beta} \right) - e. \tag{16.5}$$

Here, with probability γ^h the realization of output is high, in which case at date t the worker receives the wage w and is not fired. In consequence the worker gets the continuation value $\beta V_w(\varepsilon = e)$. With probability $1 - \gamma^h$, even though the worker exerted effort, output is low. In this case the worker gets no wage and is fired at date t, never to be re-employed. In this case from date $t + 1$ on, the worker is self-employed getting an income of $\underline{w}$ in each period and also being able to freely sell his vote to whichever party he wishes. The utility from this latter action is $\max\{p_s^R, \sigma^L + p_s^L\}$, i.e. the self-employed agent can sell his vote to the right wing party and sacrifice the utility benefit of voting for his preferred party, or he can sell his vote to the left and get the utility benefit σ^L. Now in equilibrium we showed that $p_s^R \equiv \sigma^L + p_s^L \equiv \mu W^L + \sigma^L$ and this explains the formula in (16.5). We now consider the value of shirking, which is

$$V_w(\varepsilon = 0) = \gamma^l \left(w + \beta V_w(\varepsilon = 0)\right) + (1 - \gamma^l) \left(\frac{\beta \left(\underline{w} + \mu W^L + \sigma^L\right)}{1 - \beta} \right) \tag{16.6}$$

The interpretation of (16.6) follows immediately from the discussion of (16.5), noting that now, since the worker is shirking, he does not incurr any effort cost and high output arises with probability γ^l. Hence, solving for the value functions exerting effort is optimal if

$$w \geq \beta \left(\underline{w} + \mu W^L + \sigma^L\right) + \frac{(1 - \beta\gamma^l)}{\gamma^h - \gamma^l} e \tag{16.7}$$

Next there is the participation constraint which shows that the worker prefers accepting a contract to his outside option. This implies

$$V_w(\varepsilon = e) \geq \frac{\underline{w} + \mu W^L + \sigma^L}{1-\beta} \text{ or } w \geq \frac{\underline{w} + \mu W^L + \sigma^L + e}{\gamma^h} \equiv V^*. \tag{16.8}$$

In the optimal contract, the landlord offers a wage w^* such that (16.7) holds as an equality. At this wage the participation constraint is

$$\frac{\gamma^l}{\gamma^h - \gamma^l} e \geq \underline{w} + \mu W^L + \sigma^L. \tag{16.9}$$

To show that a contract which stipulates both effort and voting behavior is optimal note that if the landlord only offered a contract which attempted to induce effort the efficiency wage would be

$$\tilde{w} = \beta \underline{w} + \frac{(1-\beta\gamma^l)}{\gamma^h - \gamma^l} e \tag{16.10}$$

and the participation constraint would be

$$\tilde{w} \geq \frac{\underline{w} + e}{\gamma^h} \equiv \tilde{V}. \tag{16.11}$$

Proposition 16.2. *If rents are initially sufficiently high and if it is optimal to induce high effort by paying such an efficiency wage which concedes such rents to workers, it is optimal for the landlord to also control political behavior.*

First note that if the wage had to be increased by the full disutility of voting behavior being controlled, namely $\mu W^L + \sigma^L$, then it could never be profitable for the landlord to offer a contract which controlled voting. This is because expected output would be the same but a vote can only be sold for μW^L.

There are two cases to consider. Note first that $w^* - \tilde{w} = \beta \left(\mu W^L + \sigma^L\right)$ which is the increase in the efficiency wage that has to be paid to deter cheating when the landlord decides to control voting behavior. If initial rents are so large that at the new wage (16.8) is slack, then the issue is whether the expected increase in the wage is less than the benefit from controlling a vote μW^L, i.e.

$$\mu W^L \geq \gamma^h \beta \left(\mu W^L + \sigma^L\right)$$

In the second case the participation constraint is violated when the efficiency wage is increased by $\beta \left(\mu W^L + \sigma^L\right)$. In this case the wage must be increased above w^* by an amount z where $w^* + z = V^*$. Can it ever be profitable for the landlord to control voting behavior in this case? The answer is yes if

$$\begin{aligned} \mu W^L &\geq w^* + z - \tilde{w} \\ &= V^* - \tilde{w} \end{aligned} \tag{16.12}$$

Now note that if $\tilde{w} = \tilde{V}$, so that initially there were no rents, then $V^* - \tilde{V} = \left(\mu W^L + \sigma^L\right)/\gamma^h$ so that (16.12) cannot possibly be satisfied. However, with rents, $\tilde{w} > \tilde{V}$, so $V^* - \tilde{w} < V^* - \tilde{V}$, and (16.12) can hold if the initial rents are sufficiently large. Let $\tilde{w} + R = \tilde{V}$ where $R > 0$ are initial rents, then (16.12) becomes

$$\mu W^L \geq V^* - \tilde{V} - R.$$

We now have to consider whether it is optimal for landlords to pay the efficiency wage w^*. To understand this we first consider the optimal demand for labor in a farm of size l with n workers. Profits are,

$$\gamma^h \theta g\left(\frac{l}{n}\right) n - \gamma^h w^* n + \mu W^L n \tag{16.13}$$

The first term in (16.13) is expected revenues, the second the expected wage bill, and the third the political rents that the landlord gets from selling the votes of his n workers at the price μW^L. The optimal demand for labor is determined by the first-order condition,

$$\gamma^h \theta \left(g\left(\frac{l}{n}\right) - g'\left(\frac{l}{n}\right)\frac{l}{n}\right) - \gamma^h w^* + \mu W^L = 0 \tag{16.14}$$

The equation (16.14) implicitly defines the optimal demand for labor as a function of parameters, which we write $n(l, \mu W^L, w^*)$.

It is always profitable for the landlord to pay this efficiency wage contract if

$$\gamma^h \left(\theta g\left(\frac{l}{n(l, \mu W^L, w^*)}\right) - w^*\right) + p_w^R \geq \gamma^l \theta g\left(\frac{l}{n(l, \mu W^L, w^*)}\right) - \underline{w}$$

Using the fact that $p_w^R = \mu W^L$ we can state this as an assumption,

$$\left(\gamma^h - \gamma^l\right)\theta g\left(\frac{l}{n(l, \mu W^L, w^*)}\right) + \mu W^L \geq \gamma^h w^* - \underline{w} \tag{16.15}$$

We therefore assume that the expected increase in profit from workers exerting effort, evaluated at the efficieny wage, plus the rents from selling their votes must be greater than the expected increase in the wage bill.

The model has interesting implications for the price of land, denoted π, which can be seen for considering how this is determined. In the model landlords hold land while workers have no access to capital markets and cannot purchase land. Nevertheless, landlords could buy land from each other. The equilibrium price of a plot of land must now adjust so that profits are zero or,

$$\left(\gamma^h \theta g\left(\frac{l}{n(l, \mu W^L, w^*)}\right) - \gamma^h w^* + \mu W^L\right)\frac{n(l, \mu W^L, w^*)}{l} = \pi \tag{16.16}$$

Equation (16.16) and (**??**) implies the following result.

PROPOSITION 16.3. *In equilibrium the price of land incorporates political rents.*

The main argument developed in the paper is that, thanks to the rents they concede to their workers, it is less costly for landowners to control the political behavior of workers than it is for the political parties. As a result, parties are ready to transfers rents to landowners in exchange for the votes they control. Acquiring land is thus desirable not only for productive purposes, but also for the political rents attached to the political control of the workforce employed on it. Equilibrium prices on the land market reflect this mechanism.

It follows from Proposition 8 that a political reform with removes electoral corruption, such as the introduction of an effective secret ballot, removes the ability of landlords to sell the votes of their workers and has interesting comparative static effects.

Proposition 16.4. *The introduction of a secret ballot leads to a fall in the price of land, reduces land concentration and the vote share of the right-wing party.*

If all agents had access to capital markets then there would be no land concentration and all land would be farmed by smallholders with no workers getting rents. The fact that, with perfect capital markets, smallholders are always willing to outbid landowners for land follows from the fact that, through the participation constraint, the economic rents that landlords transfer to workers exceed the political rents they receive from parties. Therefore, even though it is still true that the ability of landlords to sell votes increases their demand for land, land is still more valuable to smallholders.

16.5. References

(1) Acemoglu, Daron and James Robinson (2006) "Appendix" in Acemoglu and Robinson *Economic Origins of Dictatorship and Democracy*, Cambridge University Press.

(2) Acemoglu, Daron, Simon Johnson and James Robinson (2005) "The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth," *American Economic Review*, 95, 546-579.

(3) Baland, Jean-Marie and James A. Robinson (2005) "Land and Power: Theory and Evidence from Chile," Unpublished.

(4) Bates (1981) Markets and States in Tropical Africa, University of California Press, Berkeley CA.

(5) Besley and Burgess (2002) "The Political Economy of Government Responsiveness: Theory and Evidence from India," Quarterly Journal of Economics, 117, 1415 - 1452.

(6) Dixit and Londregan (1996) "The Determinants of Success of Special Interest in Redistributive Politics," Journal of Politics, LVIII, 1132-1155.

(7) Garfunkel, Michelle and Stergios Skaperdas (2006) "Economics on Conflict: An Overview," Unpublished.

(8) Grossman, Gene M., and Elhanan Helpman (1994) "Protection for Sale." *American Economic Review* 84: 833-850.

(9) Lindbeck, Assar and Jörgen Weibull (1987) "Balanced-Budget Redistribution as the Outcome of Political Competition," Public Choice, LII, 272-297.

(10) Musacchio, Aldo (2005) On reading list.

(11) Persson, Torsten and Guido Tabellini (2000) Political Economics: Explaining Economic Policy, Cambridge; MIT Press.

(12) Strömberg, David (2004) "Radio's Impact on Public Spending," Quarterly Journal of Economics, 119, 189-221.

(13) Ricardo, David (1984) See Baland-Robinson.

(14) Robinson, James A. and Thierry Verdier (2002) "The Political Economy of Clientelism," CEPR Discussion Paper #1313.

(15) Stergios Skaperdas (1992) "Cooperation, Conflict and Power in the Absence of Property Rights," *American Economic Review*, 82, 720-739.

(16) Wittman, Donald (1983) "Candidate Motivation: A Synthesis of Alternative Motivations," American Political Science Review, 77, 142-157.

CHAPTER 17

Modeling Non-Democratic Politics

We now begin our analysis of dynamics of institutions, focusing more explicitly on the commitment value of institutions and their function of allocating *de jure* political power. We will build up to our main model, which allows equilibrium changes between dictatorship and democracy by first considering a simple model of dynamic nondemocratic politics. It is worth emphasizing here that despite the fact that for most of history, and even the 20th century, democracy has not been very usual, models of dictatorial politics (especially satisfactory models of nondemocratic politics) are very rare (a notable exception is Wintrobe, 1998). There have also been only a few attempts to characterize how politics differs in dictatorships, though there is a large case study literature by political scientists. In consequence, the approach we will pursue here is very simple. This is an exciting area for research to which we will return towards the end of the course.

17.1. Basic Issues

Our approach to institutions emphasizes their consequences. People have induced preferences over institutions because they understand that different consequences flow from different sets of institutions. In order to study institutional change, we therefore need to understand political equilibria under different institutions. We have already seen a number of models of the working of democratic institutions, and a few examples of nondemocratic political equilibria. We will first present a simple model of nondemocratic politics under the threat of revolution or unrest from the excluded masses. This will be the basis of the later models of democracy-dictatorship transitions.

The underlying economic model is the very simple two-class/group structure. In particular, in this model there are two classes, the elite (the rich) with fixed income y^r and the poor citizens with income $y^p < y^r$. Total population is normalized to 1, a fraction $1 - \delta > 1/2$ of the agents are poor, with income y^p, and the remaining fraction δ are rich with income y^r. Mean income is denoted by $\bar{y}$. Let θ be the share of total income accruing to the rich:

$$y^p = \frac{(1-\theta)\bar{y}}{1-\delta} \text{ and } y^r = \frac{\theta\bar{y}}{\delta}. \tag{17.1}$$

We also have

$$\frac{(1-\theta)\bar{y}}{1-\delta} < \frac{\theta\bar{y}}{\delta} \text{ or } \theta > \delta,$$

to ensure that $y^p < \bar{y} < y^r$.

The cost of taxation is $C(\tau)\bar{y}$ when the tax rate is τ, where C is increasing and strictly convex. Then the government budget constraint is:

$$T = \tau\left((1-\delta)y^p + \delta y^r\right) - C(\tau)\bar{y} = \left(\tau - C(\tau)\right)\bar{y}, \tag{17.2}$$

where T is the lump-sum distribution.

Given this specification (especially the limited tax instruments), it is straightforward to verify that the most preferred tax rate of poor agents is given by

$$\left(\frac{\theta-\delta}{1-\delta}\right) = C'(\tau^p). \tag{17.3}$$

In contrast, rich agents prefer zero taxes. [This would not be the case, for example, if group-specific transfers were allowed. In that case, the rich agents would want to tax and redistribute to themselves].

17.2. A Simple Model of Non-Democratic Politics

Now let us in that this very simple structure in a dynamic model. Suppose that individual utility is defined over the discounted sum of post-tax incomes with discount factor $\beta \in (0,1)$, so for individual i at time $t=0$, it is

$$U^i = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \hat{y}_t^i, \tag{17.4}$$

where $\hat{y}_t^i$ denotes after-tax income.

Policy is determined by rich agents, but poor agents can undertake revolution. The rich will therefore choose policy subject to a revolution constraint.

If we restrict ourselves to sequences of events where revolution never takes place, then (17.4) can be written in a more informative way:

$$U^i = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\left(1-\tau_t\right)y^i + \left(\tau_t - C\left(\tau_t\right)\right)\bar{y}\right), \tag{17.5}$$

where the second equality substitutes for transfers from (17.2), taking into account that tax rates are potentially time-varying, hence indexed by t. However, (17.5) only applies when there is no revolution along the equilibrium path. More generally, we should have

$$U^i = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\left(1-\bar{\rho}_t\right)\left(\left(1-\tau_t\right)y^i + \left(\tau_t - C\left(\tau_t\right)\right)\bar{y}\right) + \bar{\rho}_t y_R^i\right],$$

where $\bar{\rho}_t = 1$ if there has been a revolution at any time before t, and $\bar{\rho}_t = 0$ otherwise, and y_R^i is the income of individual i after a revolution.

If a revolution is attempted, it always succeeds but a fraction μ_t of the productive capacity of the economy is destroyed forever in the process. Therefore, if there is a revolution at time t, each citizen receives a per period return of

$$\frac{(1-\mu^S)\bar{y}}{(1-\delta)}$$

in all future periods: total income in the economy is $(1-\mu^S)\bar{y}$ and is shared between $1-\delta$ agents. Here, after a revolution, μ^S is the value of μ_t at the date when the revolution took place (μ^H or μ^L). This implies that the state does not fluctuate once a revolution has taken place. This assumption is just to simplify the algebra and we could allow for the threat state to fluctuate without changing the results. μ changes between two values: $\mu^H = \mu$ and $\mu^L = 1$, with $\Pr(\mu_t = \mu) = q$ irrespective of whether $\mu_{t-1} = \mu^H$ or μ^L.

The fact that μ fluctuates will be crucial in modeling the limited ability of the elite to promise future redistribution. A change in μ corresponds to a change in the underlying environment, so the elite, who hold political power in nondemocracy, will optimize again. As a result, their promise to redistribute today may not materialize due to changes in circumstances tomorrow. A high value of μ means that a revolution is very costly, while a low value of q implies that the threat of revolution is rare, perhaps because the citizens are unorganized. Fluctuations in the threat of revolution will be the source of the commitment problems arising from political power.

- The timing of events within a period, say time t, can be summarized as follows.
- μ_t is revealed.
- The elite set the tax rate τ_t^N.
- The citizens decide whether or not to initiate a revolution, denoted by ρ_t with $\rho_t = 1$ corresponding to a revolution at time t. If there is a revolution, they obtain the remaining $1-\mu_t$ share of output in all future periods.

Let us start with the pure strategy Markov perfect equilibria of this game, in which strategies only depend on the current state of the world and not on the entire history of the game. Later we will also discuss non-Markovian equilibria.

For Markov perfect equilibria, the crucial concept is that of the "state" of the game or the system which is simply a complete specification of all payoff relevant information. Here, the state of the system consists of the current opportunity for revolution, represented by either μ^L or μ^H. Let $\sigma^r = \{\tau^N(\cdot)\}$ be the actions taken by the elite when the state is $\mu_t = \mu^H$ or

μ^L. This consists of a tax rate $\tau^N : \{\mu^L, \mu^H\} \to [0,1]$. Similarly, $\sigma^p = \{\rho(\cdot,\cdot)\}$ is the action of the citizens which consists of a decision to initiate a revolution, ρ ($\rho = 1$ representing a revolution) conditional on the current actions of the elite. Hence, as in the previous model $\rho : \{\mu^L, \mu^H\} \times [0,1] \to \{0,1\}$. Then, a Markov perfect equilibrium is a strategy combination, $\{\tilde{\sigma}^r, \tilde{\sigma}^p\}$ such that $\tilde{\sigma}^p$ and $\tilde{\sigma}^r$ are best-responses to each other for all μ. Notice that Markov perfect equilibria are a subset of subgame perfect equilibria, since they exclude any subgame perfect equilibria that feature non-Markovian strategies.

Let us start with the payoffs once there is a revolution. We define $V^p(R, \mu^S)$ as the return to poor citizens if there is a revolution starting in threat state $\mu^S \in \{\mu, 1\}$. Recall that only the value of μ^S at the time of the revolution matters, and after that, a fraction μ^S of the productive capacity of the economy is destroyed forever. This implies that the value of revolution starting in the state μ^S is

$$V^p(R, \mu^S) = \frac{(1-\mu^S)\bar{y}}{1-\delta} + \beta\frac{(1-\mu^S)\bar{y}}{1-\delta} + \beta^2\frac{(1-\mu^S)\bar{y}}{1-\delta} + ... \tag{17.6}$$

which compounds all the future returns, taking into account that the future is discounted with discount factor $\beta < 1$. We have that

$$V^p(R, \mu^S) = \frac{(1-\mu^S)\bar{y}}{(1-\delta)(1-\beta)}.$$

Also, because the rich elite lose everything, $V^r(R, \mu^S) = 0$. Next recall that we have also assumed $\mu^L = 1$, the citizens would never attempt a revolution when $\mu_t = \mu^L$. Therefore, the only relevant value is the one starting in the state $\mu^H = \mu$, which is:

$$V^p(R, \mu^H) = \frac{(1-\mu)\bar{y}}{(1-\delta)(1-\beta)}. \tag{17.7}$$

Let us next turn to the decision of the elite. First consider the state $\mu_t = \mu^L$, where there is no threat of revolution, and let us try to calculate the value to the elite and to the citizens in this state, denoted by $V^r(N, \mu^L)$ and $V^p(N, \mu^L)$. We maintain the superscripts H and L on the μ's in the value functions to facilitate the exposition. The concept of Markov perfect equilibrium implies that irrespective of promises made in the past, in this state, the elite will choose whatever policy is in their best interest at that point. Since there is no threat of revolution, this must be to set $\tau^N = \tau^r$, and engage in no redistribution. However, the state $\mu_t = \mu^L$ in nondemocracy is not permanent. Next period, we could switch to $\mu_t = \mu^H$, and in this case, the elite might have to engage in redistribution, or there might be a revolution.

Let us denote the values to the elite and to the citizens in the state $\mu_t = \mu^H$ by $V^r(N, \mu^H)$ and $V^p(N, \mu^H)$. This implies that the relevant Bellman equation determining the values

$V^r(N, \mu^L)$ and $V^p(N, \mu^L)$ can be written as:

$$\begin{aligned} V^r(N,\mu^L) &= y^r + \beta\left[qV^r(N,\mu^H) + (1-q)V^r(N,\mu^L)\right] \\ V^p(N,\mu^L) &= y^p + \beta\left[qV^p(N,\mu^H) + (1-q)V^p(N,\mu^L)\right]. \end{aligned} \tag{17.8}$$

The value functions in (17.8) say that the value to a member of the elite in a nondemocracy and in the state $\mu_t = \mu^L$ consist of two terms: (1) what happens today, the first term y^r; and (2) what is expected to happen tomorrow, or the continuation value, represented by the second term, $\beta\left[qV^r(N,\mu^H) + (1-q)V^r(N,\mu^L)\right]$. Today, given the decision $\tau^N = \tau^r$, there is no redistribution, and a member of the elite obtains y^r, which is the first term. The second term is multiplied by β, since it starts tomorrow, and therefore is discounted back to today by the discount factor β. Tomorrow, there is a new draw from the distribution of μ, and with probability $1-q$, the state μ^L recurs, so we have $\mu_{t+1} = \mu^L$. In this case, exactly the same reasoning as today implies that the value to an elite agent from that point onwards will be $V^r(N,\mu^L)$, hence this term is multiplied by $1-q$ and included as part of the future value. The value $V^r(N,\mu^L)$ recurs because the world looking forward into the infinite future from state $\mu_t = \mu^L$ looks identical to the world looking forward into the infinite future from state $\mu_{t+1} = \mu^L$. With the remaining probability, q, there is a change in the state, and we have $\mu_{t+1} = \mu^H$, and in this case, we will have a different value for a member of the elite tomorrow, denoted by $V^r(N,\mu^H)$.

The same argument also applies for citizens, and gives the corresponding expression for $V^p(N,\mu^L)$, again consisting of two terms, what they receive today, y^p, and what they will receive tomorrow, $\beta\left[qV^p(N,\mu^H) + (1-q)V^p(N,\mu^L)\right]$.

Naturally, (17.8) is not sufficient to characterize the equilibrium, since we do not know what happens in the state $\mu_t = \mu^H$, or in other words we do not know what is $V^r(N,\mu^H)$ and similarly what is $V^p(N,\mu^H)$. In this state, there may be an effective threat of revolution. So the first thing to check is whether the revolution constraint is binding. To do so, define $V^r(N)$ and $V^p(N)$ as the payoffs that would apply if the society remains in nondemocracy all the time (i.e., no revolution) and the elite never redistribute to the citizens, i.e., $\tau^N = \tau^r$. We clearly have:

$$V^r(N) = \frac{y^r}{1-\beta},$$

since the elite always receive the income y^r as there is no taxation, and this future income stream is discounted to the present at the discount factor β. Similarly:

$$V^p(N) = \frac{y^p}{1-\beta} \tag{17.9}$$

We say that the revolution constraint binds if the poor citizens prefer a revolution in the state $\mu_t = \mu^H$ rather than to live in nondemocracy without any redistribution, i.e., if

$$V^p(R, \mu^H) > V^p(N)$$

where $V^p(R, \mu^H)$ is given by (17.7). Using the definitions in (17.1), the revolution constraint is equivalent to

$$\theta > \mu. \tag{17.10}$$

In other words, inequality needs to be sufficiently high, i.e., θ sufficiently high, for the revolution constraint to bind. If inequality is not that high, so that we have $\theta \leq \mu$, there is no threat of revolution even in the state $\mu_t = \mu^H$, even with no redistribution ever. In this case, the elite will always set their unconstrained best tax rate, $\tau^N = \tau^r$, and we have no revolution along the equilibrium path.

The more interesting case is the one where the revolution constraint (17.10) binds. If in this case, the elite set $\tau^N = \tau^r$ in the threat state $\mu_t = \mu^H$, there will be a revolution. So the elite will make some concessions by setting a tax rate $\tau^N = \hat{\tau} > 0$. Let us denote the values to the elite and the citizens in the state $\mu_t = \mu^H$ when the elite set a tax rate $\hat{\tau}$ and are expected to do so in the future, and there is no revolution, by $V^r(N, \mu^H, \tau^N = \hat{\tau})$ and $V^p(N, \mu^H, \tau^N = \hat{\tau})$. At this tax rate, we have that an agent of type i has net income of $(1 - \hat{\tau})\, y^i$, plus he receives a lump sum transfer of $\hat{T}$. From the government budget constraint, this lump-sum transfer is $\hat{T} = (\hat{\tau} - C(\hat{\tau}))\, \bar{y}$, where $\hat{\tau}\bar{y}$ is total tax revenue, and $C(\hat{\tau})\, \bar{y}$ is the cost of taxation.

By the same argument as before, we have that the value functions $V^r(N, \mu^H, \tau^N = \hat{\tau})$ and $V^p(N, \mu^H, \tau^N = \hat{\tau})$ are given by:

$$\begin{aligned} & V^r(N, \mu^H, \tau^N = \hat{\tau}) \\ = \;& y^r + (\hat{\tau}(\bar{y} - y^r) - C(\hat{\tau})\bar{y}) + \beta \left[q V^r(N, \mu^H, \tau^N = \hat{\tau}) + (1-q) V^r(N, \mu^L) \right], \\ & V^p(N, \mu^H, \tau^N = \hat{\tau}) \\ = \;& y^p + (\hat{\tau}(\bar{y} - y^p) - C(\hat{\tau})\bar{y}) + \beta \left[q V^p(N, \mu^H, \tau^N = \hat{\tau}) + (1-q) V^p(N, \mu^L) \right]. \end{aligned} \tag{17.11}$$

For the purposes of illustration, let us focus on the value function for a member of the elite. The first term is now $y^r + (\hat{\tau}(\bar{y} - y^r) - C(\hat{\tau})\bar{y})$, which is his net income after taxation at the rate $\hat{\tau}$. The second term is again the continuation value, $\beta \left[q V^r(N, \mu^H, \tau^N = \hat{\tau}) + (1-q) V^r(N, \mu^L) \right]$. With probability q, the state μ^H arises again tomorrow, and in this case, the rich continue to set $\tau^N = \hat{\tau}$ and receive $V^r(N, \mu^H, \tau^N =$

$\hat{\tau}$). With probability $1-q$, the state switches to μ^L, and the corresponding value is $V^r(N,\mu^L,\tau^N=\hat{\tau})$. The whole term is multiplied by β to discount it to the present.

A similar argument underlies the expression for $V^p(N,\mu^H,\tau^N=\hat{\tau})$. A citizen receives a relatively high income today, because there is redistribution at the rate $\hat{\tau}$. But what is going to happen in the future is uncertain. If the state remains at μ^H, redistribution continues. However, there is no guarantee of this, and in fact the threat state could switch to μ^L where the threat of revolution disappears, and as we saw above, now irrespective of what they promise, the elite will stop redistributing, and set $\tau^N=\tau^r$. Therefore, the expression for $V^p(N,\mu^H,\tau^N=\hat{\tau})$ already incorporates the potential "non-credibility" of the promise of future redistribution made today. Today's redistribution arises because the citizens have de facto political power: they have a relatively effective revolution threat, and if the elite do not make some concessions in the form of redistribution, they can overthrow the system. This political power therefore gets them additional income. This redistribution might cease tomorrow, however, if what gives political power to the citizens, the revolution threat, disappears. This is the essence of the problem of commitment in this society.

Now returning to the analysis of the current game, we still need to determine the action of the citizens after the elite decide to redistribute at the tax rate $\hat{\tau}$ in the state μ^H. Clearly, they have a choice between no revolution, $\rho=0$, and revolution $\rho=1$. If they decide to undertake a revolution, then once the game reaches this point, the value functions for revolution, $V^r(R,\mu^H)$ and $V^p(R,\mu^H)$, will apply. Otherwise, we will have $V^r(N,\mu^H,\tau^N=\hat{\tau})$ and $V^p(N,\mu^H,\tau^N=\hat{\tau})$. Moreover, clearly, a citizen will choose ρ depending on whether $V^p(N,\mu^H,\tau^N=\hat{\tau})$ or $V^p(R,\mu^H)$ is greater. Hence, we can write:

$$\rho\begin{cases} =0 & \text{if } V^p(R,\mu^H)\leq V^p(N,\mu^H,\tau^N=\hat{\tau}) \\ =1 & \text{if } V^p(R,\mu^H)> V^p(N,\mu^H,\tau^N=\hat{\tau}) \end{cases} \tag{17.12}$$

Note that this decision calculus is the same for all citizens. In other words, a citizen will take part in a revolution if he gets a higher return with a revolution than with redistribution at the rate $\hat{\tau}$ today, which again can be thought of as a "semi-credible promise of redistribution by the elite"—there will be redistribution today at the tax rate $\hat{\tau}$, and there might be tomorrow if nature determines that there is an effective threat of revolution tomorrow. We shall proceed by assuming in (17.12) that if $V^p(R,\mu^H)=V^p(N,\mu^H,\tau^N=\hat{\tau})$ then $\rho=0$ so that indifference is broken by not undertaking a revolution.

With ρ given by (17.12), we also have that

$$\begin{aligned} V^r(N,\mu^H) &= \rho V^r(R,\mu^H)+(1-\rho)\,V^r(N,\mu^H,\tau^N=\hat{\tau}) \\ V^p(N,\mu^H) &= \max_{\rho\in\{0,1\}}\rho V^p(R,\mu^H)+(1-\rho)\,V^p(N,\mu^H,\tau^N=\hat{\tau}) \end{aligned} \tag{17.13}$$

As we know, the elite would like to prevent revolution if they can. The question is whether they will be able to do so. To determine the answer to this question, we need to see what is the maximum value that the elite can promise to the citizens. Clearly this will be when they set the tax most preferred by the citizens, τ^p, given by (17.3). Hence the relevant comparison is between $V^p(R,\mu^H)$ and $V^p(N,\mu^H,\tau^N=\tau^p)$. If $V^p(N,\mu^H,\tau^N=\tau^p)\geq V^p(R,\mu^H)$, then a revolution can be averted, but not otherwise.

Note that, as one would expect, the value function $V^p(N,\mu^H,\tau^N=\tau^p)$ crucially depends on q, the probability that the state will be μ^H in the future, since this is the extent to which redistribution will recur in the future (in some sense, how much future redistribution the rich can credibly promise). To derive an expression for $V^p(N,\mu^H,\tau^N=\tau^p)$ we substitute $V^p(N,\mu^H,\tau^N=\tau^p)=V^p(N,\mu^H)$ in (17.8) and note that (17.8) and (17.11) are two linear equations in two unknowns, the value functions $V^p(N,\mu^H,\tau^N=\tau^p)$ and $V^p(N,\mu^L)$. Solving these two equations we find

$$V^p(N,\mu^H,\tau^N=\tau^p)=\frac{y^p+(1-\beta(1-q))\left(\tau^p(\bar{y}-y^p)-C(\tau^p)\bar{y}\right)}{1-\beta}. \tag{17.14}$$

(17.14) has a straightforward interpretation. It says that $V^p(N,\mu^H,\tau^N=\tau^p)$ is equal to the present discounted value of y^p, the pre-tax income of a citizen, plus the expected present value of net redistribution. Net redistribution is $\tau^p(\bar{y}-y^p)-C(\tau^p)\bar{y}$, but this only occurs when the state is μ^H, something which happens a proportion q of the time. However, in (17.14), $(\tau^p(\bar{y}-y^p)-C(\tau^p)\bar{y})$ is multiplied by $(1-\beta(1-q))$ not by q. This reflects the fact that today we start in the state μ^H, and given that today is more important than the future because of discounting (i.e., because $\beta<1$), the state μ^L, where there will be no redistribution, gets the weight $\beta(1-q)$, not $(1-q)$, and as a result, the state μ^H, received the remaining weight, $1-\beta(1-q)$. (Expressed differently, because we start in the high state, the citizens receive transfers today, and a fraction q of the time in the future, so the net present discounted value of the transfer is multiplied by $1+\beta q/\left(1-\beta\right)=\left(1-\beta(1-q)\right)/\left(1-\beta\right)$). Notice also that as $\beta\to 1$, i.e., as discounting disappears, the weight of the state μ^H indeed converges to q.

Given this value function, we can see that the revolution can be averted if $V^p(N,\mu^H,\tau^N=\tau^p)\geq V^p(R,\mu^H)$, or if

$$\frac{y^p+(1-\beta(1-q))\left(\tau^p(\bar{y}-y^p)-C(\tau^p)\bar{y}\right)}{1-\beta}\geq\frac{(1-\mu)\bar{y}}{\left(1-\delta\right)\left(1-\beta\right)},$$

which can be simplified to

$$\mu\geq\theta-(1-\beta(1-q))\left(\tau^p(\theta-\delta)-(1-\delta)C(\tau^p)\right). \tag{17.15}$$

If this condition does not hold, even the maximum credible transfer to a citizen is not enough, and there will be a revolution along the equilibrium path. We can now use (17.15) to define a critical value of μ^H, again denoted μ^* such that $V^p(N, \mu^*, \tau^N = \tau^p) = V^p(R, \mu^*)$, or

$$\mu^* = \theta - (1 - \beta(1-q))\left(\tau^p(\theta - \delta) - (1-\delta)C(\tau^p)\right). \tag{17.16}$$

where $\mu^* < \theta$. Naturally, we have that when $\mu \geq \mu^*$, $V^p(N, \mu^H, \tau^N = \tau^p) \geq V^p(R, \mu^H)$, and the revolution is averted. Whereas when $\mu < \mu^*$, $V^p(N, \mu^H, \tau^N = \tau^p) < V^p(R, \mu^H)$, future transfers are expected to be sufficiently rare that even at the best possible tax rate for the citizens, there isn't enough redistribution in the future, and the citizens prefer a revolution rather than to live under nondemocracy with political power in the hands of the elite.

Using (17.7) and (17.14), we have that $\hat{\tau}$ is given by:

$$\mu = \theta - (1 - \beta(1-q))\left(\hat{\tau}(\theta - \delta) - (1-\delta)C(\hat{\tau})\right). \tag{17.17}$$

This analysis than leads to:

Proposition 17.1. *There in a unique Markov perfect equilibrium $\{\tilde{\sigma}^r, \tilde{\sigma}^p\}$ of the game $G^\infty(\beta)$. Let μ^* and $\hat{\tau}$ be given by (17.16) and (17.17). Then in this equilibrium:*

- *If $\theta \leq \mu$, the elite never redistribute and the citizens never undertake a revolution.*
- *If $\theta > \mu$, then we have that:*
- *If $\mu < \mu^*$, promises by the elite are insufficiently credible to avoid a revolution. In the low state, the elite do not redistribute and there is no revolution, but in the high state a revolution occurs whatever tax rate the elite set.*
- *If $\mu \geq \mu^*$, the elite do not redistribute in the low state and set the tax rate $\hat{\tau}$ in the high threat state, just sufficient to stop a revolution. The citizens never revolt.*

It is most interesting to focus on the cases where $\theta > \mu$. Starting with the elite in power, if $\mu < \mu^*$, then they set a zero tax rate when $\mu_t = \mu^L$, but when the state transits to μ^H they are swept away by a revolution. The problem here is that although the elite would like to stay in power by offering the citizens redistribution, they cannot offer today enough to make the present value of nondemocracy to the citizens as great as the present value of revolution. To avoid a revolution they would have to pay not just now, but also in the future. Unfortunately, however, they cannot credibly promise to pay enough in the future and as a result the citizens find it optimal to revolt. In contrast, when $\mu \geq \mu^*$, the elite can prevent a revolution by redistributing. So in the state $\mu_t = \mu^L$, they set $\tau^N = 0$, and when $\mu_t = \mu^H$, they set a tax rate, $\tau^N = \hat{\tau}$, just high enough to prevent the revolution.

This proposition therefore shows how in a dynamic setting the ability of the elite to transfer resources to the citizens, in other words, the "credibility" of their promises, depends on the future allocation of political power. When q is very low, the citizens may have de facto political power today because of an effective revolution threat, but are very unlikely to have it again in the future. In this case, any promises made by the elite are not credible, and the citizens prefer to use their political power to transform society towards one that's more beneficial for themselves, taking a bigger slice of the cake. It is only when q is high, so that the de facto political power of the citizens is likely to recur in the future, that the promises made by the elite are sufficiently credible that a revolution can be averted.

There is an interesting paradox here. When q is high, so that the de facto political power of the citizens is more permanent, it is easier to avoid a revolution. This follows from the fact that μ^* defined by (17.16) is decreasing in q. This is because when the power of the citizens is not transitory it is easier for the elite to make credible promises of redistribution in the future. This is somewhat counterintuitive because a simple intuition might have been that when the citizens were better organized and more powerful, a revolution would have been more of a threat. Here this is not the case because the future threat of revolution also enables more credible promises by the elite to stave off the revolution.

Notice also that the critical threshold μ^* depends on the extent of inequality in society. In particular, the more unequal is society, i.e., the higher is θ, the higher is μ^*, and the more likely are revolutions. The reason is simple: with greater inequality, revolution is more attractive, and a greater amount of credible redistribution is necessary to avert the revolution.

17.3. Incentive Compatible Promises

The analysis in the previous section focused on Markov perfect equilibria, and showed how a revolution may arise as an equilibrium outcome. Since the political power of the citizens in the future was limited, any promise made by the elite when they keep political power in their own hands is imperfectly credible, and the citizens may prefer to take power today by revolution. An important ingredient of this story was the commitment problem: the elite find it optimal to revert back to their most preferred tax rate as soon as the threat of revolution disappears. This was a consequence of our restricting attention to Markovian strategies, since we imposed that, once the threat of revolution subsides, the elite would always choose the strategy that is in their immediate interests.

It is possible, however, that the elite can make certain other promises, for example, they might promise to redistribute in the future even if this is not in their immediate interests, and they can support this by the implicit understanding that if they deviated from this promise, when the threat of the revolution recurs again, the citizens would undertake a revolution, giving the elite a very low payoff—in other words, these promises could be supported by the threat of future punishments, or by "repeated-game" strategies. Punishments here correspond to actions that the citizens will take in the future (i.e., revolution), once the elite deviate from their prescribed behavior (renege on their promises), that will hurt the elite. When we allow players to play non-Markovian strategies the result will be the survival of nondemocracy for a larger set of parameter values. The important difference between Markovian and non-Markovian strategies is that the latter allow players to condition their actions at date t not only on the state at that date, but also on the previous history of play until that date.

Let us now take a situation where, in ter,s of the previous Proposition, $\theta > \mu$ and $\mu < \mu^*$, so with the restriction to Markov perfect equilibria, the unique equilibrium involves a revolution. Let us see whether the elite can avert the revolution by using incentive-compatible promises supported by future punishments. To do this, we first find the maximum value that the elite can give to the citizens, once we take into account potential punishment strategies. Since in general repeated games have many subgame perfect equilibria, we will focus on the subgame perfect equilibrium that is best for the elite. This subgame perfect equilibrium will prevent revolution for the largest possible set of parameter values, but there are other subgame perfect equilibria, which also prevent revolution for the same set of parameter values but give the citizens more than this. Nevertheless, the analysis of the specific equilibrium here will give the flavor of what types of outcomes can be supported in non-Markovian equilibria.

Suppose also that we start when the state is μ^L. We first calculate the value to the elite if they redistribute at the rate $\tau^N = \tau^H \leq \tau^p$ in the state $\mu_t = \mu^H$ and at the rate $\tau^N = \tau^L \leq \tau^p$ in the state $\mu_t = \mu^L$ (since we are no longer looking at Markovian strategies, $\tau^L > 0$ is now possible). We also suppose for now that the citizens will not undertake a revolution (later we will impose this as a constraint on the tax vector). By the same arguments as above, this value is given by

(17.18)

$$\begin{aligned} & V^r(N, \mu^L, \left[\tau^L, \tau^H\right]) \\ = \quad & y^r + \left(\tau^L \left(\bar{y} - y^r\right) - C\left(\tau^L\right)\bar{y}\right) + \beta\left[qV^r(N, \mu^H, \left[\tau^L, \tau^H\right]) + (1-q)V^r(N, \mu^L, \left[\tau^L, \tau^H\right])\right]. \end{aligned}$$

Note that we are now using a different notation, $V^r(N, \mu^L, \left[\tau^L, \tau^H\right])$, rather than $V^r(N, \mu^L)$ as we did in the previous section. This is because while in the MPE, the elite always set $\tau^N = 0$ when $\mu_t = \mu^L$, this is no longer true. In particular, we are looking at situations in which the elite make credible promises of a tax rate of τ^L when $\mu_t = \mu^L$ and set a tax rate of τ^H when $\mu_t = \mu^H$. Our new notation captures this. The term μ^L refers to the fact that we are in state $\mu_t = \mu^L$, and $\left[\tau^L, \tau^H\right]$ is the vector of promised taxes starting with the tax rate in the state $\mu_t = \mu^L$.

The intuition for equation (17.18) is straightforward; the first term, $y^r + \left(\tau^L\left(\bar{y} - y^r\right) - C\left(\tau^L\right)\bar{y}\right)$, is again the current return to the elite, given that there is taxation at the rate τ^L, and the second term is the continuation value, taking into account the fact that taxation will change to τ^H if the state switches to μ^H. By the same token, we also have

$$\begin{aligned} & V^r(N, \mu^H, \left[\tau^L, \tau^H\right]) \\ = \;& y^r + \left(\tau^H\left(\bar{y} - y^r\right) - C\left(\tau^H\right)\bar{y}\right) + \beta\left[qV^r(N, \mu^H, \left[\tau^L, \tau^H\right]) + (1-q)V^r(N, \mu^L, \left[\tau^L, \tau^H\right])\right] \end{aligned}$$

as the value starting in the state μ^H. Combining these two expressions, we obtain

$$V^r(N, \mu^L, \left[\tau^L, \tau^H\right]) = \frac{y^r + (1-\beta q)\left(\tau^L\left(\bar{y} - y^r\right) - C\left(\tau^L\right)\bar{y}\right) + \beta q\left(\tau^H(\bar{y} - y^r) - C(\tau^H)\bar{y}\right)}{1-\beta} \tag{17.19}$$

as the value that the elite will receive if they stick to their 'promised' behavior summarized by the tax vector $\left[\tau^L, \tau^H\right]$. The key will be whether this behavior is "incentive compatible" for them, that is, whether they will wish to deviate from it now or in the future.

What happens if they deviate? Clearly, the answer depends on how the citizens react. We want to see whether we can make the promise by the elite to redistribute at the tax rate $\tau^L > 0$ in state μ^L credible. It is more likely to be credible, when deviation from it is less profitable, or when deviation from this prescribed behavior will be met by a severe punishment. The most severe punishment is that of a revolution by the citizens when the opportunity occurs again (it is never profitable for the citizens to undertake a revolution in the state $\mu_t = \mu^L$, since $\mu^L = 1$, so the threat to undertake such a revolution in the state $\mu_t = \mu^L$ will not be credible, and therefore never part of a subgame perfect equilibrium). Consequently, the best way to ensure that the elite do not deviate from their promises is to threaten them (credibly) with as severe a punishment as possible, that is, a revolution as soon as the state switches to $\mu_t = \mu^H$. So there will be a revolution the first time the state is $\mu_t = \mu^H$. What will happen until then? The elite are now deviating from their promised behavior, so in the meantime, they will adopt the best policy for themselves, so $\tau^N = \tau^r = 0$. Thus, what we have is a value $V_d^r(N, \mu^L)$ for the elite, where the subscript d denotes that they have deviated from their prescribed behavior, and this value is given by the following recursion:

$$V_d^r(N,\mu^L) = y^r + \beta\left[qV^r(R,\mu^H) + (1-q)\,V_d^r(N,\mu^L)\right],$$

where we know that $V^r(R,\mu^H)=0$. Using this fact, we have that

$$V_d^r(N,\mu^L) = \frac{y^r}{1-\beta\,(1-q)}. \tag{17.20}$$

This analysis immediately establishes that only redistribution at the rate τ^L in the state $\mu_t = \mu^L$ such that

$$V^r(N,\mu^L,\left[\tau^L,\tau^H\right]) \geq V_d^r(N,\mu^L). \tag{17.21}$$

is credible. If the inequality were reversed, the elite would prefer to deviate and give the citizens no redistribution in the state μ^L, and suffer the consequences, rather than tax themselves at the rate τ^L now (and at the rate τ^H when the state becomes high). Therefore, (17.21) is necessary for redistribution at the tax rate τ^L to be 'incentive compatible' for the elite and thus a credible promise to the citizens. The reader can also note that the tax rate $\tau^H \leq \tau^p$ in the state $\mu_t = \mu^H$ is automatically credible, because we are looking at the part of the parameter space where $\mu < \mu^*$, so any deviation by the elite from their promised actions in the high state can be immediately punished.

Now the subgame perfect equilibrium that is best for the elite, starting in the state μ^L, can be characterized as the solution to the following maximization problem:

$$\max_{\tau^L,\tau^H} V^r(N,\mu^L,\left[\tau^L,\tau^H\right]) \tag{17.22}$$

subject to (17.21) and

$$V^p(N,\mu^H,\left[\tau^L,\tau^H\right]) \geq V^p\left(R,\mu^H\right), \tag{17.23}$$

where $V^p(N,\mu^H,\left[\tau^L,\tau^H\right])$ is the value to the citizens starting in the state μ^H from the tax vector $\left[\tau^L,\tau^H\right]$, and $V^p\left(R,\mu^H\right)$, as usual, is the value to the citizens from the revolution in the state μ^H given by (17.7) in the previous section.

While the first constraint ensures that the elite do not wish to renege on their promises, the second constraint requires that the citizens do not wish to undertake a revolution in the high state.

The value $V^p(N,\mu^H,\left[\tau^L,\tau^H\right])$ is obtained analogously to the values for the elite. In particular, we have the following value functions for the citizens. In the low state:

$$\begin{aligned} & V^p(N,\mu^L,\left[\tau^L,\tau^H\right]) \\ = \;& y^p + \left(\tau^L\left(\bar{y}-y^p\right) - C\left(\tau^L\right)\bar{y}\right) + \beta\left[qV^p(N,\mu^H,\left[\tau^L,\tau^H\right]) + (1-q)V^p(N,\mu^L,\left[\tau^L,\tau^H\right])\right], \end{aligned}$$

and in the high state,

$$\begin{aligned}&V^p(N,\mu^H,\left[\tau^L,\tau^H\right])\\ =\;& y^p+\left(\tau^H\left(\bar{y}-y^p\right)-C\left(\tau^H\right)\bar{y}\right)+\beta\left[qV^p(N,\mu^H,\left[\tau^L,\tau^H\right])+(1-q)V^p(N,\mu^L,\left[\tau^L,\tau^H\right])\right].\end{aligned}$$

Combining the two expressions, we obtain:

$$\begin{aligned}&V^p(N,\mu^H,\left[\tau^L,\tau^H\right])\\ =\;&\frac{y^p+\beta\left(1-q\right)\left(\tau^L\left(\bar{y}-y^p\right)-C\left(\tau^L\right)\bar{y}\right)+\left(1-\beta\left(1-q\right)\right)\left(\tau^H(\bar{y}-y^p)-C(\tau^H)\bar{y}\right)}{1-\beta}.\end{aligned} \tag{17.24}$$

Before providing a full solution to this maximization problem, it is straightforward to characterize the minimum value of μ^H, such that a revolution can be averted. We denote this threshold by μ^{**} with an analogy with the the threshold μ^* in the previous section. Formally, this threshold corresponds to the minimum value of μ^H such that the constraint set of the above optimization problem is non-empty. When the constraint set is empty, this implies that there is no tax vector $\left[\tau^L,\tau^H\right]$ that is simultaneously credible and can convince the citizens not to undertake a revolution, so there has to be an equilibrium revolution in the state μ^H.

To calculate this threshold, note that the largest value that τ^H can take is τ^p. Intuitively, in the high state, the elite are willing to give the maximum redistribution to avoid a revolution. What about τ^L? Once $\tau^H=\tau^p$, τ^L is then given by setting the incentive compatibility constraint of the elite, (17.21), as equality. Therefore, the largest amount of redistribution that can credibly be promised is that which stems from levying the tax rate $\bar{\tau}'$ in the state $\mu_t=\mu^L$ such that: $V^r_d(N,\mu^L)=V^r(N,\mu^L,[\bar{\tau}',\tau^p])$, or

$$\frac{y^r+(1-\beta q)\left(\bar{\tau}'\left(\bar{y}-y^r\right)-C\left(\bar{\tau}'\right)\bar{y}\right)+\beta q\left(\tau^p(\bar{y}-y^r)-C(\tau^p)\bar{y}\right)}{1-\beta}=\frac{y^r}{1-\beta\left(1-q\right)}.$$

Substituting for the definition of y^r and simplifying terms we obtain the maximum credible tax rate $\bar{\tau}'$ as:

$$\bar{\tau}'\left(\theta-\delta\right)+\delta C\left(\bar{\tau}'\right)=\frac{\beta q}{(1-\beta q)}\left[\frac{\theta}{1-\beta\left(1-q\right)}-\left(\tau^p\left(\theta-\delta\right)+\delta C(\tau^p)\right)\right]. \tag{17.25}$$

This tax rate, $\bar{\tau}'$, can be shown to be an increasing function of β; the more valuable is the future, the less attractive it is for the elite to deviate from the promised behavior, so the higher will be the maximum tax rate they can promise. This is intuitive, and in fact, a fundamental principle of analyses of repeated games; for a player not to take the action that is in their immediate interest, the benefits from this action need to be counterbalanced by some other, future, considerations. Here, if they take these actions, they will be punished in the future. The more a player discounts the future or the less severe is the expected punishment, the harder it will be to convince him to stick to these promises.

The important point highlighted by (17.25) is that the elite do not have unrestricted powers to make promises: they have a limited capability, supported by the threat of future punishments. Any promises they make will be credible only if it is in their interests to carry out this promise at the time. Here, some positive redistribution even without the threat of revolution might be in their interests because otherwise they know that they will have to put up with a revolution later down the line. Nevertheless, this threat of future punishments can support only a limited amount of redistribution (the elite cannot credibly promise a tax rate greater than $\bar{\tau}'$ in the low state).

This analysis then implies that the question of whether a revolution can be averted boils down to whether the value to the citizens from redistribution at the tax rate $\bar{\tau}'$ in the state $\mu_t = \mu^L$ and at the tax rate τ^p in the state $\mu_t = \mu^H$, starting in the state $\mu_t = \mu^H$, is better than a revolution for the citizens. Or put differently, this is equivalent to whether the tax vector $[\bar{\tau}', \tau^p]$ is in the constraint set of the above maximization problem given by inequalities (17.21) and (17.23).

By analogy to the analysis in the previous section, we can see that the tax vector $[\bar{\tau}', \tau^p]$ is in the constraint set for all $\mu \geq \mu^{**}$, where μ^{**} is such that $V^p(N, \mu^H, [\bar{\tau}', \tau^p]) = V^p(R, \mu^H)$ when $\mu^H = \mu^{**}$. More explicitly, we have the threshold μ^{**} is the solution to:

$$\begin{aligned}\mu^{**} = {} & \theta - \beta\,(1-q)\left(\bar{\tau}'\,(\theta-\delta) - (1-\delta)C\left(\bar{\tau}'\right)\right) \\ & - \left(1-\beta\,(1-q)\right)\left(\tau^p(\theta-\delta) - (1-\delta)C(\tau^p)\right),\end{aligned} \tag{17.26}$$

where $\bar{\tau}'$ is given by (17.25).

Recall that, using the notation in this section, μ^* is defined by $V^p(N, \mu^H, [0, \tau^p]) = V^p(R, \mu^H)$, so for all $\bar{\tau}' > 0$, we have

$$\mu^{**} < \mu^*,$$

which is clear from the formulas (17.16) and (17.26).

This implies that once we allow for the use of punishment strategies, there will be situations in which a revolution can be averted by incentive compatible promises, but could not have been averted otherwise. This will be true when $\mu \in [\mu^{**}, \mu^*)$. Nevertheless, as long as $\mu^{**} > 0$, there will still be situations, i.e. when $\mu < \mu^{**}$, in which the best that the elite can promise is not enough to avert a revolution. This again underlines the limited capability of the elite to make credible promises: only promises that will eventually be in their interest to carry out are credible.

This discussion leads to the main result of this section, which we informally state as:

> **Result:** When we allow non-Markovian strategies, a revolution can be averted for all $\mu \geq \mu^{**}$. Here $\mu^{**} < \mu^{*}$, which means that greater redistribution is now possible, but $\mu^{**} > 0$, which means that there are limits how much credible redistribution the elite can promise.

Now to state the results of this section more carefully and to complete the characterization of the equilibrium, we must define what a strategy is in this game. The main difference with the previous section is that we have dropped the restriction to Markov strategies and now a strategy can depend not just on the state at any date t but also on the history of play up to that date. Let $\mathcal{H}^{t-1}$ denote the set of all possible histories of play up to $t-1$ with a particular history being denoted $h^{t-1} \in \mathcal{H}^{t-1}$. The actions of the elite and citizens are now denoted by $\sigma^r = \{\tau^N(\cdot,\cdot)\}$ and $\sigma^p = \{\rho(\cdot,\cdot,\cdot)\}$ where $\tau^N(\mu_t, h^{t-1})$ is the tax rate set by the elite at date t when the current state is $\mu_t = \mu^H$ or μ^L and the observed history is h^{t-1}. Hence, $\tau^N : \{\mu^L, \mu^H\} \times \mathcal{H}^{t-1} \to [0,1]$. Similarly, $\rho(\mu_t, \tau^N, h^{t-1})$ is the decision by the citizens to initiate a revolution conditional on the the current state, the current actions of the elite, and the history. We have that $\rho : \{\mu^L, \mu^H\} \times [0,1] \times \mathcal{H}^{t-1} \to \{0,1\}$. Then, a subgame perfect equilibrium is a strategy combination, $\{\hat{\sigma}^r, \hat{\sigma}^p\}$ such that $\hat{\sigma}^r$ and $\hat{\sigma}^p$ are best-responses to each other for all possible histories $h^{t-1} \in \mathcal{H}^{t-1}$ and prior actions taken within the same stage game.

When $\mu < \mu^{**}$, the following strategy profile is the unique subgame perfect equilibrium: $\tau^N(\mu_t, h^{t-1}) = 0$ for $\mu_t \in \{\mu^L, \mu^H\}$ and any h^{t-1}, $\rho(\mu^L, \cdot, h^{t-1}) = 0$ and $\rho(\mu^H, \cdot, h^{t-1}) = 1$ for any h^{t-1}. For this set of parameter values, a revolution is sufficiently attractive that concessions will not work and the first time μ^H arises there will be a revolution whatever the previous history of play or the current tax rate. Since the elite know this they simply set zero taxes when μ^L occurs.

To understand the nature of the subgame perfect equilibrium when $\mu \geq \mu^{**}$, it is also useful to note that in this case there is an additional motive for the elite, "tax smoothing". Intuitively, the elite would like to deliver a given amount of redistribution to the citizens at the minimum cost to themselves. Since the cost of taxation given by the function $C(\cdot)$ is convex, this implies that taxes should exhibit as little variability as possible, in other words, they should be smooth. This idea was first suggested by Barro (1979) in the context of optimal fiscal policy, but it applies equally here. Such tax smoothing was not possible before, because the elite could never promise to redistribute in the state μ^L. Now that this type of redistribution is possible, tax smoothing also emerges as a possibility.

To see the tax smoothing more explicitly here, consider a pair of taxes, τ^L and $\tau^H > \tau^L$ that satisfy (17.23). Now imagine we construct a weighted average of these two taxes, $\tilde{\tau} = \left[\beta(1-q)\tau^L + (1-\beta(1-q))\tau^H\right]$. Inspection of (17.24) together with the (strict) convexity of $C(\cdot)$ immediately establishes that $V^p(N,\mu^H,[\tilde{\tau},\tilde{\tau}]) > V^p(N,\mu^H,\left[\tau^L,\tau^H\right])$ so the tax vector $[\tilde{\tau},\tilde{\tau}]$ also avoids a revolution. Moreover, again by the convexity of $C(\cdot)$, $V^r(N,\mu^L,[\tilde{\tau},\tilde{\tau}]) > V^r(N,\mu^L,\left[\tau^L,\tau^H\right])$, so the tax vector $[\tilde{\tau},\tilde{\tau}]$ also gives higher utility to the elite. This establishes that tax smoothing is preferable (if it is incentive compatible).

The tax smoothing argument makes it clear that the cheapest way of providing utility of $V^p(R,\mu^H)$, is to set a constant tax rate, τ^S, such that

$$V^p(N,\mu^H,\left[\tau^S,\tau^S\right]) = V^p(R,\mu^H), \tag{17.27}$$

or, more explicitly, τ^S is given by:

$$\begin{aligned}\mu = \theta &- \beta(1-q)\left(\tau^S(\theta-\delta) - (1-\delta)C\left(\tau^S\right)\right) \\ &- (1-\beta(1-q))\left(\tau^S(\theta-\delta) - (1-\delta)C(\tau^S)\right).\end{aligned} \tag{17.28}$$

Therefore, redistributing at this rate is the best possible strategy for the elite. The question is whether this tax vector is incentive compatible, i.e., whether it satisfies (17.21). The same arguments as above immediately imply that the vector $\left[\tau^S,\tau^S\right]$ will be incentive compatible as long as $\tau^S \leq \bar{\tau}^S$ where $\bar{\tau}^S$ is given by

$$\bar{\tau}^S(\theta-\delta) + \delta C\left(\bar{\tau}^S\right) = \frac{\beta q}{(1-\beta q)}\left[\frac{\theta}{1-\beta(1-q)} - \left(\bar{\tau}^S(\theta-\delta) + \delta C(\bar{\tau}^S)\right)\right], \tag{17.29}$$

which is similar to (17.25) above with the vector $\left[\bar{\tau}^S,\bar{\tau}^S\right]$ replacing $[\bar{\tau}',\tau^p]$.

Then the question of whether perfect tax smoothing can be achieved simply boils down to whether any tax rate $\tau^S \leq \bar{\tau}^S$ satisfies (17.27). Again similar arguments to before immediately establish that there exists a level of μ^H, here denoted $\bar{\mu}^S$, and given by

$$\begin{aligned}\bar{\mu}^S = \theta &- \beta(1-q)\left(\bar{\tau}^S(\theta-\delta) - (1-\delta)C\left(\bar{\tau}^S\right)\right) \\ &- (1-\beta(1-q))\left(\bar{\tau}^S(\theta-\delta) - (1-\delta)C(\bar{\tau}^S)\right),\end{aligned} \tag{17.30}$$

such that when $\mu \geq \bar{\mu}^S$, a perfectly smooth credible tax policy will prevent revolution.

Clearly $\bar{\mu}^S > \mu^{**}$ (on the other hand, $\bar{\mu}^S$ can be greater than or less than μ^*). When $\mu \geq \bar{\mu}^S$, the best possible subgame perfect equilibrium for the elite is a strategy combination that corresponds to the tax vector $\left[\tau^S,\tau^S\right]$ (which, by construction, prevents revolution at the lowest possible cost). More explicitly, let us define the history $\hat{h}^t$ such that $h^t = \hat{h}^t$ if for all $s \leq t$, $\tau^N(\mu^L,h^s) = \tau^S$ where τ^S is given by (17.27) above. Then, the subgame perfect equilibrium is given by the following strategy combination. For the elite:

$$\tau^N(\mu_t, h^{t-1}) = \begin{cases} \tau^S \text{ if } h^{t-1} = \hat{h}^{t-1} \\ 0 \text{ if } h^{t-1} \neq \hat{h}^{t-1} \end{cases} \tag{17.31}$$

for $\mu_t \in \{\mu^L, \mu^H\}$, and for the citizens: $\rho(\mu^L, \cdot, h^{t-1}) = 0$, and

$$\rho(\mu^H, \tau^N, h^{t-1}) = \begin{cases} 0 \text{ if } h^{t-1} = \hat{h}^{t-1} \text{ and } \tau^N \geq \tau^S \\ 1 \text{ if } h^{t-1} \neq \hat{h}^{t-1} \text{ or } \tau^N < \tau^S \end{cases}.$$

Note that in this case, as before, strategies specify how a player will play even off the equilibrium path, which now includes all possible histories up to that point. In particular, here $\hat{h}^{t-1}$ denotes the equilibrium path. Then as long as play is on this path, the elite set τ^S in both states, and the citizens never revolt. However, if the elite ever set a tax rate less than τ^S, then we will move along some history $h^{t-1} \neq \hat{h}^{t-1}$ and the strategies say that the first time the state is $\mu_t = \mu^H$ the citizens undertake a revolution. How do we know that in such a situation it will actually be credible for the citizens to undertake a revolution? This comes from (17.31), which states that if the elite find themselves setting the tax rate after some history different from $\hat{h}^{t-1}$, then they set the tax rate to zero. Thus the poor understand that if they do not undertake a revolution following a deviation from the prescribed behavior, they will never get any redistribution from that point on in the game. Therefore, as long as the revolution constraint $\theta > \mu$ holds, it will be optimal to undertake a revolution following a deviation by the elite.

Finally, when $\mu \in [\mu^{**}, \bar{\mu}^S)$, revolution can be averted, but perfect tax smoothing is no longer possible. In this case, it can be seen that the best subgame perfect equilibrium for the elite is a tax vector $\left[\hat{\tau}^L, \hat{\tau}^H\right]$ which is the solution to (17.22) and satisfies:

$$\hat{\tau}^L(\theta-\delta) + \delta C\left(\hat{\tau}^L\right) = \frac{\beta q}{(1-\beta q)}\left[\frac{\theta}{1-\beta(1-q)} - \left(\hat{\tau}^H(\theta-\delta) + \delta C(\hat{\tau}^H)\right)\right], \tag{17.32}$$

and

$$\begin{aligned} \mu = {} & \theta - \beta(1-q)\left(\hat{\tau}^L(\theta-\delta) - (1-\delta)C\left(\hat{\tau}^L\right)\right) \\ & - (1-\beta(1-q))\left(\hat{\tau}^H(\theta-\delta) - (1-\delta)C(\hat{\tau}^H)\right), \end{aligned} \tag{17.33}$$

and the corresponding subgame perfect strategies are:

$$\tau^N(\mu^L, h^{t-1}) = \begin{cases} \hat{\tau}^L \text{ if } h^{t-1} = \hat{h}^{t-1} \\ 0 \text{ if } h^{t-1} \neq \hat{h}^{t-1} \end{cases}, \tau^N(\mu^H, h^{t-1}) = \begin{cases} \hat{\tau}^H \text{ if } h^{t-1} = \hat{h}^{t-1} \\ 0 \text{ if } h^{t-1} \neq \hat{h}^{t-1} \end{cases},$$

$\rho(\mu^L, \cdot, h^{t-1}) = 0$, and

$$\rho(\mu^H, \tau^N, h^{t-1}) = \begin{cases} 0 \text{ if } h^{t-1} = \hat{h}^{t-1} \text{ and } \tau^N \geq \hat{\tau}^H \\ 1 \text{ if } h^{t-1} \neq \hat{h}^{t-1} \text{ or } \tau^N < \hat{\tau}^H \end{cases}.$$

Summarizing this discussion, we have

Proposition 17.2. *Assume $\theta > \mu$. Let μ^{**} and $\bar{\mu}^S > \mu^{**}$ be given by (17.26) and (17.30). Then, the subgame perfect equilibrium that is best from the viewpoint of the elite, $\{\hat{\sigma}^r, \hat{\sigma}^p\}$, of the game $G^\infty(\beta)$ is such that:*

(1) *if $\mu < \mu^{**}$, then $\tau^N(\mu_t, h^{t-1}) = 0$ for $\mu_t \in \{\mu^L, \mu^H\}$ and any h^{t-1}; and $\rho(\mu^L, \cdot, h^{t-1}) = 0$ and $\rho(\mu^H, \cdot, h^{t-1}) = 1$ for any $h^{t-1} \in \mathcal{H}^{t-1}$.*
(2) *if $\mu \geq \bar{\mu}^S$, $\tau^N(\mu_t, h^{t-1}) = \tau^S$ for $\mu_t \in \{\mu^L, \mu^H\}$ and $h^{t-1} = \hat{h}^{t-1}$, where τ^S is given by (17.28); $\tau^N(\mu_t, h^{t-1}) = 0$ for $\mu_t \in \{\mu^L, \mu^H\}$ and $h^{t-1} \neq \hat{h}^{t-1}$, $\rho(\mu^L, \tau^N, h^{t-1}) = 0$; $\rho(\mu^H, \tau^N, h^{t-1}) = 0$ for $h^{t-1} = \hat{h}^{t-1}$ and $\tau^N \geq \tau^S$; and $\rho(\mu^H, \tau^L, h^{t-1}) = 1$ for any $h^{t-1} \neq \hat{h}^{t-1}$ or $\tau^N < \tau^S$.*
(3) *if $\mu \in [\mu^{**}, \bar{\mu}^S)$, then $\tau^N(\mu^L, h^{t-1}) = \hat{\tau}^L$ and $\tau^L(\mu^H, h^{t-1}) = \hat{\tau}^H$ for $h^{t-1} = \hat{h}^{t-1}$ where $\hat{\tau}^L$ and $\hat{\tau}^H$ are given by (17.32) and (17.33); $\tau^N(\mu_t, h^{t-1}) = 0$ for $\mu_t \in \{\mu^L, \mu^H\}$ and $h^{t-1} \neq \hat{h}^{t-1}$; $\rho(\mu^L, \cdot, h^{t-1}) = 0$; $\rho(\mu^H, \tau^N, h^{t-1}) = 0$ if $h^{t-1} = \hat{h}^{t-1}$ and $\tau^N \geq \hat{\tau}^H$; and $\rho(\mu^H, \tau^N, h^{t-1}) = 1$ if $h^{t-1} \neq \hat{h}^{t-1}$ or $\tau^N < \hat{\tau}^H$.*

The important point that emerges from this Proposition is that there is now a larger set of parameter values that will allow the elite to avoid revolution. In other words, in societies with μ such that $\mu^{**} \leq \mu < \mu^*$, there will be equilibrium revolutions if we do not allow the elite to make incentive compatible promises of redistribution in future low-revolution threat periods, but these revolutions can be avoided once we allow such promises. Moreover, even when $\mu \geq \mu^*$, the elite can achieve a better outcome for themselves by smoothing taxes because of the possibility of using incentive compatible promises.

Nevertheless, it is important to emphasize that the elite still have limited abilities to make credible promises. Only promises of redistribution at the tax rate τ^L that satisfy $V^r(N, \mu^L, [\tau^L, \tau^H]) \geq V^r_d(N, \mu^L)$ are incentive compatible, and this implies that in societies with $\mu < \mu^{**}$ the same considerations as in the Proposition of the last section will apply and credible redistribution will not be enough to convince the citizens to live under nondemocracy, and they will prefer alternative routes. Here the only option open to them is revolution. In the next chapter, we will see how the elite can try to convince them not to undertake a revolution by offering a change in political institutions to make future redistribution more credible. Democratization will give the citizens political power, and thereby make much higher levels of future redistribution credible.

17.4. References

Acemoglu, Daron and James Robinson (2006) "Chapter 5: Nondemocratic Politics" in Acemoglu and Robinson *Economic Origins of Dictatorship and Democracy*, Cambridge University Press.

Barro, Robert J. (1979) "On the Determination of the Pubic Debt," *Journal of Political Economy*, 87, 940-971.

Wintrobe, Ronald (1998) *The Political Economy of Dictatorship*, New York: Cambridge University Press.

CHAPTER 18

Democratization

We will now discuss a dynamic model in which endogenous democratizations can happen. While far from a complete theory of institutional change or institutional origins, this is a useful first step, since the form of aggregating individual preferences into economic and social policies is one of the important institutional features, as we have seen above.

The approach is to construct a simple forward-looking rational model in which individuals value different political institutions differently, because they anticipate their influence on future equilibrium behavior.

18.1. The Emergence of Democratic Institutions

Democratic institutions emerged in most countries during the 19th or early 20th centuries with extensions of the franchise to portions of the population that had previously no votes. The notion of democracy as a way of making decisions is much older of course and goes back at least to Ancient Greece (and was discussed at length by Plato and Aristotle). During the colonial period in Mexico, indian cmmunities elected mayors democratically and in the English Civil War of the 1640s radical groups such as the 'Diggers' and the 'Levellers' pushed for universal male suffrage.

Existing evidence suggests that, as may have been expected, this change in voting rights led to important policy changes. For example, greater redistribution, better organization for labor, more investment in the education of the masses (see Lindert, 2004).

The emergence of democracies therefore offers an example of major institutional change, and gives us an opportunity to understand the process of how institutions get determined.

Here are a number of different theories of democratization (see below) the one we will develop is based on the idea that democratization occurs because of competition between those who do not have the vote and those who do. In this view, democratization is a change in political institutions that affects how political power is allocated in the future.

We will now discuss a model formalizing the view that democratization emerges as a result of the conflict between ins and outs. Clearly, this falls in the broad category of social conflict views of institutions.

This is partly motivated by the historical evidence. Most historical evidence suggests an important role for conflict between those who controlled political power and those without political power. For example, in Europe the preponderance of case study and historical evidence suggests that most major extensions of the franchise happened amidst social unrest or even in the face of a severe threat of revolution by the disenfranchised (the workers, the poor or the lower middle-class) - see Acemoglu and Robinson (2006).

There are a number of conceptual issues that this discussion brings out. In particular:

Why would groups that have political power now transfer to others? The answer to this is related to the notion that the outs can threaten the system, or cause a revolution.

But then why not simply give them redistribution or adopt the policies that they want rather than change the whole political system? The answer to this question will relate to the issue of political institutions as a "commitment device"—democratization not only changes policies today, but also regulates political power in the future. Recall that this is a key role of political institutions, so we have an example here where this role of political institutions is essential.

18.2. A Model of Democratization

The model is a direct extension of the one developed before and we will use the same notation and we again refer to the infinite horizon discounted repeated game as $G^{\infty}(\beta)$. There are two big extensions because we endow the elite with two new instruments to stay in power. The first is repression, which is clearly used often in reality. The second is giving away their power - democratization. There is again a continuum 1 of agents with a rich elite and poor citizens as before, with fractions, δ and $1-\delta$. Initially there is a nondemocracy but the citizens can contest power through collective action, and in a democracy the median voter will be a poor citizen. The structure of de facto power is exactly as before so that the cost of a revolution is μ_t where $\mu_t \in \{\mu^L, \mu^H\}$ and $\Pr(\mu_t = \mu^H) = q$ irrespective of whether $\mu_{t-1} = \mu^H$ or μ^L. We again normalize so that $\mu^L = 1$ and use the notation $\mu^H = \mu$.

The timing of the stage game is an extension to what we had before. In each period the elite can decide whether or not to create democracy and whether or not to repress. If democracy is created, the median voter, a poor citizen, sets the tax rate. We assume that if democracy is created, it cannot be rescinded, so the society always remains a democracy. We assume that if repression is chosen a revolution cannot be undertaken and the stage game is over for that period with agents getting the repression payoffs.

As a result, utilities are now given by $U^i = \sum_{t=0}^{\infty} \beta^t \hat{y}_t^i$ where U^i applies only when there is no revolution in equilibrium. Pre-tax incomes are given by (17.1), except that now there can also be costs due to repression which affect net income. In particular, the post-tax net return of agent i is

$$\hat{y}^i = \omega \Delta y^i + (1-\omega)\left((1-\tau)\, y^i + (\tau - C(\tau))\, \bar{y}\right), \tag{18.1}$$

where Δ is the cost due to repression with $\omega = 0$ denoting no repression and $\omega = 1$ denoting repression. We model the cost of repression as we did the costs of revolution. If the elite decides to repress then all agents lose some fraction of their income in the period of repression. We assume that $\Delta = 1 - \kappa$, which makes the effective cost of repression is equal to κy^i. We adopt the assumption that the citizens lose the same fraction of income as the elite only for symmetry, and this plays no major role in the analysis, since the repression decision is taken by the elite.

The timing of moves in the stage game is now as follows.

- The state $\mu_t \in \{\mu^L, \mu^H\}$ is revealed.
- The elite decide whether or not to use repression, $\omega \in \{0,1\}$. If $\omega = 1$, the poor cannot undertake a revolution and the stage game ends.
- If $\omega = 0$, the elite decide whether or not to democratize, $\phi \in \{0,1\}$. If they decide not to democratize, they set the tax rate τ^N.
- The citizens decide whether or not to initiate a revolution, $\rho \in \{0,1\}$. If $\rho = 1$ they share the remaining income forever. If $\rho = 0$ and $\phi = 1$ the tax rate τ^D is set by the median voter (a poor citizen). If $\rho = 0$ and $\phi = 0$, then the tax rate is τ^N.

We initially characterize Markov perfect equilibria of this game where players are restricted to playing Markov strategies which are functions only of the current state of the game.

The state of the game consists of the current opportunity for revolution, represented by either μ^L or μ^H, and the political state P which is either N (nondemocracy) or D (democracy). More formally, let $\sigma^r = \{\omega(\cdot), \phi(\cdot), \tau^N(\cdot)\}$ be the notation for the actions taken by the elite, while $\sigma^p = \{\rho(\cdot), \tau^D\}$ are the actions of the poor. σ^r consists of a decision to repress $\omega : \{\mu^L, \mu^H\} \to \{0,1\}$, or to create democracy $\phi : \{\mu^L, \mu^H\} \to \{0,1\}$, when $P = N$, and a tax rate $\tau^N : \{\mu^L, \mu^H\} \to [0,1]$, when $\phi = 0$ (i.e., when democracy is not extended). Clearly, if $\phi = 0$, P remains at N, and if $\phi = 1$, P switches to D forever, thus we do not make these strategies explicit functions of the political state. The actions of the citizens consist of a decision to initiate a revolution, $\rho : \{\mu^L, \mu^H\} \times \{0,1\}^2 \times [0,1] \to \{0,1\}$, and possibly a tax rate $\tau^D \in [0,1]$ when the political state is $P = D$. Here $\rho(\mu, \omega, \phi, \tau^N)$ is the revolution decision of the citizens which is conditioned on the current actions of the elite, as well as on the state, since the elite move before the citizens in the stage game according to the timing of events above. Then, a Markov perfect equilibrium is a strategy combination, $\{\tilde{\sigma}^r, \tilde{\sigma}^p\}$ such that $\tilde{\sigma}^p$ and $\tilde{\sigma}^r$ are best-responses to each other for all μ_t and P.

We can characterize the equilibria of this game by writing the appropriate Bellman equations. Define $V^p(R, \mu^S)$ as the return to citizens if there is a revolution starting in state $\mu^S \in \{\mu^L, \mu^H\}$. This value is naturally given by

$$V^p(R, \mu^S) = \frac{(1-\mu^S)\bar{y}}{(1-\delta)(1-\beta)}, \tag{18.2}$$

which is the per-period return from revolution for the infinite future discounted to the present. Also, because the elite lose everything, $V^r(R, \mu^S) = 0$ whatever is the value of μ^S. Moreover, recall that we have assumed $\mu^L = 1$, so $V^p(R, \mu^L) = 0$, and the citizens would never attempt a revolution when $\mu_t = \mu^L$.

In the state (N, μ^L) the elite are in power and there is no threat of revolution, so in any Markov Perfect Equilibrium, $\phi = \omega = 0$ and $\tau^N = \tau^r = 0$. This just says that when the elite are in power and the citizens cannot threaten them, the elite do not repress and set their preferred tax rate which is zero. Therefore, the values of citizens and elite agents, $i = p$ or r, are given by:

$$V^i(N, \mu^L) = y^i + \beta \left[qV^i(N, \mu^H) + (1-q)V^i(N, \mu^L) \right]. \tag{18.3}$$

Now (18.3) says that the value to an agent of type i in a nondemocracy when there is no threat of a revolution is equal to a payoff of y^i today, plus the expected continuation value discounted back to today (which is why it is multiplied by β). The payoff today is y^i because taxes are set at zero and each person simply consumes their income. The continuation value is made up of two terms; the second, $(1-q)V^i(N, \mu^L)$ is the probability that μ^L arises tomorrow, times the value of being in that state $V^i(N, \mu^L)$. In this case tomorrow is the same as today and this is why the same value 'recurs'. The first term, $qV^i(N, \mu^H)$, is the probability that μ^H arises tomorrow, multiplied by the value of that state, $V^i(N, \mu^H)$. This value is different because now there is a potential threat to the regime. To see how this will play out we need to understand what the value $V^i(N, \mu^H)$ looks like.

Consider the state (N, μ^H), where there is a nondemocracy, but it is relatively attractive to mount a revolution. Suppose that the elite play $\phi = \omega = 0$ and $\tau^N = \tau^r$, that is, they neither create democracy nor repress nor redistribute to the citizens. Then, we would have

$$V^p(N,\mu^H) = \frac{y^p}{1-\beta}.$$

The *revolution constraint* is equivalent to: $V^p(R,\mu^H) > V^p(N,\mu^H)$, so that without any redistribution or democratization, the citizens prefer to initiate a revolution when $\mu_t = \mu^H$. This is equivalent to $\theta > \mu$, and says that revolution becomes attractive when θ is sufficiently high, i.e. when inequality is sufficiently high.

Since the revolution is the worst outcome for the elite, they will try to prevent it. They can do this in three different ways. First, the elite can choose to maintain political power, $\phi = 0$, but redistribute through taxation. In this case, the poor obtain $V^p(N,\mu^H,\tau^N)$ where τ^N is the specific value of the tax rate chosen by the elite. Second, the elite can create democracy.

Finally, let us introduce another option: repression (oppression). The elite can use repression, which is costly, but prevents the revolution threat without having to democratize. Let $V^i(O,\mu \mid \kappa)$ be the value function of agent $i = p,r$ in state μ when the elite pursue the strategy of repression and the cost of repression is κ. We condition these values explicitly on κ to emphasize the importance of the cost of repression, and to simplify notation when we define threshold values below.

If the elite create democracy or attempt to stay in power by redistributing, the citizens may still prefer a revolution, thus:

$$\begin{aligned} V^p(N,\mu^H) &= \omega V^p(O,\mu^H \mid \kappa) + \\ &\quad (1-\omega) \max_{\rho \in \{0,1\}} \rho V^p(R,\mu^H) + (1-\rho)(\phi V^p(D) + (1-\phi)V^p(N,\mu^H,\tau^N)), \end{aligned}$$

where $V^p(D)$ is the return to the citizens in democracy. Note here how the value of the citizens depends on the decision variables ω and ϕ of the elite. If $\omega = 1$ the the elite choose to repress, citizens cannot revolt and get the continuation value $V^p(O,\mu^H \mid \kappa)$. If $\omega = 0$ then what the citizens compare $V^p(R,\mu^H)$ to depends on the decision by the elite as to whether or not create democracy. If $\phi = 1$ then they choose between revolution and democracy. If $\phi = 0$ they choose between revolution and accepting the promise of redistribution at the tax rate τ^N.

We first focus on the trade-off for the elite between redistribution and democratization and then integrate repression into the analysis. The return to the citizens when the elite choose the redistribution strategy is:

$$V^p(N,\mu^H,\tau^N) = y^p + \tau^N(\bar{y} - y^p) - C(\tau^N)\bar{y} + \beta\left[qV^p(N,\mu^H,\tau^N) + (1-q)V^p(N,\mu^L)\right]. \tag{18.4}$$

The elite redistribute to the citizens, taxing all income at the rate τ^N. The citizens therefore receive their income y^p from their own earnings and a net transfer of $\tau^N(\bar{y} - y^p) - C(\tau^N)\bar{y}$. If in the next period we are still in state $\mu_{t+1} = \mu^H$, redistribution continues. But, if the state switches to $\mu_{t+1} = \mu^L$, redistribution stops and the citizens receive $V^p(N, \mu^L)$. This captures our intuitive ideas that the elite cannot commit to future redistribution, unless the future also poses an effective revolution threat.

The second strategy to prevent the revolution is to democratize, $\phi = 1$. Since $1 - \delta > 1/2$, in a democracy the median voter is a citizen and the equilibrium tax rate is τ^p and $T = (\tau^p - C(\tau^p))\,\bar{y}$. The returns to citizens and elite agents in democracy are therefore:

$$V^p(D) = \frac{y^p + \tau^p(\bar{y} - y^p) - C(\tau^p)\bar{y}}{1-\beta} \text{ and } V^r(D) = \frac{y^r + \tau^p(\bar{y} - y^r) - C(\tau^p)\bar{y}}{1-\beta}. \tag{18.5}$$

These expressions exploit the fact that once created democracy consolidates and there are never any coups.

Will democratization prevent a revolution? The answer is not obvious. It might be that revolution in the state $\mu_t = \mu^H$ is so attractive that even democratization is not sufficient to prevent revolution. It is straightforward to see that the condition for democratization to prevent revolution is $V^p(D) \geq V^p(R, \mu^H)$.

To determine whether the elite can prevent the revolution with the redistribution strategy, let $V^p(N, \mu^H, \tau^N = \tau^p)$ be the maximum utility that can be given to the citizens without democratizing. This maximum utility is achieved by setting $\tau^N = \tau^p$ in (18.4). Therefore, combining (18.3) and (18.4), we obtain:

$$V^p(N, \mu^H, \tau^N = \tau^p) = \frac{y^p + (1 - \beta(1-q))\,(\tau^p(\bar{y} - y^p) - C(\tau^p)\bar{y})}{1-\beta}. \tag{18.6}$$

(18.6) has a nice interpretation. It says that $V^p(N, \mu^H, \tau^N = \tau^p)$ is equal to the present discounted value of y^p, the pre-tax income of citizens, plus the expected present value of net redistribution from the elite to the citizens. Net redistribution is given by the expression $(\tau^p(\bar{y} - y^p) - C(\tau^p)\bar{y})$ but this only occurs today, and a proportion q of the time in the future when the state is μ^H.

If $V^p(N, \mu^H, \tau^N = \tau^p) < V^p(R, \mu^H)$, then the maximum transfer that can be made when $\mu_t = \mu^H$ is not sufficient to prevent a revolution. Notice that as long as

$$\mu \geq \theta - (\tau^p(\theta - \delta) - (1-\delta)C(\tau^p))\,. \tag{18.7}$$

holds, we have that $V^p(D) \geq V^p(R, \mu^H)$. It is clear that we have $V^p(N, \mu^H = 1, \tau^N = \tau^p) > V^p(R, \mu^H = 1)$ since a revolution generates a zero payoff to the citizens forever. This implies that when $\mu^H = 1$ it must be the case that the value to the citizens of accepting redistribution at the rate τ^p in state μ^H is greater than the value from having a revolution. Also note that,

$$
\begin{aligned}
(18.8) \qquad V^p(N,\mu^H &= 0,\tau^N=\tau^p) = y^p + (1-\beta(1-q))\left(\tau^p(\bar{y}-y^p) - C(\tau^p)\bar{y}\right) \\
&< V^p(R,\mu^H=0) = \frac{\bar{y}}{1-\delta}
\end{aligned}
$$

so that the payoff from a revolution must be greater when $\mu^H = 0$. Since $V^p(R,\mu^H)$ is monotonically increasing and continuous in μ, by the intermediate value theorem there exists a unique $\mu^* \in (0,1)$ such that

$$
(18.9) \qquad V^p(N,\mu^*,\tau^N=\tau^p) = V^p(R,\mu^*).
$$

When $\mu < \mu^*$, concessions do not work so that the elite are forced to either democratize or repress. When $\mu \geq \mu^*$, they can prevent revolution by temporary redistribution, which is always preferable to them when the alternative is democratization (since with democratization, redistribution is not temporary but permanent). In this case the tax which the elite set, which is denoted by $\hat{\tau}$, will be set exactly to leave the citizens indifferent between revolution and accepting concessions under a nondemocratic regime, i.e. $\hat{\tau}$ satisfies the equation $V^p(N,\mu^H,\tau^N=\hat{\tau}) = V^p(R,\mu^H)$.

To determine equilibrium actions, we need to compare the payoffs to the elite from staying in power using redistribution and from democracy to the costs of repression. Without loss of generality we limit attention to situations where the elite play a strategy of always repressing, rather than more complicated strategies of repressing sometimes and using redistribution some other time. By standard arguments, these values satisfy the Bellman equations:

$$
\begin{aligned}
(18.10) \qquad V^i(O,\mu^H &\mid \kappa) = \Delta y^i + \beta\left[qV^i(O,\mu^H\mid\kappa) + (1-q)V^i(O,\mu^L\mid\kappa)\right], \\
V^i(O,\mu^L &\mid \kappa) = y^i + \beta\left[qV^i(O,\mu^H\mid\kappa) + (1-q)V^i(O,\mu^L\mid\kappa)\right],
\end{aligned}
$$

which take into account that the cost of repression will only be incurred in the state where the revolution threat is active, i.e., when $\mu_t = \mu^H$.

Together with the definition for Δ, these Bellman equations can be solved simultaneously to derive the vales to the elite and citizens from repression,

$$
\begin{aligned}
(18.11) \qquad V^r(O,\mu^H &\mid \kappa) = \frac{y^r - (1-\beta(1-q))\kappa y^r}{1-\beta} \text{ and} \\
V^p(O,\mu^H &\mid \kappa) = \frac{y^p - (1-\beta(1-q))\kappa y^p}{1-\beta}.
\end{aligned}
$$

The value function $V^r(O,\mu^H \mid \kappa)$ has a clear interpretation. It says that the payoff to the elite from a strategy of repression is the discounted sum of their income, $y^r/(1-\beta)$ minus the expected cost of repressing. The net present value of the cost of repressing is

$(1-\beta(1-q))\,\kappa y^r/\,(1-\beta)$ for the elite, because they pay this cost today and a fraction q of the time in the future.

To understand when repression occurs we need to compare $V^r(O,\mu^H \mid \kappa)$ to $V^r(D)$ when $\mu < \mu^*$; and to $V^r(N,\mu^H,\tau^N = \hat{\tau})$ when $\mu \geq \mu^*$. We will now determine two threshold values for the cost of repression, this time called κ^* and $\bar{\kappa}$, such that the elite are indifferent between their various options at these threshold levels. More specifically, let κ^* be such that the elite are indifferent between promising redistribution at the tax rate $\tau^N = \hat{\tau}$ and repression, $V^r(O,\mu^H \mid \kappa^*) = V^r(N,\mu^H,\tau^N = \hat{\tau})$. This equality implies

$$\kappa^* = \frac{1}{\theta}\left(\delta C(\hat{\tau}) - \hat{\tau}\left(\delta - \theta\right)\right). \tag{18.12}$$

Similarly, let $\bar{\kappa}$ be such that at this cost of repression, the elite are indifferent between democratization and repression, i.e., $V^r(O,\mu^H \mid \bar{\kappa}) = V^r(D)$, which implies that

$$\bar{\kappa} = \frac{1}{\theta(1-\beta(1-q))}\left(\delta C(\tau^p) - \tau^p\left(\delta - \theta\right)\right). \tag{18.13}$$

It is immediate that $\bar{\kappa} > \kappa^*$, i.e., if the elite prefer repression to redistribution, then they also prefer repression to democratization. Therefore, we have that the elite will prefer repression when $\mu \geq \mu^*$ and $\kappa < \kappa^*$, and also when $\mu < \mu^*$ and $\kappa < \bar{\kappa}$.

Proposition 18.1. *There is a unique Markov perfect equilibrium $\{\tilde{\sigma}^r, \tilde{\sigma}^p\}$ in the game $G^\infty(\beta)$, and it is such that:*

- *If $\theta \leq \mu$, then the revolution constraint does not bind and the elite can stay in power without repressing, redistributing or democratizing.*
- *If $\theta > \mu$, then the revolution constraint binds. In addition, let be μ^* defined by (18.9), and κ^* and $\bar{\kappa}$ be defined by (18.12) and (18.13). Then:*
  - *If $\mu \geq \mu^*$ and $\kappa \geq \kappa^*$, repression is relatively costly and the elite redistribute income in state μ^H to avoid revolution.*
  - *If $\mu < \mu^*$ and $\kappa < \bar{\kappa}$, or $\kappa \geq \bar{\kappa}$ and (18.7) does not hold, or if $\mu \geq \mu^*$ and $\kappa < \kappa^*$, the elite use repression in state μ^H.*
  - *If $\mu < \mu^*$, (18.7) holds, and $\kappa \geq \bar{\kappa}$, concessions are insufficient to avoid a revolution and repression is relatively costly. In this case, in state μ^H the elite democratize.*

Democracy arises only if $\mu < \mu^*$, repression is relatively costly, i.e., $\kappa \geq \bar{\kappa}$ and if (18.7) holds. Notice that this critical threshold for the cost of repression, $\bar{\kappa}$, is increasing in inequality (increasing in θ), more specifically we can again show by an argument identical to the one we used before:

$$\frac{d\bar{\kappa}}{d\theta} > 0.$$

Intuitively, when inequality is higher, democracy is more redistributive, i.e., τ^p is higher, and hence more costly to the rich elite. They are therefore more willing to use repression.

Democracy therefore emerges as an equilibrium outcome only in societies with intermediate levels of inequality. In very equal or very unequal societies, democracy does not arise as an equilibrium phenomenon. In very equal societies, there is little incentive for the disenfranchised to contest power and the elite do not have to make concessions, neither do they have to democratize. In very unequal societies the elite cannot use redistribution to hang onto power, but since in such a society democracy is very bad for the elite, they use repression rather than having to relinquish power. It therefore tends to be in societies with intermediate levels of inequality that democracy emerges. Here inequality is sufficiently high for challenges to the political status quo to emerge, but not high enough that the elite find repression attractive.

We will see next that even without the restriction to Markov Perfect equilibria, similar results obtain: revolution can be stopped with temporary redistribution when $\mu \geq \mu^{**}$ where $\mu^{**} < \mu^{*}$, hence for a larger range of parameters, but if $\mu < \mu^{**}$, the elite cannot use concessions to avoid a revolution.

Perhaps paradoxically, a high q makes franchise extension less likely. A high q corresponds to an economy in which the citizens are well organized, so they frequently pose a revolutionary threat. Alternatively, if μ^L is sufficiently less than one, then even in this state, the elite have to redistribute to the citizens. In this case, a low value of μ^L would also lead to the same result. A naive intuition may have been that in this case franchise extension would be more likely. This is not the case, however, because with a frequent revolutionary threat, future redistribution becomes *credible*. When the citizens have the power to oversee the promises made to them, then there is less need for the elite to undertake a change in institutions in order to increase the future political power of the citizens.

18.3. Subgame Perfect Equilibria

In the previous section we characterized a subset of the subgame perfect equilibria of $G^{\infty}(\beta)$. In this section we analyze our basic dynamic model of democratization without the restriction to Markovian strategies. More specifically, we look for subgame perfect equilibria. In general there are many subgame perfect equilibria of this game which are supported by various history dependent strategies and our analysis here mirrors that of previous chapter. We are interested in understanding the extent to which punishment strategies can make redistribution in state μ^L credible. Thus we look for the best possible equilibrium for the elite, which will be the one that prevents democratization for the largest set of parameter values. Therefore, implicitly we are interested in the maximum possible amount of credible redistribution to the citizens in the nondemocratic regime. To keep things simple in this section, we abstract from the use of repression though this can be easily added. As in the previous chapter, the analysis in this section focuses on showing that there exists a cutoff level of μ, $\tilde{\mu}^{**} < \mu^*$ such that when $\mu \geq \tilde{\mu}^{**}$, there will be redistribution without democratization, preventing a revolution. In contrast when $\mu < \tilde{\mu}^{**}$, the equilibrium will feature democratization when $\mu_t = \mu^H$.

Exactly as in the analysis of nondemocratic politics in the previous chapter, we study the circumstances under which the elite can redistribute at some tax rate $\tau^L > 0$ in state μ^L, and thus avoid the transition away from the nondemocratic regime even when $\mu < \mu^*$. There we saw that the limitation on such redistribution was that it had to be incentive compatible for the elite, i.e., it had to be such that the payoff to the elite from redistributing according to the vector $\left[\tau^L, \tau^H\right]$, given by the value $V^r(N, \mu^L, \left[\tau^L, \tau^H\right])$, had to be greater than the payoff from deviating, $V_d^r(N, \mu^L)$.

There is only one substantive difference between the game we studied in the previous chapter and the one here. This difference is that, as long as (18.7) holds, when the nondemocratic regime collapses, there will be a transition to democracy. Therefore, the value $V_d^r(N, \mu^L)$ here will take into account that when the elite deviate in state μ^L, their "punishment" in state μ^H will be democratization instead of revolution as before. This is because it is not a subgame perfect strategy for the citizens to threaten a revolution after the elite democratize, since they obtain greater payoff from democracy than revolution. Consequently, if the elite democratizes, it in effect forestalls revolution. This implies that the value $V_d^r(N, \mu^L)$ for the elite is given by the following recursion:

$$V_d^r(N, \mu^L) = y^r + \beta \left[qV^r(D) + (1-q)\, V_d^r(N, \mu^L) \right]$$

where $V^r(D)$ is as in (18.5).

As before, only redistribution at the tax vector $\left[\tau^L, \tau^H\right]$ such that

$$V^r(N, \mu^L, \left[\tau^L, \tau^H\right] \geq V_d^r(N, \mu^L)$$

is credible. In addition, it is straightforward to see that the derivations leading up to $V^p(N, \mu^H, \left[\tau^L, \tau^H\right])$ in (17.24) from the previous chapter still apply. So the incentive compatibility constraint for the elite will only differ from before because of the change in $V_d^r(N, \mu^L)$.

Just as in the analysis of nondemocratic politics in the previous chapter, in general the best equilibrium for the elite will need to take into account the incentives to smooth taxes over time. However, in order to simplify the discussion here, and because the concept of tax smoothing is not central to our analysis, we simply focus on characterizing the minimum value of μ^H, such that the elite can avoid democratizing. We denote this $\tilde{\mu}^{**}$, such that when $\mu \geq \tilde{\mu}^{**}$ nondemocracy can be maintained with promises of redistribution. It is still the case that the maximum tax rate in the state μ^H is τ^p. So we only need to find the maximum incentive compatible redistribution in state μ^L, which we now denote by $\tilde{\tau}'$. By an identical argument to before, it is given by

$$V^r(N, \mu^L, [\tilde{\tau}', \tau^p]) = V^r_d(N, \mu^L).$$

Since $V^r(D) > 0$, the citizens can punish deviation less when the elite can democratize and this implies that deviation is more attractive for the elite. In consequence it is immediate that $\tilde{\tau}' < \bar{\tau}'$ which satisfies (17.25).

In addition, since the value of a revolution to the citizens is also the same, the formula for the critical value of the cost of revolution, $\tilde{\mu}^{**}$ must be identical to the one we derived for μ^{**} before, with the value of $\bar{\tau}'$ we derived there replaced by the new value of $\tilde{\tau}'$. Thus, the critical value $\tilde{\mu}^{**}$ can be easily found so that $V^p(N, \mu^H, [\tilde{\tau}', \tau^p]) = V^p(R, \mu^H)$ at $\mu^H = \tilde{\mu}^{**}$. This is:

$$\begin{aligned} \tilde{\mu}^{**} &= \theta - \beta(1-q)\left(\tilde{\tau}'(\theta - \delta) - (1-\delta)C(\tilde{\tau}')\right) \\ &\quad - (1 - \beta(1-q))\left(\tau^p(\theta - \delta) - (1-\delta)C(\tau^p)\right). \end{aligned} \tag{18.14}$$

The value of $\tilde{\mu}^{**}$ implied by (18.14) is greater than the value of μ^{**} in the previous chapter because, here, the potential punishments on the elite are less severe.

More important, it is clear that $\tilde{\mu}^{**} < \mu^*$ (where μ^* is given by (18.9)), and we have as before that if $\mu \geq \tilde{\mu}^{**}$ the elite can stay in power by redistributing. Equally important, when $\mu < \tilde{\mu}^{**}$, contrary to the previous chapter, there is no revolution, because the elite have an extra instrument—they can democratize.

In summary, allowing the elite and citizens to play non-Markovian strategies has very similar implications in this model as it did in the previous chapter. The threat of punishments by the citizens, in particular, the threat that they will undertake a revolution, implies that some amount of redistribution can be sustained in state μ^L. Interestingly, this amount is actually lower here since the possibility for the elite to democratize limits the punishment that the citizens can inflict on them. Most important however is that the main thrust of the analysis of the previous chapter applies. Though the ability to use punishment strategies increases the circumstances under which the elite can stay in power by making concessions, this does not eliminate the problem of credibility. When $\mu < \tilde{\mu}^{**}$, concessions do not work because of the absence of sufficient future credibility and the elite will be forced to democratize.

18.4. Alternative Approaches

The model we discussed above is based on a few key presuppositions. Clearly there are alternative conceptualizations of the mechanisms that lead to democracy, though these are seldom discussed in a parsimonious way (Huntington, 1991, pp. 37-38 lists 27 different factors that he claims have been said to promote democracy). The approach we took builds on several pillars. First, it places central emphasis on the fact that democracy is conceded in the face of potential conflict which is internal to a society (as do Therborn, 1977, and Rueschemeyer, Stephens and Stephens, 1992). Second, conflict over political institutions is instrumental - people fight over political institutions because of the different allocation of resources that different institutions lead to. Thus the framework stresses the economic benefits of different political regimes rather than people's intrinsic preference for one type of institution or another. Third, the demand for changes in political institutions comes because they influence the future distribution of political power and help to solve problems of commitment (building on the work of North and Weingast, 1989, and Weingast, 1997). In this section we discuss some other potential mechanisms.

Nearly all recent research accepts the second of these premises, while it abstracts from the latter. The emphasis on conflict and the form it takes varies. Closest to the spirit of the above model, Rosendorff (2001) examines the trade-off between fighting and democracy arguing that elites fight if the cost is lower than having to accept the policy preferred by the median voter. Rosendorf (2001) independently made the connection between inequality and democratization. Interestingly, in some sense all of this work stems not from Linz and Stepan (1978), often regarded as the key work in political science which has formed the modern work on democratization, but rather from Chapter 1 of Dahl (1971).

The political science literature has downplayed the collective action of the disenfranchised as the force leading to the creation or collapse of democracy. In its place it has put the idea that both democratic and nondemocratic regimes have an intrinsic propensity to self-destruct. In the context of democratization this approach is most associated with O'Donnell and Schmitter (1986) who play down the role of outside social pressure and instead emphasizes conflict within ruling authoritarian regimes. In their view democracy arises when some subset of the authoritarian coalition (the 'soft-liners') joins with the disenfranchised. Collier

(1999) develops a similar approach arguing that democracy arises as an 'elite project.' It is natural to think of the elite as heterogeneous and in this case one can imagine scenarios where one faction of the elite favors giving political rights to the disenfranchised because this will help to move policy or institutional choices in direction favored by them. Llavador and Oxoby (2005) present a model along these lines. Think of a situation where the elite favor some positive levels of taxation T. Imagine that T is the level of taxation which can be spent on the provision of public goods which increases the income of the elite. However, there are two factions of the elite, manufacturers and landowners and the public goods can benefit only manufacturers. Thus while landowners prefer $T = 0$, manufacturers prefer $T^m > 0$. Imagine a situation such that the revolution constraint does not bind so that the elite do not have to create democracy. Even though manufacturers prefer a lower rate of taxation that the citizens, they prefer the citizens tax rate, T^c to $T = 0$. Hence manufacturers prefer democracy to a nondemocratic regime controlled by landowners and if they get the chance will favor democratization. These ideas complement those I focused on in the sense that they still emphasize conflict over social choices as a driving force behind democracy.

The recent book of Bueno de Mesquita, Smith, Siverson and Morrow (2003) presents a theory of democratization which combines elements of both of these approaches. Like the model I develop above they emphasize the role of the threat of force in the creation of democracy since dictators oppose it while the disenfranchised certainly favor it. Nevertheless, in their model members of the "winning coalition" can favor democratization because of the way this influences the equilibrium public policy - in particular in the direction of the greater provision of public goods.

Going further than this, another set of scholars have argued that democratization may in certain circumstances be Pareto improving in the sense of being better for both the elite and the citizens. This research includes non-formal work by Kiser and Barzel (1991) and mathematical models by Green (1993), Weingast (1997), and Lizzeri and Persico (2004). For instance, Green (1993) argues that the creation of legislative institutions was a way for rulers to credibly signal information. Lizzeri and Persico's paper (2004) is based on the idea that when the franchise is restricted elites compete for a limited number of votes by providing private rather than (socially desirable) public goods. Democratization, by increasing the number of voters who must be attracted, induces competing parties to choose strategies with greater provision of public goods which, because this is socially efficient, can make everyone better off. One can think of this as a situation where all members of the elite prefer (T^m, T^m) but in their competitive struggle they can only find it optimal to provide some lower level (for convenience think of this as zero). In this case democratization, by shifting the policy to that preferred by the citizens, while it might not be as good for the elite as (T^m, T^m), is better than $(0, 0)$. The essence of this provocative set of ideas is that the origins of democracy may actually be consensual and may serve to solve a problem of coordination or commitment within the elite, rather than between the elite and the citizens (as emphasized above).

An alternative theoretical approach stems from the sociological literature on the origins of state institutions which has inspired an analyses of democratization by Bates (1991), Rogowski (1998) and Tilly (2004). These scholars argue that democracy, like the origins of representative institutions more generally, is a concession from authoritarian rulers necessary to raise taxation. The more elastic is the tax base, the harder it is for authoritarian rulers to raise taxes without agreement, and the greater the likelihood of concessions—here democracy. Hence Bates (1991, p. 25) points out that democracy is less likely in an agrarian society since land is easier to tax, than it is in a society dominated by physical or human capital. Moreover, he makes the argument that authoritarian rulers will be more willing to abide by democracy if they fear it less. He connects this to their economic power with respect to democracy—democrats cannot hurt previous elites a lot if they have sufficient economic strength, perhaps because taxing the elite leads to a collapse in the economy. Rogowski (1998) similarly emphasizes the impact of the ability of citizens to exit as leading to democracy.

A final and interesting approach has been developed by Ticchi and Vindigni (2003) who analyze a model where countries are engaged in inter-state warfare and political elites democratize in order to give their citizens greater incentives to fight.

Of course one could think of other ideas, such as changes in ideology (as a consequence of the Enlightenment?).

18.5. References

(1) Acemoglu, Daron and James A. Robinson (2000) "Why Did the West Extend the Franchise? Growth, Inequality and Democracy in Historical Perspective", *Quarterly Journal of Economics*, CXV, 1167-1199.

(2) Acemoglu, Daron and James A. Robinson (2005) *Economic Origins of Dictatorship and Democracy*, forthcoming Cambridge University Press.

(3) Bueno de Mesquita, Bruce D., James D. Morrow, Randolph M. Siverson and Alastair Smith (2003) *The Logic of Political Survival*, MIT Press; Cambridge.

(4) Collier, Ruth Berins (1999) *Paths Towards Democracy: The Working Class and Elites in Western Europe and South America*, Cambridge University Press, New York.

(5) Dahl, Robert A. (1971) *Polyarchy: Participation and Opposition*, Yale University Press, New Haven.

(6) Green, Edward J. (1993) "On the Emergence of Parliamentary Government: The Role of Private Information," *Federal Reserve Bank of Minneapolis Quarterly Review*, 17, 1-12.

(7) Huntington, Samuel P. (1991) *The Third Wave: Democratization in the Late Twentieth Century*, University of Oklahoma Press, Norman OK.

(8) Kiser, Edgar and Yoram Barzel (1991) "The Origins of Democracy in England," *Rationality and Society*, 3, 396-422.

(9) Lagunoff, Roger (2004) "Markov Equilibrium in Models of Dynamic Endogenous Political Institutions" http://www.georgetown.edu/faculty/lagunoff/

(10) Lagunoff, Roger (2006) "Dynamic Stability and reform of Political Institutions," http://www.georgetown.edu/faculty/lagunoff/

(11) Linz, Juan J. and Alfred Stepan (1978) *The Breakdown of Democratic Regimes*, Johns Hopkins University Press, Baltimore MD.

(12) Lizzeri, Alessandro and Nicola Persico (2004) "Why Did the Elites Extend the Suffrage? Democracy and the Scope of Government, With an Application to Britain's 'Age of Reform," *Quarterly Journal of Economics*, 119, 707-765.

(13) Llavador Humberto and Robert Oxoby (2005) "Partisan competition, growth and the franchise," *Quarterly Journal of Economics,* 120, 1155-89.

(14) North, Douglass C. and Barry R. Weingast (1989) "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England" *Journal of Economic History*, 49, 803-832.

(15) O'Donnell, Guillermo and Philippe C. Schmitter (1986) *Transitions from Authoritarian Rule: Tentative Conclusions about Uncertain Democracies*, Johns Hopkins University Press, Baltimore MD.

(16) Rogowski, Ronald (1998) "Democracy, Capital, Skill and Country Size: Effects os Asset Mobility and Regime Monopoly on the odds of Democratic Rule," in Paul W. Drake and Mathew D. McCubbins eds. *The Origins of Liberty*, Princeton; Princeton University Press.

(17) Rosendorff, B. Peter (2001) "Choosing Democracy," *Economics and Politics*, 13, 1-29.

(18) Rueschemeyer, Dietrich, Evelyn H. Stephens and John D. Stephens (1992) *Capitalist Development and Democracy*, University of Chicago Press, Chicago IL.

(19) Tarrow, Sidney (1998) *Power in Movement: Social Movements and Contentious Politics*, Second Edition, New York; Cambridge University Press.

(20) Tilly, Charles (1995) *Popular Contention in Britain, 1758-1834*, Cambridge; Harvard University Press.

(21) Tilly, Charles (2004) *Contention and Democracy in Europe, 1650-2000*, New York; Cambridge University Press.

(22) Therborn, Goran (1977) "The Rule of Capital and the Rise of Democracy," *New Left Review*, 103, 3-41.

(23) Ticchi, Davide and Andrea Vindigni (2003) "On Wars and Political Development. The Role of International Conflicts in the Democratization of the West," Unpublished, Department of Politics, Princeton University.

(24) Weingast, Barry R. (1997) "The Political Foundations of Democracy and the Rule of Law," *American Political Science Review*, 91, 245-263.

CHAPTER 19

Political Instability and Coups

The above analysis presumed that once democracy is created is permanent. In practice, as in many Latin American countries, there are frequent coups against democracy. Therefore a theory of changes in political institutions also needs to incorporate the possibility of such coups and reversions back to nondemocratic regimes. This will allow us to develop ideas not just about the circumstances under which democracy is created, but also why it is that some democracies concolidate and others do not.

19.1. Basic Model

As before, we consider an infinite horizon model, denoted $G^{\infty}(\beta)$, with a continuum 1 of agents divided into $1-\delta > 1/2$ poor citizens, while the remaining δ form a rich elite. Initially, political power is concentrated in the hands of the elite, but the median voter is a poor agent. Agents' expected utility at time $t=0$ is again given by $U^i = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta \hat{y}_t^i$ where again $\hat{y}_t^i$ is post tax income (inclusive of costs of political conflict).

The collective action technology via which the citizens can mount a revolution and the payoffs to revolution are identical to those we have specified previously. In a democracy, the elite can attempt a coup. After the coup every agent loses a fraction φ^S of their income, where the threat state is $S = H, L$ and $\varphi^H < \varphi^L$, and the political situation reverts back to the initial status quo with the elite controlling political power. Similar to our analysis of the revolution threat, we assume that in the low threat state, the coup threat is not active, so we set $\varphi^L = 1$. The relevant cost will therefore be the cost of the elite in the state $S = H$, $\varphi^H = \varphi$, and only in this state will the elite want to mount a coup. We assume that $\Pr(\varphi_t = \varphi) = s$. We assume that both q and s are less than $1/2$ so that crises which facilitate the exercise of de facto power are relatively rare events.

If a coup is mounted, then $\mu_t = \mu^L$ at first so that there is no revolution immediately. Similarly, if democratization occurs, then democracy starts with the coup cost at 1, implying that a democracy has at least some window of opportunity before a coup can occur. Finally, in each nondemocratic period the elite have to decide whether or not to democratize, and if they do, the society becomes a democracy, and the median voter, a citizen, sets the tax rate.

The timing of events within a period can be summarized as follows.

- the state μ_t or φ_t is revealed.
- the citizens set the tax rate, τ_t^D, if we are in a democracy, and the elite set τ_t^N, otherwise.
- in a nondemocratic regime, the elite decide whether or not to repress, ω, or democratize, ϕ. In a democracy, they decide whether to mount a coup, ζ. If they democratize or undertake a coup, the party that comes to power decides whether to keep the tax τ_t set at stage 2 or set a new tax rate.
- if $P = N$ and $\omega = 0$ the citizens decide whether or not to initiate a revolution, ρ. If there is a revolution, they share the remaining income of the economy. If there is no revolution, the tax rate decided at 2 or 3 gets implemented.
- incomes are realized and consumption takes place.

We will again characterize the Markov perfect equilibria of this game in which strategies only depend on the current state of the world. The state is one of (D, φ^H), (D, φ^L), (N, μ^L), or (N, μ^H), where N denotes elite in power (nondemocratic regime) and D denotes democracy. Let $\sigma^r = \{\omega(\cdot), \phi(\cdot), \tau^N(\cdot), \zeta(\cdot), \tau^N\}$ be the notation for the actions taken by the elite, while $\sigma^p = \{\rho(\cdot), \tau^D(\cdot)\}$ are the actions of the citizens. σ^r consists of a decision to repress $\omega : \{\mu^L, \mu^H\} \to \{0, 1\}$, or to create democracy $\phi : \{\mu^L, \mu^H\} \to \{0, 1\}$, when $P = N$, and a tax rate $\tau^N : \{\mu^L, \mu^H\} \to [0, 1]$ when $\phi = 0$ (i.e., when democracy is not created). Clearly, if $\phi = 0$, P remains at N, and if $\phi = 1$, P switches to D. When $P = D$ the elite make a coup decision which is a function $\zeta : \{\varphi^L, \varphi^H\} \times [0, 1] \to \{0, 1\}$ where $\zeta(\varphi, \tau^D)$ is the coup decision when the state is φ and the median voter sets the tax rate τ^D. If $\zeta = 1$ then the political state switches to $P = N$ and the elite also get to re-set the tax rate, $\tau^N \in [0, 1]$. The actions of the citizens consist of a decision to initiate a revolution, $\rho : \{\mu^L, \mu^H\} \times \{0, 1\}^2 \times [0, 1] \to \{0, 1\}$. Here $\rho(\mu, \omega, \phi, \tau^N)$ is the revolution decision of the citizens which is conditioned on the current actions of the elite, as well as on the state. When $P = D$ the citizens set the tax rate, $\tau^D : \{\varphi^L, \varphi^H\} \to [0, 1]$.

Then, a Markov perfect equilibrium is a strategy combination, $\{\tilde{\sigma}^r, \tilde{\sigma}^p\}$ such that $\tilde{\sigma}^p$ and $\tilde{\sigma}^r$ are best-responses to each other for all μ_t, φ_t and P. As usual, we will characterize the Markov perfect equilibria by writing the appropriate Bellman equations.

Let $V^i(D, \varphi^L)$ be the value of an agent of type $i = p, r$ when there is democracy and when the cost of mounting a coup is φ^L. Similarly, let $V^i(\varphi^H)$ be the value of agent i when the cost is φ^H (in which case there may be a switch to a nondemocratic regime, as a result of a coup).

When the state is (D, φ^L), there are no constraints on the median voter, so he will choose the tax rate $\tau^D = \tau^p$. The returns to citizens and elite agents are:

$$V^i(D, \varphi^L) = y^i + \tau^p \left(\bar{y} - y^i\right) - C(\tau^p)\bar{y} + \beta \left[sV^i(\varphi^H) + (1-s)V^i(D, \varphi^L)\right], \tag{19.1}$$

for $i = p, r$, and as before $\tau^p \left(\bar{y} - y^i\right) - C(\tau^p)\bar{y}$ represents the net amount of redistribution at tax rate τ^p faced by agent i.

Next consider the state (D, φ^H), where the poor may set a tax rate different from the one they prefer in an attempt to prevent a coup. Denote the values in the state (D, φ^H), when the tax rate is τ^D by $V^i(D, \varphi^H, \tau^D)$, which are given as:

$$V^i(D, \varphi^H, \tau^D) = y^i + \tau^D \left(\bar{y} - y^i\right) - C(\tau^D)\bar{y} + \beta \left[sV^i(D, \varphi^H, \tau^D) + (1-s)V^i(D, \varphi^L)\right]. \tag{19.2}$$

Clearly, $\tau^D (\bar{y} - y^p) - C(\tau^D)\bar{y} \leq \tau^p (\bar{y} - y^p) - C(\tau^p)\bar{y}$ and $\tau^D (\bar{y} - y^r) - C(\tau^D)\bar{y} \geq \tau^p (\bar{y} - y^r) - C(\tau^p)\bar{y}$, for $\tau^D \leq \tau^p$.

After observing the tax rate τ^D, the elite may still decide to mount a coup, so the values in the state (D, φ^H) are not necessarily equal to $V^i(D, \varphi^H, \tau^D)$. Instead, we denote them by $V^i(\varphi^H)$ such that:

$$\begin{aligned} V^r(\varphi^H) &= \max_{\zeta \in \{0,1\}} \zeta \left(V^r(N, \mu^L) - \varphi y^r\right) + (1-\zeta)V^r(D, \varphi^H, \tau^D) \\ V^p(\varphi^H) &= \zeta \left(V^p(N, \mu^L) - \varphi y^p\right) + (1-\zeta)V^p(D, \varphi^H, \tau^D), \end{aligned} \tag{19.3}$$

where recall that $\zeta = 1$ implies a coup. The first line of (19.3) says that the value $V^r(\varphi^H)$ for the elite in the high threat state depends on their own choice about whether or not to mount a coup. In making this decision they compare the value from not mounting a coup and accepting a concession of a tax rate of τ^D from the citizens, which is $V^r(D, \varphi^H, \tau^D)$, to the value from mounting a coup. This value is $V^r(N, \mu^L) - \varphi y^r$ which is the value of being in nondemocracy when there is no threat of a revolution, $V^r(N, \mu^L)$, minus the cost of a coup φy^r. The second line states that the value for the citizens in this state, $V^p(\varphi^H)$, also depends on what the elite do. If $\zeta = 1$, then the citizens find themselves in a nondemocracy and their continuation value if $V^p(N, \mu^L)$, minus the cost of the coup φy^p, while if $\zeta = 0$, there is no coup, democracy persists, and the citizens' value is $V^p(D, \varphi^H, \tau^D)$.

We shall now derive the *coup constraint*, which parallels the revolution constraint before. This constraint immediately follows from (19.3) by checking when a coup will be attractive, provided that the median voter sets his preferred tax rate $\tau^D = \tau^p$. It is therefore:

$$V^r(N, \mu^L) - \varphi y^r > V^r(D, \varphi^H, \tau^D = \tau^p). \tag{19.4}$$

This coup states that a coup occurs if the gain to the elite of capturing political power and reducing taxation, $V^r(N,\mu^L) - V^r(D,\varphi^H,\tau^D = \tau^p)$, is greater than the cost of the coup, φy^r.

We can now determine a critical value of φ, denoted $\hat{\varphi}$, such that as long as $\varphi \geq \hat{\varphi}$, a coup is never beneficial for the elite, even if the citizens tax at $\tau^D = \tau^p$ in state (D,φ^H). This critical value will clearly satisfy inequality (19.4) as an equality with $\tau^D = \tau^p$. Therefore,

$$\hat{\varphi} = \frac{V^r(N,\mu^L) - V^r(D,\varphi^H,\tau^D = \tau^p)}{y^r}. \tag{19.5}$$

In words, this equation specifies that the critical threshold is such that the loss of current income for the elite is equivalent to the discounted loss of living forever under democracy with the tax rate most preferred by the citizens, $V^r(D,\varphi^H,\tau^D = \tau^p)$, versus undertaking a coup and switching to a nondemocratic regime, which will give the value $V^r(N,\mu^L)$.

However, equation (19.5) is not informative unless we obtain expressions for $V^r(N,\mu^L)$ and $V^r(D,\varphi^H,\tau^D = \tau^p)$. The return to the elite of always remaining in democracy with a tax rate $\tau^D = \tau^p$ is simply:

$$V^r(D,\varphi^H,\tau^D = \tau^p) = \frac{y^r + \tau^p\left(\bar{y} - y^r\right) - C(\tau^p)\bar{y}}{1-\beta}. \tag{19.6}$$

We next compute the value of nondemocracy to the elite $V^r(N,\mu^L)$. First, with the standard arguments, we have:

$$V^i(N,\mu^L) = y^i + \beta\left[qV^i(N,\mu^H) + (1-q)V^i(N,\mu^L)\right] \tag{19.7}$$

for $i = p$ or r, where $V^i(N,\mu^H)$ refers to values in nondemocracy when $\mu_t = \mu^H$. In this expression, we already used the fact that when $\mu_t = \mu^L (= 1)$, the elite will choose no redistribution in a nondemocratic regime.

Next we note that since society starts in a nondemocracy, if a coup ever happens, then democratization must have previously arisen. Thus it is natural to assume that we are in the part of the parameter space where, if coups happen and the state moves to $\mu_t = \mu^H$, then following a coup, a re-democratization must take place, and therefore it must take place again when $\mu_t = \mu^H$. Therefore, we can impose $V^r(N,\mu^H) = V^r(D,\varphi^L)$.

The issue, however, is that once democracy has been reached again, the state (D,φ^H) will also be reached, and we have to make some conjectures about whether there will be another coup or not. However, the logic of dynamic programming dictates that what conjectures we make about future coups is not important. In other words, we can compute $V^r(N,\mu^L)$ and $V^r(D,\varphi^L)$ in two different ways, with identical implications for the threshold $\hat{\varphi}$. In the first, and possibly more natural way, we assume that once (D,φ^H) has been reached, there will be another coup. The second way looks only at a 'one-shot deviation,' and assumes that, even though the elite undertake a coup today, in the future they will never again do so, and democracy would survive even in the state (D, φ^H).

To illustrate the working of the model and this principle, we now derive the critical value $\hat{\varphi}$ using both approaches. Let us start with the first. In that case, the relevant values can be written as:

$$V^r(N, \mu^L) = y^r + \beta \left[qV^r(D, \varphi^L) + (1-q)V^r(N, \mu^L) \right], \tag{19.8}$$

and

$$V^r(D, \varphi^L) = y^r + \tau^p (\bar{y} - y^r) - C(\tau^p)\bar{y} + \beta \left[s \left(V^r(N, \mu^L) - \varphi y^r \right) + (1-s)V^r(D, \varphi^L) \right]. \tag{19.9}$$

Notice that equation (19.8) imposes that there will be a switch to democracy in the state (N, μ^H) for the reasons discussed already (we are in the part of the parameter space where there is an equilibrium switch to democracy). Equation (19.9), on the other hand, imposes that whenever state (D, φ^H) comes, there will be a coup, hence there is a switch to non-democracy, giving the value $V^r(N, \mu^L) - \varphi y^r$ to the elite, which takes into account the fact that they incur the cost of coup, φy^r. To solve for $V^r(N, \mu^L)$, we treat (19.8) and (19.9) as two equations in two unknowns, $V^r(N, \mu^L)$ and $V^r(D, \varphi^L)$ which we can solve for $V^r(N, \mu^L)$.

Substituting this into (19.4), using (19.6) and solving for φ gives the critical value as

$$\hat{\varphi} = \frac{1}{\theta} \left(\frac{\delta C(\tau^p) - \tau^p (\delta - \theta)}{1 - \beta(1-q)} \right). \tag{19.10}$$

The second method of looking at one shot deviations is often simpler. In this case, since a coup takes place only once and never again, when democracy is reached, there will never again be a coup despite the fact that the citizens always tax at the rate τ^p. This implies that in the equation (19.8), we have

$$V^r(D, \varphi^L) = \frac{y^r + \tau^p (\bar{y} - y^r) - C(\tau^p)\bar{y}}{1 - \beta}.$$

Substituting this into (19.8) we can solve for $V^r(N, \mu^L)$, which gives

$$V^r(N, \mu^L) = \frac{(1 - \beta(1-q))y^r + \beta q \left(\tau^p (\bar{y} - y^r) - C(\tau^p)\bar{y} \right)}{(1-\beta)(1 - \beta(1-q))}.$$

Substituting this into (19.4), using (19.6) and solving for φ gives the same critical value as in (19.10).

When $\varphi \geq \hat{\varphi}$, the coup threat does not play a role, and democracy is fully consolidated. The tax rate, $\tau^D = \tau^p$, is always determined by the usual trade-off for the median voter, balancing transfers against the deadweight losses of taxation. Observe that $d\hat{\varphi}/d\theta > 0$, which implies that a more unequal society is less likely to achieve a fully consolidated democracy.

This is intuitive since a greater level of inequality makes democracy less attractive for the elite and generalizes the results from the static model.

We can next determine the value of the cost of coup, φ^*, such that if $\varphi \geq \varphi^*$, the citizens can stop a coup by setting a low enough tax rate in the state (D, φ^H) (or conversely, when $\varphi < \varphi^*$, even a policy of setting $\tau^D = 0$ in state φ^H does not stop a coup). Since the lowest tax rate that the citizens can set is $\tau^D = 0$, φ^* is given by $V^r(N, \mu^L) - V^r(D, \varphi^H, \tau^D = 0) = \varphi^* y^r$.

Combining (19.1) and (19.2), and setting $V^r(\varphi^H) = V^r(D, \varphi^H, \tau^D = 0)$, we can calculate the value of always remaining in democracy for the elite. From this, we define

$$V^r(D, \varphi^H, \tau^D = 0) = \frac{y^r + \beta(1-s)\left(\tau^p\left(\bar{y} - y^r\right) - C(\tau^p)\bar{y}\right)}{1-\beta}$$

as the maximum value the median voter can credibly commit to give to a member of the elite under democracy.

To solve for $V^r(N, \mu^L)$, we use the one-shot deviation approach again. To do this we again work with (19.8) and substitute $V^r(N, \mu^H) = V^r(D, \varphi^L)$. As before we assume that a coup is only undertaken once and if there is redemocratization, there is never a coup again. However, the formula for $V^r(N, \mu^L)$ is different because when democracy is re-created after a coup this will a democracy in which the median voter sets $\tau^D = 0$ when $\varphi_t = \varphi^H$. Hence

$$V^r(D, \varphi^L) = \frac{y^r + (1-\beta s)\left(\tau^p\left(\bar{y} - y^r\right) - C(\tau^p)\bar{y}\right)}{1-\beta}.$$

Using this in (19.8) we find

$$V^r(N, \mu^L) = \frac{(1-\beta(1-q))y^r + \beta q(1-\beta s)\left(\tau^p\left(\bar{y} - y^r\right) - C(\tau^p)\bar{y}\right)}{(1-\beta)(1-\beta(1-q))}$$

Therefore $V^r(\varphi^H) = V^r(N, \mu^L) - \varphi y^r$ implies

$$\varphi^* = \frac{1}{\theta}\left(\frac{\beta(q+s-1)\left(\tau^p\left(\delta - \theta\right) - \delta C(\tau^p)\right)}{1-\beta(1-q)}\right), \tag{19.11}$$

where $\tau^p\left(\delta - \theta\right) - \delta C(\tau^p) < 0$ and $q + s - 1 < 0$, so φ^* is decreasing in q and s. If q is high, then a nondemocratic regime following a coup will be short lived because a revolutionary threat will reoccur quickly. This reduces the expected benefits from a coup. Similarly, if s is high, the coup constraint binds regularly, and because in this state the elite pay relatively low taxes, democracy is less costly to them. Also clearly, $\varphi^* < \hat{\varphi}$.

More important for the focus of here is that $d\varphi^*/d\theta > 0$: higher inequality decreases the threshold φ^* and makes a coup more likely because in an unequal society the elite lose more under democracy.

If $\varphi \geq \varphi^*$, then democracy is semi-consolidated: the citizens can avoid a coup by reducing the tax rate below τ^p in state (D, φ^H) and setting $\tau^D = \tilde{\tau} \leq \tau^p$ such that

$$V^r(N, \mu^L) - \varphi y^r = V^r(D, \varphi^H, \tau^D = \tilde{\tau}). \tag{19.12}$$

Although society always remains democratic, the threat of a coup is still important and influences taxes: the tax rate $\tilde{\tau}$ is less than τ^p, which the citizens would have set in the absence of this threat. Now $V^r(D, \varphi^H, \tau^D = \tilde{\tau})$ is solved for from the equations,

(19.13)

$$\begin{aligned} V^r(D, \varphi^H, \tau^D &= \tilde{\tau}) = \\ & y^r + \tilde{\tau}\left(\bar{y} - y^r\right) - C(\tilde{\tau})\bar{y} + \beta\left[sV^i(D, \varphi^H, \tau^D = \tilde{\tau}) + (1-s)V^i(D, \varphi^L)\right], \\ V^i(D, & \varphi^L) = \\ & y^r + \tau^p\left(\bar{y} - y^r\right) - C(\tau^p)\bar{y} + \beta\left[sV^i(D, \varphi^H, \tau^D = \tilde{\tau}) + (1-s)V^i(D, \varphi^L)\right]. \end{aligned}$$

which gives

$$\begin{aligned} V^r(D, \varphi^H, \tau^D &= \tilde{\tau}) \\ &= \frac{y^r + (1 - \beta(1-s))\left(\tilde{\tau}\left(\bar{y} - y^r\right) - C(\tilde{\tau})\bar{y}\right) + \beta(1-s)\left(\tau^p\left(\bar{y} - y^r\right) - C(\tau^p)\bar{y}\right)}{1-\beta} \end{aligned}$$

To calculate $V^r(N, \mu^L)$ we again use (19.8). The one-shot deviation approach implies that we should replace $V^r(D, \varphi^L)$ in (19.8) with the value of democracy to the elite when the citizens set the tax rate $\tau^D = \tilde{\tau}$ when $\varphi_t = \varphi^H$ and set $\tau^D = \tau^p$ when $\varphi_t = \varphi^L$. This value is just $V^r(D, \varphi^L)$ calculated from (19.13):

$$V^r(D, \varphi^L) = \frac{y^r + \beta s\left(\tilde{\tau}\left(\bar{y} - y^r\right) - C(\tilde{\tau})\bar{y}\right) + (1 - \beta s)\left(\tau^p\left(\bar{y} - y^r\right) - C(\tau^p)\bar{y}\right)}{1-\beta}$$

Using this to solve for $V^r(N, \mu^L)$ from (19.8) we find

$$V^r(N, \mu^L) = \frac{(1 - \beta(1-q))y^r + \beta^2 qs\left(\tilde{\tau}\left(\bar{y} - y^r\right) - C(\tilde{\tau})\bar{y}\right) + \beta q(1 - \beta s)\left(\tau^p\left(\bar{y} - y^r\right) - C(\tau^p)\bar{y}\right)}{(1-\beta)(1 - \beta(1-q))},$$

and substituting the results of these calculations into (19.12), we find that the tax rate $\tilde{\tau}$ is given implicitly by the equation

$$\begin{aligned} \varphi &= \frac{1}{\theta}\left(\frac{(\beta(1-q-s)-1)\left(\tilde{\tau}\left(\delta - \theta\right) - \delta C(\tilde{\tau})\right)}{1 - \beta(1-q)}\right. \\ &\quad \left. + \frac{\beta(q+s-1)\left(\tau^p\left(\delta - \theta\right) - \delta C(\tau^p)\right)}{1 - \beta(1-q)}\right). \end{aligned}$$

Implicit differentiation shows that $\tilde{\tau}$ is decreasing in θ: $d\tilde{\tau}/d\theta < 0$, so higher inequality reduces the tax rate that is required to prevent a coup.

If $\varphi < \varphi^*$, democracy is unconsolidated; even a strategy of setting $\tau^D = 0$ by the citizens will not prevent a coup. In this case, the society will revert back to a nondemocratic regime when $\varphi_t = \varphi$. The citizens would like to prevent such an outcome, and if they could, they would promise lower tax rates in the future. However, such promises are not credible because future tax rates are determined in future political equilibria, and once the threat of coup disappears, the tax rate will rise back to τ^p. Forward-looking elites, realizing this, prefer a coup, even though this is a costly outcome for the society as a whole.

This discussion leads to the concepts of fully and semi-consolidated democracies. In a fully-consolidated democracy there is never any threat of a coup, whereas in the semi-consolidated democracy, coup threats happen along the equilibrium past, but enough concessions are made so that coups are avoided.

Consider the state (N, μ^H). If the citizens did not attempt a revolution in this state, the elite would stay in power forever and set $\tau^N = 0$, so the citizens would receive utility equal to $y^p/(1-\beta)$. In contrast, with a revolution in state $\mu_t = \mu^H$, they would obtain $V^p(R, \mu^H) = (1-\mu)\bar{y}/(1-\delta)(1-\beta)$, the per-period return from revolution for the infinite future discounted to the present. Recall that revolution is an absorbing state in the sense that once a revolution occurs society stays like that forever and that only the value of μ at the time of the revolution matters, hence the per-period return is constant over time (and this also implies that in the state $\mu_t = \mu^L (= 1)$, a revolution will never occur). For the sake of reducing the number of cases we have to consider, we now impose $\theta > \mu$, which implies that when $\mu_t = \mu^H (= \mu)$, the revolution threat is binding.

In case of a revolution, the elite lose everything, i.e. $V^r(R, \mu^H) = 0$. They will therefore attempt to prevent it at all costs. They can do this in three different ways. First, they can democratize, $\phi = 1$, giving the citizens their return under democracy, $V^p(D, \varphi^L)$. Second, they can use repression giving a citizen the value $V^p(O, \mu^H \mid \kappa)$ which identical to the one we derived in the last chapter. Thirdly, they can choose to maintain political power, $\phi = 0$, but redistribute through taxation. In this case, the elite impose a tax rate τ^N and give the citizens a return $V^p(N, \mu^H, \tau^N)$ where

$$V^i(N, \mu^H, \tau^N) = y^i + \tau^N \left(\bar{y} - y^i\right) - C(\tau^N)\bar{y} + \beta \left[qV^i(N, \mu^H, \tau^N) + (1-q)V^i(N, \mu^L)\right] \tag{19.14}$$

So agent i receives income y^i from his own earnings and also a net income transfer $\tau^N \left(\bar{y} - y^i\right) - C(\tau^N)\bar{y}$. If next period, we are still in state $\mu_{t+1} = \mu^H$, then redistribution continues. But, if in the next period the economy switches to $\mu_{t+1} = \mu^L$, redistribution stops. This captures the notion that the elite cannot commit to future redistribution, unless the future also poses an effective revolution threat. Also note that $\tau^N \leq \tau^p$, that is, the elite will not tax themselves at a rate higher than τ^p, since this is the rate that maximizes redistribution to a citizen. If this tax rate is not sufficient to stop a revolution, then no tax rate $\tau^N \in [0,1]$ will do so.

With either democratization or redistribution by the elite, the citizens may still prefer a revolution. Thus, given the actions ϕ and τ^N of the elite, the value to the citizens in the state (N, μ^H) is

$$\begin{aligned} V^p(N,\mu^H) &= \omega V^p(O,\mu^H \mid \kappa) + \\ &(1-\omega) \max_{\rho\in\{0,1\}} \rho V^p(R,\mu^H) + (1-\rho)(\phi V^p(D,\varphi^L) + (1-\phi)V^p(N,\mu^H,\tau^N)). \end{aligned}$$

Combining (19.7) and (19.14), we calculate the maximum utility that can be given to the citizens without democratizing. This involves the elite setting the tax rate $\tau^N = \tau^p$ when there is a threat of revolution so that the continuation value for the citizens will be $V^p(N,\mu^H,\tau^N = \tau^p)$. This value satisfies:

$$V^p(N,\mu^H,\tau^N=\tau^p) = \frac{y^p + (1-\beta(1-q))\left(\tau^p\left(\bar{y}-y^p\right) - C(\tau^p)\bar{y}\right)}{1-\beta}, \tag{19.15}$$

which is of course the same as (18.6) derived in the last chapter. The citizens compare (19.15) to $V^p(R,\mu^H)$. This defines a critical value of μ^H,

$$\mu^* = \theta - (1-\beta(1-q))\left(\tau^p\left(\theta-\delta\right) - (1-\delta)C(\tau^p)\right) \tag{19.16}$$

such that $V^p(N,\mu^*,\tau^N=\tau^p) = V^p(R,\mu^*)$. For $0<\mu<\mu^*$, a revolution is so attractive for the citizens in state $\mu_t = \mu^H$ that even the maximum amount of redistribution by the elite cannot stop it. Democratization is therefore the only option left to the elite. Notice also that

$$\frac{d\mu^*}{d\theta} = 1 - (1-\beta(1-q))\frac{d\left(\tau^p\left(\theta-\delta\right) - (1-\delta)C(\tau^p)\right)}{d\theta} > 0,$$

so high inequality increases the revolution threshold because the citizens are worse off in a nondemocratic regime. Citizens are now willing to undertake revolutions when the cost of doing so is higher.

For $\mu \geq \mu^*$, democratization can be avoided by redistributing to the citizens in state (μ^H, N). In this case, the tax rate that the elite have to set in order to avoid revolution is $\tau^N = \hat{\tau}$, such that $V^p(N,\mu^H,\tau^N=\hat{\tau}) = V^p(R,\mu^H)$, which is decreasing in μ, and increasing in θ (i.e., increasing in the level of inequality).

Having determined the conditions under which a nondemocratic regime will be able to stay in power by making concessions and when a democracy is or is not consolidated, it remains to consider the implications of repression. Our assumptions about repression are identical to before so that the payoffs from repression as given by (18.10) as in the last section of the notes. There are again two situations to consider. If $\mu \geq \mu^*$ then a nondemocratic regime never needs to democratize in which case repression is used in equilibrium if it is cheaper than making policy concessions. The conditions under which this is so, and indeed the threshold level κ^* at which the elite are indifferent between promising redistribution at the tax rate $\hat{\tau}$ and repression, are identical to those we derived previously. In particular, κ^* is again given by (18.12). If $\mu < \mu^*$, then the elite cannot use concessions to stay in power and they compare the cost of repression to the cost of democracy. In the previous analysis the cost of democracy was uniquely defined because we assumed that democracy was fully consolidated. However, this is not the case now and the cost of democracy to the elite, and therefore the attractiveness of repression, depends on the nature of democracy.

If $\varphi \geq \hat{\varphi}$ so that democracy is fully consolidated, then the threshold at which the elite are just indifferent between repression and democratization is $\bar{\kappa}$ given by (18.13), in the previous section. If $\varphi \in [\varphi^*, \hat{\varphi})$ then democracy is partially consolidated and when there is the threat of a coup the tax rate is cut. In this case we can define a threshold level $\kappa(\varphi)$ such that the elite are just indifferent between repressing and creating a semi-consolidated democracy. To see the formula for this first note that the value of repression is $V^r(O, \mu^H \mid \kappa)$ given by (18.11) in the previous section. The value of being in an unconsolidated democracy is $V^r(D, \varphi^L)$ which satisfies (19.8) and (19.9). Thus $\kappa(\varphi)$ is such that $V^r(O, \mu^H \mid \kappa(\varphi)) = V^r(D, \varphi^L)$. Note that the higher is φ, the more costly a coup, the higher the tax in this state and the greater the cost of creating democracy to the elite. Hence $\kappa(\varphi)$ is a strictly increasing function of φ since as φ increases, the burden of democracy increases for the elite and they are more inclined to use repression. Finally if $\varphi < \varphi^*$ democracy is unconsolidated and we can define a threshold $\breve{\kappa}$ such that elite are just indifferent between repressing and creating an unconsolidated democracy.

We restrict attention to the area of the parameter space where democratization prevents a revolution, that is $V^p(D, \varphi^L) \geq V^p(R, \mu^H)$. Since democracy is not necessarily an absorbing state, the value function $V^p(D, \varphi^L)$ takes into account the future possibility of coups. The value to the citizens of a semi-consolidated democracy is higher than that of a democracy subject to coups, so it suffices to ensure that the value to the citizens of an unconsolidated democracy is greater than $V^r(R, \mu^H)$. To derive a formula for the value of a citizen of an unconsolidated democracy, we use (19.1) and (19.7) with $V^p(N, \mu^H) = V^p(D, \varphi^L)$ and $V^p(\varphi^H) = V^p(N, \mu^L)$, giving

$$\begin{aligned}
V^p(N, \mu^L) &= y^p + \beta\left[qV^p(D, \varphi^L) + (1-q)V^p(N, \mu^L)\right], \\
V^p(D, \varphi^L) &= y^p + \tau^p(\bar{y} - y^p) - C(\tau^p)\bar{y} + \beta\left[s\left(V^p(N, \mu^L) - \varphi y^p\right) + (1-s)V^p(D, \varphi^L)\right],
\end{aligned}$$

which are the same as (19.8) and (19.9) from the point of view of the citizens. Solving for $V^p(D,\varphi^L)$ we find

$$V^p(D,\varphi^L) = \frac{y^p\left(\left(1-\varphi\beta s\right)\left(1-\beta(1-q)\right)+\beta s\right)+\left(1-\beta(1-q)\right)\left(\tau^p\left(\bar{y}-y^p\right)-C(\tau^p)\bar{y}\right)}{\left(1-\beta(1-s)\right)\left(1-\beta(1-q)\right)-\beta^2 sq}$$

and the condition $V^p(D,\varphi^L) \geq V^p(R,\mu^H)$ is therefore equivalent to:

$$\frac{\left(1-\beta+\beta(q+s)\right)(1-\theta)(1-\varphi\beta s)+\left(1-\beta(1-q)\right)\left(\tau^p\left(\theta-\delta\right)-(1-\delta)C(\tau^p)\right)}{1-\beta+\beta(q+s)} \geq 1-\mu$$

which is a condition on the parameters that we shall simply assume holds. As with the corresponding condition in the previous section of the notes, this will hold when democracy is sufficiently redistributive. Note that this leads to an interesting trade-off: a highly redistributive democracy leads to political instability, but if the potential for redistribution is too limited, democratization does not prevent revolution.

PROPOSITION 19.1. *There is a unique Markov perfect equilibrium* $\{\tilde{\sigma}^r, \tilde{\sigma}^p\}$ *in the game described in* $G^\infty(\beta)$. *Let* $\hat{\varphi}$, φ^*, κ^*, $\bar{\kappa}$ *and* $\breve{\kappa}$ *be as defined above. Then in this equilibrium:*

- *If* $\mu \geq \mu^*$, *the society remains nondemocratic. When* $\mu_t = \mu^L$, $\tau^N = \tau^r$, *and there is no redistribution. If* $\kappa < \kappa^*$, *then when* $\mu_t = \mu^H$, *the rich use repression. If* $\kappa \geq \kappa^*$, *then when* $\mu_t = \mu^H$, $\tau^N = \hat{\tau}$ *where* $V^p(N,\mu^H,\tau^N=\hat{\tau}) = V^p(R,\mu^H)$.
- *If* $\mu < \mu^*$, *then we have:*
  - *If* $\varphi \geq \hat{\varphi}$ *and* $\kappa \geq \bar{\kappa}$, *we are in a fully consolidated democracy. The society switches to democracy the first time* $\mu_t = \mu^H$, *and remains democratic thereafter, and taxes are always given by* $\tau^D = \tau^p$.
  - *If* $\varphi^* \leq \varphi < \hat{\varphi}$ *and* $\kappa \geq \kappa(\varphi)$, *we are in a semi-consolidated democracy. The society switches to democracy the first time* $\mu_t = \mu^H$, *and remains democratic thereafter. When* $\varphi_t = \varphi^L$, $\tau^D = \tau^p$. *When* $\varphi_t = \varphi^H$, *democracy sets the tax rate* $\tau^D = \tilde{\tau} < \tau^p$ *such that* $V^r(N,\mu^L) - \varphi\bar{y} = V^r(\varphi^H, D, \tau^D = \tilde{\tau})$.
  - *If* $\varphi < \varphi^*$ *and* $\kappa \geq \breve{\kappa}$, *we are in an unconsolidated democracy. The society continuously switches regimes. In a nondemocratic regime, when* $\mu_t = \mu^L$, *the elite set* $\tau^N = \tau^r$, *and when* $\mu_t = \mu^H$, *they democratize. In a democracy, when* $\varphi_t = \varphi^L$, $\tau^D = \tau^p$, *and when* $\varphi_t = \varphi^H$, *there is a coup.*
  - *If* $\varphi \geq \hat{\varphi}$ *and* $\kappa < \bar{\kappa}$, *or* $\varphi^* \leq \varphi < \hat{\varphi}$ *and* $\kappa < \kappa(\varphi)$, *or if* $\varphi < \varphi^*$ *and* $\kappa < \breve{\kappa}$, *when* $\mu_t = \mu^L$, $\tau^N = \tau^r$, *and there is no redistribution and when* $\mu_t = \mu^H$, *the elite use repression to stay in power.*

19.2. Discussion

The main message from this proposition is that, as before, democracy arises because the elite cannot commit to future policies while they maintain a monopoly of political power. However, once created democracy is not necessarily consolidated. Nevertheless, despite the fact that rational individuals anticipate that in the future coups may occur against democracy, the creation of democracy may nevertheless stop a revolution in the same way as before. This is because to mount a coup the elite must have de facto power and whether or not they will have this power in the future is uncertain. This being the case the citizens value the creation of democracy which moves de jure power in their direction, even when they understand that democracy will not be permanent.

We can now discuss the conditions in the proposition in more detail. In the first type of equilibrium where $\mu \geq \mu^*$, a revolution is sufficiently costly that given the amount of inequality and the value of q, the elite can avoid it by redistributing. Therefore, in state $\mu_t = \mu^L$, the elite set their preferred tax rate of zero, i.e., $\tau^N = \tau^r = 0$, while in state μ^H, if repression is sufficiently costly, they redistribute by setting the tax rate $\tau^N = \hat{\tau}$, which is just enough to stop a revolution. If repression is relatively cheap, however, the elite respond to the threat of revolution by repressing the citizens. In this equilibrium, there is never democratization and the amount of redistribution is relatively limited and zero if the elite choose repression. If redistribution takes place inequality nonetheless increases the level of redistribution in this regime because the elite are forced to choose higher taxes to prevent a revolution in the state (N, μ^H).

Now consider what happens when $\mu < \mu^*$. When the society transits into state μ^H, the elite can no longer maintain their political power via redistribution, and must either repress or democratize. There are four types of equilibria depending on the values of φ and κ. The first possibility is that $\varphi \geq \hat{\varphi}$ and $\kappa \geq \bar{\kappa}$. Democracy, once created, is fully consolidated and repression is sufficiently costly that democracy will be created even though the elite know that the citizens will always set $\tau^D = \tau^p$ from then on. In this type of society, the amount of redistribution is at its highest level, there is very little or no fiscal volatility, and the threat of a coup plays no role once the society becomes democratic. We interpret this case as similar to the situation in most OECD countries. It is more likely to arise when θ is low, that is when the society is fairly equal as long as $\theta > \mu$ so that the revolution constraint binds.

The second possibility is that $\varphi^* \leq \varphi < \hat{\varphi}$, and $\kappa \geq \kappa(\varphi)$. Then democracy is semi-consolidated and only survives by making concessions in some states. In particular, if in democracy the citizens were to set a tax rate τ^p in the state (D, φ^H), a coup would occur.

The citizens avoid this by setting a lower tax $\tau^D = \tilde{\tau}$ in state (D, φ^H), which is just sufficient to dissuade the elite from mounting a coup. Although the society always remains democratic, it is in some sense "under the shadow of a coup", as the threat of a coup keeps overall redistribution below the level of a fully consolidated democracy.

The third type of equilibrium involves $\varphi < \varphi^*$ and $\kappa \geq \check{\kappa}$ so that democracy is unconsolidated: when the state moves to φ^H, a coup is relatively attractive for the elite, and cannot be halted by reducing taxes. As a result, the economy will fluctuate randomly between democracy and nondemocracy. More specifically, when repression is not attractive, the economy starts with the elite in power and they set $\tau^N = \tau^r$. Whenever the state moves to μ^H, they democratize after which the citizens set $\tau^D = \tau^p$. But as soon as the state goes from (D, φ^L) to (D, φ^H), they mount a coup, regain political power, and set $\tau^N = 0$. The variability of policy is therefore highest in this equilibrium, and the amount of redistribution is less than in cases 2 and 3, but more than in case 1. Higher inequality increases redistribution in this regime because it increases the tax rate when there is democracy, while there is never any redistribution in nondemocracy. Notice that in this case, when the citizens are in power, they set the maximum tax rate, fully anticipating that redistribution will eventually come to an end as a result of a coup. This result may help to explain the existence of highly redistributive, but relatively short-lived, populist regimes of Latin America (see, for example, Kaufman and Stallings, 1991, on populism).

The final type of equilibrium involves repression by the elite to maintain the nondemocratic regime. This arises in a variety of circumstances if the cost of repression is sufficiently low. Note that since $\bar{\kappa} > \kappa(\varphi) > \check{\kappa}$, repression is most attractive for the elite when they anticipate that they will have to create a fully consolidated democracy. Interestingly, therefore, our analysis suggests that it is more likely that an unconsolidated democracy will be created than a semi or a fully consolidated one.

As with democratizations, coups happen only in the high state, which can be interpreted as a relatively unlikely or unusual state. In this context, one appealing interpretation is that the high state corresponding to periods of recession or economic crises. During such crises, undertaking a coup may be less costly because society is in disarray and a proportional loss of income or output due to turbulence and political instability may be less severe because output is already low. This interpretation, which suggests that regime changes, and in particular coups, are more likely during recessionary periods, is in line with the broad patterns in the data. Many coups happen during recessions or during periods of economic difficulties, such as those in Brazil in 1964, Chile in 1973 and Argentina in 1976.

There are four other conclusions to be drawn from this analysis. The first links inequality to regime changes. An increase in θ increases μ^*, φ^*, $\hat{\varphi}$, κ^*, $\kappa(\varphi)$ and $\breve{\kappa}$. Thus higher inequality makes revolutions, coups and repression all more attractive. As discussed above, there is an inverted-U shaped relationship between inequality and democratization. Highly equal or highly unequal societies are unlikely to democratize. Rather, it is societies at intermediate levels of inequality in which we will observe democratization. The model of this chapter predicts that having democratized, democracy is also more likely to consolidate in more equal societies. Thus we might expect to see very equal societies, such as Singapore remain non-democratic. Societies with higher levels of inequality will democratize and become (semi- or fully) consolidated democracies, while societies with greater inequality may democratize but be unconsolidated. These two cases may fit the historical evolution of Britain and Argentina. Finally a very inegalitarian society may never democratize in the first place, which fits the South African experience. Of course these statements apply holding other things equal.

The second conclusion pertains to the link between inequality and redistribution. To see this, fix the cost of a coup φ, and define $\theta^H > \theta^L$ such that $\varphi = \hat{\varphi}(\theta^L)$ and $\varphi = \varphi^*(\theta^H)$. Moreover, suppose that $\mu < \mu^*(\theta^H)$. When $\theta < \theta^L$, $\varphi \geq \hat{\varphi}(\theta)$, so inequality is sufficiently low that democracy is fully consolidated. Now consider an increase in inequality (an increase in θ). This will increase redistribution at first as in the standard models of voting over redistribution, since $d\tau^p/d\theta > 0$. However, as θ rises above θ^L, democracy is no longer fully consolidated, but semi-consolidated, i.e., $\varphi \in [\varphi^*(\theta), \hat{\varphi}(\theta))$. In this case, the citizens are forced to reduce taxes from τ^p to $\tilde{\tau}$ in the state (D, φ^H), so overall redistribution falls. In fact, in a semi-consolidated democracy, the relationship between inequality and taxation is ambiguous. The average tax rate is $\tau^a = (1-s)\tau^p + s\tilde{\tau}$. τ^p is increasing in inequality while $\tilde{\tau}$ is decreasing. If the cost of taxation $C(\tau)$ is highly convex, then the second effect dominates and the average tax rate falls as inequality rises. Intuitively, higher inequality makes a coup more attractive for the elite, so to prevent the coup, the citizens have to reduce the tax rate substantially in the state φ, leading to lower redistributive taxation on average. As inequality increases further, we have $\theta > \theta^H$ so $\varphi < \varphi^*$, and democracy is now unconsolidated with lower overall redistribution than both in fully and semi-consolidated democracies. Therefore, there is a nonmonotonic relationship between inequality and redistribution, with societies at intermediate levels of inequality redistributing more than both very equal and very unequal societies.

The third implication of our analysis is related to fiscal volatility. The relationship between fiscal volatility and inequality is likely to be increasing. Within each regime, higher inequality leads to more variability. Moreover, higher inequality makes unconsolidated democracy, which has the highest amount of fiscal variability, more likely. This may explain why fiscal policy has been much more volatile in Latin America than in the OECD.

The final implication of our analysis is that the costs of redistribution will also have an impact on the equilibrium political system. Suppose that the cost of taxation becomes less convex, so that $C(\tau^p)$ is unchanged, but $C'(\tau^p)$ decreases. Since deadweight losses from taxation are now lower, the median voter will choose a higher level of taxation. However, as τ^p increases, so will $-(\tau^p(\bar{y}-y^r)-C(\tau^p)\bar{y})$, so democracy becomes more costly to the elite, and hence less likely to be consolidated. This implies that in societies where taxation creates less economic distortions, for example in societies where a large fraction of the GDP is generated from natural resources, democracies may be harder to consolidate. This result has an obvious parallel to the result discussed below, that targeted transfers also make coups more likely. These two results together imply that a more efficient or flexible fiscal system may not always be preferable once its implications for the political equilibrium are taken into account.

19.3. References

(1) Acemoglu, Daron and James Robinson (2006) "Chapter 7: Coups and Consolidation" in Acemoglu and Robinson *Economic Origins of Dictatorship and Democracy*, Cambridge University Press.

(2) ** Acemoglu, Daron and James A. Robinson (2001) "A Theory of Political Transitions," *American Economic Review*, Vol. 91, pages 938-963.

CHAPTER 20

Economic Structure and Democracy

The analysis so far showed how political institutions played a key role of allocating de jure political power in dynamic environments and the need to regulate the future allocation of de jure power shapes equilibrium political institutions and whether they endure. However any theory of institutions (political or economic) is most useful for its comparative static results. So far the comparative static results have been limited to inequality and cost of repression (and coups and revolution). More interesting comparative static results relate to the structure of the economy. For example, our societies where land is the major asset more or less like you to be democratic? Is industrialization like you to encourage democratization ? What is the relationship between human capital and democracy? The classic by Barrington Moore (1966) was explicitly concerned with these questions. In fact, Moore argued that the emergence of industrial bourgeoisie was essential for democracy, and in societies dominated by landowners either fascist or communist dictatorships were more likely to emerge. His analysis was purely historical and sociological, however, and did not clarify the mechanisms linking either the industrial bourgeoisie or the landowners with the structure of political institutions. Indeed, it is hard to tell from his discussion why one group favors one set of institutions rather than another. Our purpose in this chapter is to highlight some mechanisms that may be important in thinking about these questions. To do this, we now present a simple model of economic structure embedded in the political environment analyzed in the previous two sections of the notes.

20.1. A Simple Model of Economic Structure

Let us now introduce an explicit economic structure which will enable us to endogenize income distribution and talk about the political implications of different factor endowments. We want this structure to include labor (as the source of income for the citizens), physical capital and land. Consider a fully competitive economy with unique final consumption good, produced via the aggregate production function

$$Y = F(K, L, N)$$

where K is the capital stock, L is total amount of productive land and N is the labor force. Y is aggregate output which will be the physical quantity that people will have to consume. All of these factors are fully employed, and we assume that the production function F exhibits constant returns to scale, so that when all three factors are doubled, total output will be doubled. Constant returns to scale is important because it implies that all revenues from production are distributed as incomes to the factors of production, capital, land and labor. Fully competitive markets imply that all factors of production will be paid their marginal products. Holding institutions constant, inequality will result because these marginal products differ, and there are different scarcities for different factors.

The simplest way to provide a microfoundations for the framework we have used so far is to assume that the aggregate production function takes the special Cobb-Douglas form

$$Y = (K + \sigma L)^{\theta} N^{1-\theta} \tag{20.1}$$

where $0 < \theta < 1$, and $\sigma > 0$. As will become clear later when we calculate the distribution of income in this model, the choice of $1 - \theta$ as the power to which N is raised is chosen deliberately to relate this model to the ones we have used so far in the book.

There are two features implicit in this function. First, there is a limited amount of substitution between labor and the other factors of production (more precisely, the elasticity of substitution between labor and the other factors is exactly equal to 1). Second, there is a much higher level of substitution between capital and land. Both of these assumptions are plausible. For instance, they imply that the share of labor in national income is constant when income grows as a result of capital accumulation, while the share of land falls and the share of capital rises. This is roughly consistent with empirical evidence.

Let us see this in greater detail. First, let us assume, like before, that there are $1 - \delta$ citizens and now these agents correspond to wage earners. Hence, $N = 1 - \delta$, and we can write the production function as:

$$Y = (K + \sigma L)^{\theta} (1 - \delta)^{1-\theta}.$$

Moreover, the remaining δ agents, who constitute the elite, do not own any labor, and each of them holds a fraction δ of total capital stock, K, and a fraction δ of total land stock, L.

We assume that the final good Y is the numeraire (i.e., its price is normalized to 1). Throughout, all other prices are therefore relative to the price of the final good. Exploiting the fact that in competitive labor markets all factors of production will be paid their full marginal products, we have the following expressions for factor prices

$$(20.2) \qquad \begin{aligned} w &= (1-\theta)\left(\frac{K+\sigma L}{1-\delta}\right)^{\theta}, \\ r &= \theta\left(\frac{K+\sigma L}{1-\delta}\right)^{\theta-1}, \\ v &= \sigma\theta\left(\frac{K+\sigma L}{1-\delta}\right)^{\theta-1}. \end{aligned}$$

Here w denotes the wage rate, r the return to capital, and v is the rental rate of land. These prices are all 'real' or relative prices because they are measured in terms of the final good.

The shares of national income accruing to three factors are given as

$$(20.3) \qquad \begin{aligned} s_N &\equiv \frac{wN}{Y} = 1-\theta \\ s_K &\equiv \frac{rK}{Y} = \theta\frac{K}{K+\sigma L} \\ s_L &\equiv \frac{vL}{Y} = \theta\frac{\sigma L}{K+\sigma L} \end{aligned}$$

The interesting thing here is that the share of national income accruing to labor is a constant equal to $1-\theta$. This stems directly from the functional form of the Cobb-Douglas production function (20.1). Note for instance that even if capital accumulates, and from (20.2) real wages increase, the share of labor in national income is nevertheless constant. At the same time, the share of capital in national income increases, and that of land declines.

Now total income is $(K+\sigma L)^{\theta}(1-\delta)^{1-\theta}$ and since total population is 1, this is also average income $\bar{y}$. Hence:

$$(20.4) \qquad \bar{y} = (K+\sigma L)^{\theta}(1-\delta)^{1-\theta}$$

Exploiting the fact that citizens only have labor income, we can derive an expression for the income of a citizen, denoted y^p:

$$(20.5) \qquad \begin{aligned} y^p &= (1-\theta)\left(\frac{K+\sigma L}{1-\delta}\right)^{\theta} = \frac{(1-\theta)(K+\sigma L)^{\theta}(1-\delta)^{1-\theta}}{1-\delta} \\ &= \left(\frac{1-\theta}{1-\delta}\right)\bar{y}, \end{aligned}$$

which is the exact expression for y^p we have used throughout the book.

Recall that, for now, we are assuming that all members of the elite are homogeneous and own both capital and land. Therefore, we have

$$(20.6) \qquad y^r = \frac{rK+vL}{\delta} = \frac{\theta}{\delta}(K+\sigma L)^{\theta}(1-\delta)^{1-\theta} = \frac{\theta}{\delta}\bar{y}$$

as the expression which gives the income of a member of the elite.

We assume that the parameters are such that average incomes are less than the incomes of the rich, or in other words: $\delta < \theta$, which is identical to the assumption we made in the model where incomes are exogenous.

20.2. Political Conflict

Let us now link the previous analysis of political conflict and policy determination in democracy to this setup. As before, all individuals have utility functions that are linear in consumption (or income, since all income is directly consumed). They also discount the future at the rate β. We continue to assume that there are two policy instruments, a tax rate which is proportional to income and a lump-sum transfer which all agents receive. As before, it is costly to redistribute income. Although we now have a model with a richer set of underlying institutions, we assume these to be exogenous in the analysis now.

The utility of an individual i is now, $(1-\tau)y^i + T$ or $i = p, r$ where the government budget constraint again implies that

$$T = \delta\tau y^r + (1-\delta)\tau y^p - C(\tau)\bar{y} = (\tau - C(\tau))\,\bar{y}.$$

Incorporating the costs of taxation, we have the indirect utility of a poor agent as $V(y^p \mid \tau) = (1-\tau)y^p + (\tau - C(\tau))\,\bar{y}$. The first order condition of maximizing this indirect utility is identical to that which we have derived before and since we know that preferences are single peaked, we can apply the median voter theorem to determine the (unconstrained) democratic equilibrium tax rate again denoted τ^p. Using the fact that the incomes of the poor are given (20.5) and average income is given by (20.4), this equilibrium tax, τ^p, is identical to above, namely (17.3).

20.3. Capital, Land and the Transition to Democracy

Let us now use this framework how the structure of the economy influences the costs of repression. We assume that repression creates costs for the elites depending on the sources of their incomes, in particular, whether they rely more on income from capital or income from land. It is plausible to presume that the disruption associated with putting down the threat of revolution and an uprising by the citizens will be much more costly for industrialists, for factories, for commerce, than for land and landowners. As a result, when land is very important for the elites, they will be more willing to bear the cost of repression to avoid democratization. In a society where income from capital becomes more important than income from land, it is more likely that the potential costs of repression exceed those of democracy, and the elite prefer to give democracy to the dissatisfied citizens rather than use force against them.

The underlying economic model is the same as the one above. The elite own capital and land. Moreover, all members of the elite have identical endowments, so that there is no heterogeneity among the elite (we return for the distinction between industrialists and landowners below). As before, the payoff to the citizens from a revolution is $V^p(R,\mu) = (1-\mu)\bar{y}/(1-\delta)(1-\beta)$, while the elite always have $V^r(R,\mu) = 0$. We consider a situation where democracy once created is consolidated so there are no coups.

Let us assume that if the elite choose to repress in order to avoid a revolution and democratization they will lose a fraction κ_K of the capital stock and a fraction κ_L of land. Moreover, we assumed

$$\kappa_K \geq \kappa_L.$$

To reduce notation we set $\kappa_L = \kappa$ and $\kappa_K = \varrho\kappa$ where $\varrho \geq 1$.

With similar steps to before, the values to the citizens and to the elite if there is democracy are given by

(20.7)

$$\begin{aligned}
V^p(D) &= \frac{1}{1-\beta}\left[w + \tau^p(\bar{y}-w) - C(\tau^p)\bar{y}\right] \\
&= \frac{1}{1-\beta}\left[\frac{1}{1-\delta}\left(1-\theta+\tau^p(\theta-\delta)-(1-\delta)\,C(\tau^p)\right)(K+\sigma L)^{\theta}(1-\delta)^{1-\theta}\right] \\
V^r(D) &= \frac{1}{1-\beta}\left[\frac{rK+vL}{\delta} + \tau^p\left(\bar{y}-\frac{rK+vL}{\delta}\right) - C(\tau^p)\bar{y}\right] \\
&= \frac{1}{1-\beta}\left[\frac{1}{\delta}\left(\theta+\tau^p(\delta-\theta)-\delta C(\tau^p)\right)(K+\sigma L)^{\theta}(1-\delta)^{1-\theta}\right],
\end{aligned}$$

where the factor prices w, r and v are given by (20.2), and the most preferred tax rate of the citizens is τ^p. These expressions take into account that once there is democratization, the citizens set their most preferred tax rate unconstrained.

If, on the other hand, the elite choose repression, the payoffs are

$$\begin{aligned}
V^p(O\,|\kappa) &= \frac{(1-\theta)}{(1-\beta)}\left(\frac{(1-\varrho\kappa)K+\sigma(1-\kappa)L}{1-\delta}\right)^{\theta}, \\
V^r(O\,|\kappa) &= \frac{\theta}{\delta}\left((1-\varrho\kappa)K+\sigma(1-\kappa)L\right)^{\theta}(1-\delta)^{1-\theta}/(1-\beta).
\end{aligned} \tag{20.8}$$

Finally, the elite could offer redistribution under the existing regime, without democratizing and without resorting to repression. The best they can do in this case is to offer redistribution at the favorite tax rate of the citizens, τ^p, given by (17.3), and in this case the values are

(20.9)

$$V^p(N,\tau^N = \tau^p) = \frac{(1-\theta+\beta(1-q)\left(\tau^p(\theta-\delta)-(1-\delta)\,C(\tau^p)\right))\,(K+\sigma L)^\theta\,(1-\delta)^{1-\theta}}{(1-\beta)\,(1-\delta)},$$

$$V^r(N,\tau^N = \tau^p) = \frac{(\theta+\beta(1-q)\left(\tau^p\,(\delta-\theta)-\delta C(\tau^p)\right))\,(K+\sigma L)^\theta\,(1-\delta)^{1-\theta}}{\delta\,(1-\beta)}$$

which incorporates the fact that this promise will be realized only with probability p.

As before, if $\theta \leq \mu$, the revolution threat is absent. The more interesting case for the discussion here is the one where $\theta > \mu$, which for the sake of simplicity we assume to be the case. The promise to redistribute will prevent a revolution if we have that $V^p(N,\tau^N = \tau^p) \geq V^p(R,\mu)$. By the same arguments as before, this is equivalent to $\mu \geq \mu^*$ where μ^* is defined as above.

If $\mu < \mu^*$, the elite cannot prevent a revolution by promising redistribution, so they have to resort either to democratization or to repression. We assume as usual that $V^p(D) \geq V^p(R,\mu)$ so that democratization prevents revolution and the formula for this is identical to (18.7).

When will the elite prefer repression? This depends on whether $\mu \geq \mu^*$ or not. When $\mu \geq \mu^*$, the relevant comparison is between redistribution and repression since, for the elite, redistribution is always preferable to democratization when it is feasible. The case that is more interesting for us is when $\mu < \mu^*$, so that there is a trade-off between repression and democratization. In this case, the elite simply compare $V^r(D)$ and $V^r(O\,|\kappa)$ the jobas given by (20.7) and (20.8). It is clear that they will prefer repression if $V^r(D) < V^r(O\,|\kappa)$ or if

$$\theta + \tau^p\,(\delta-\theta) - \delta C(\tau^p) < \theta\left(\frac{(1-\varrho\kappa)K+\sigma(1-\kappa)L}{K+\sigma L}\right)^\theta. \tag{20.10}$$

It is useful to re-write (20.10) in terms of the capital to land ratio, $k = K/L$. This gives

$$\theta + \tau^p\,(\delta-\theta) - \delta C(\tau^p) < \theta\left(\frac{(1-\varrho\kappa)k+\sigma(1-\kappa)}{k+\sigma}\right)^\theta, \tag{20.11}$$

as the condition under which repression takes place. We shall say that when k is higher, the economy is more 'capital intensive' whereas low values of k correspond to relatively 'land intensive' societies. Condition (20.11) makes it clear that capital intensity of a society will be a crucial determinant of whether repression will be attractive for the elite or not.

Proposition 20.1. *Assume that (18.7) holds, $\theta > \mu$, and $\mu < \mu^*$. Then we have that*

- *If (20.11) does not hold, then democratization happens as a credible commitment to future redistribution by the elite.*

- *If (20.11) holds, then the elite use repression to prevent revolution.*

Whether repression is desirable, (20.11), holds or not depends on how capital intensive the economy is (i.e., on the level of k). The easiest way to see this is to consider the case where $\varrho = 1$ so that the costs of repression fall equally on capital and land.

Proposition 20.2. *Consider the above game with $\varrho = 1$. Then (20.11) is independent of k, so the political equilibrium is unaffected by the capital intensity of the economy.*

In contrast, if $\varrho > 1$, it is straightforward to verify that (20.11) is less likely to hold as k increases. Therefore, let us define k^* such that

$$\theta + \tau^p (\delta - \theta) - \delta C(\tau^p) = \theta \left(\frac{(1 - \rho\kappa)k^* + \sigma(1 - \kappa)}{k^* + \sigma} \right)^\theta . \tag{20.12}$$

Proposition 20.3. *Consider a society described by the above game with $\varrho > 1$ and define k^* by (20.12). Then in the unique subgame perfect equilibrium we have that: if $k < k^*$, then the elite will meet the threat of revolution with repression, and if $k \geq k^*$, they will democratize in response to the threat to revolution.*

This result applies because the use of force by the elite is more costly in such a society compared to a land intensive society, or expressed differently, capital investments make the elites more pro-democractic than land holdings (and in a sense to made precise below, industrialists are more pro-democratic than landowners).

20.4. Costs of Coup on Capital and Land

We now move to extend these ideas to coups. Because of the parallels between using repression and mounting coups, there also appear to be natural reasons for these costs to also depend on how capital intensive the economy is. In particular, suppose that during a coup a certain fraction of the productive assets of the economy get destroyed. Let the fraction of physical capital that is destroyed be φ_K and that of land be φ_L if a coup is undertaken. It is natural to think that

$$\varphi_K \geq \varphi_L,$$

in other words, that the disruptions associated with a coup are more destructive to capital than to land. The reasons for why this is plausible are similar to those we discussed above. Coups and the associated turbulence and disruption will lead to the breakdown of complex economic relations. These are much more important for capitalist production than agrarian production. This is natural, since there is less concern about the quality of products in agriculture than in manufacturing. Moreover, the importance of complex relationships between buyer and supplier networks, and of investments in skills and in relationship-specific capital is far greater in more industrialized activities. Therefore, land will be hurt less as a result of a coup than capital.

Let $\varphi_L = \varphi$ and let $\varphi_K = \xi\varphi$ where $\xi \geq 1$. Given this assumption, we can write the incomes after coups as:

$$\tilde{y}^p = (1-\theta)\left(\frac{(1-\xi\varphi)K + \sigma(1-\varphi)L}{1-\delta}\right)^{\theta} \tag{20.13}$$

$$\tilde{y}^r = \frac{\theta}{\delta}\left((1-\xi\varphi)K + \sigma(1-\varphi)L\right)^{\theta}(1-\delta)^{1-\theta} \tag{20.14}$$

Clearly, both of these expressions are less than the corresponding ones before the coup, (20.5) and (20.6), since the disruptions associated with a coup will typically lead to the destruction of a certain fraction of the productive assets of the economy.

Armed with this specification of the costs of coups we can now analyze the impact of economic structure on coups and democratic consolidation.

Whether or not the elite wish to mount a coup will depend on the continuation value in democracy and nondemocracy. Faced with the threat of a coup, the median voter will wish to make a concession to avoid a coup, i.e., set $\tau^D < \tau^p$. After this, the elite decide whether to undertake the coup. If they do so, society switches to nondemocracy, and the elite set the tax rate. Naturally, after a successful coup, they will choose their most preferred tax rate, $\tau^N = 0$. As a result, the game ends with respective payoffs for the citizens and the elite, $V^p(C,\varphi) = \tilde{y}^p/(1-\beta)$ and $V^r(C,\varphi) = \tilde{y}^r/(1-\beta)$, where $\tilde{y}^p$ and $\tilde{y}^r$ are given by (20.13) and (20.14).

The analysis of the political economy's again similar, and implies that the coup constraint can be expressed as

$$\theta + \tau^p(\delta-\theta) - \delta C(\tau^p) < \theta\left(\frac{(1-\xi\varphi)k + \sigma(1-\varphi)}{k+\sigma}\right)^{\theta}. \tag{20.15}$$

where we again write the expression in terms of the capital intensity of the economy $k = K/L$. When this constraint does not bind coups are sufficiently costly that the elite never find a coup profitable —democracy is *fully consolidated*. (20.15) is fairly intuitive and responds to changes in parameters in the way we should expect. For example, a greater democratic tax rate, τ^p, makes it more likely to hold, since only the left-hand side depends on τ^p and is decreasing in it, and greater level of φ make it less likely to hold, since a greater fraction of the assets of the elite will be destroyed in the process of a coup.

In contrast, when this constraint binds, the democratic regime is not fully consolidated: if the citizens do not deviate from their most preferred to tax rate, there will be a coup along the equilibrium path. We can therefore define a critical value of the fraction of assets destroyed in a coup, denoted φ^*, such that when $\varphi < \varphi^*$, (a coup is not too costly) the promise of limited redistribution by the citizens is not sufficient to dissuade the elite from a coup. Of course, the most attractive promise that the citizens can make to the elite is to stop redistribution away from them totally, i.e., $\tau^D = 0$. Therefore, we must have that at φ^*, $V^r(D, \tau^D = 0) = V^r(C, \varphi^*)$, or:

$$\varphi^* = \left[1 - \left(1 + \frac{(1-\beta(1-q))}{\theta}\left(\tau^p(\delta - \theta) - \delta C(\tau^p)\right)\right)^{\frac{1}{\theta}}\right]\left(\frac{k+\sigma}{\xi k + \sigma}\right) \tag{20.16}$$

This expression implies as usual that a higher level of τ^p makes democracy worse for the elite, and therefore increases φ^*, that is, the elite are willing to undertake more costly coups when τ^p is higher. We now have the following result.

PROPOSITION 20.4. *In the game described above, there is a unique Markov perfect equilibrium such that:*

- *If the coup constraint (20.15) does not bind, democracy is fully consolidated and the citizens set their most preferred tax rate $\tau^p > 0$.*
- *If the coup constraint (20.15) binds and $\varphi \geq \varphi^*$, then democracy is semi-consolidated. The citizens set a tax rate $\tau^D = \tilde{\tau} < \tau^p$ such that $V^r(D, \tau^D = \tilde{\tau}) = V^r(C, \varphi)$.*
- *If the coup constraint (20.15) binds and $\varphi < \varphi^*$, then democracy is unconsolidated. There is a coup, the elite come to power and set their most preferred tax rate, $\tau^N = 0$.*

The novel part of this result is that the likelihood of a coup is now affected by the economic structure, in particular, by whether or not society is capital or land intensive. However, the only reason why the degree of capital intensity will affect the propensity of the elite to mount coups is that different fractions of capital and land are destroyed in the process of the coup, i.e., $\varphi_K > \varphi_L$. To emphasize this, we can also state:

PROPOSITION 20.5. *Consider the above game with $\xi = 1$. Then (20.15) is independent of k, so the political equilibrium is unaffected by the capital intensity of the economy.*

The proof of this result follows from (20.16) since when $\xi = 1$ the term $(k + \sigma) / (\xi k + \sigma) = 1$ and cancels from the right side. This proposition states that there is no link between economic structure and capital intensity when costs of coups are the same for capital and land holders.

This picture changes substantially when $\xi > 1$, however. With a greater cost of coups on capital than land, (20.15) implies that as k increases the coup constraint becomes less tight and from (20.16), φ^* decreases. This implies that we can define two threshold levels $\hat{k}$ and k^* such that at $k = \hat{k}$, (20.15) holds with equality. On the other hand, $k = k^*$ is such that when democracy promises $\tau^D = 0$ the elite are indifferent between a coup and living in democracy. Naturally, $k^* < \hat{k}$. This discussion establishes the next result:

Proposition 20.6. *Consider a society described by the above game and assume that $\xi > 1$. Let $\hat{k}$ and k^* be as described above. In the unique subgame perfect equilibrium we have that: if $k < k^*$, then society is an unconsolidated democracy. If $k^* \leq k < \hat{k}$, then society is a semi-consolidated democracy. If $k \geq \hat{k}$, then society is a fully consolidated democracy.*

Therefore, in a very land intensive society where k is low, then there will be a coup during periods of crises. However, when the structure of production is different—i.e., when capital is relatively more important in the production process and in the asset portfolios of the elite, as captured by the threshold level of capital intensity k^*—then coups will no longer happen along the equilibrium path, and democracy persists. But because $k < \hat{k}$, democracy is not a fully consolidated political institution, and survives only by making concessions to the elite who pose an effective coup threat. As society becomes even more capital intensive and k increases, it will eventually become a fully consolidated democracy without the shadow of a coup affecting equilibrium tax rates and redistributive policies.

This model therefore illustrates how the structure of the economy, in particular the extent of capital intensity, influences the propensity of democracy to consolidate. The underlying idea is that in a more industrialized society with a greater fraction of the assets of the elite is in the form of physical capital, the turbulence and disruption associated with coups, like those created by repression, are more damaging. In consequence coups, as well as repression, will be less attractive in a capital intensive society.

20.5. Capital, Land and the Burden of Democracy

An even more important channel via which the economic structure may affect democracy is that the elite's attitudes toward democracy will also vary with the structure of the economy, because there are typically different burdens of taxation on capital and land. Let us now analyze a model with this feature. For the sake of brevity, we focus only on coups and democratic consolidations Given the analysis before, it should be clear that the analysis of transition to democracy is very similar, and factors discouraging coups will also discourage repression, facilitating transition to democracy.

The key idea here is that because land is supplied more inelastically, when allowed, the citizens will impose higher taxes on land than on capital. Thus, everything else equal, the elite will be more opposed to democracy when land is more important for their incomes. This gives us another reason for land intensive economies to be less likely to consolidate democracy (and also to transition to democracy).

Let us now discuss this issue by assuming that there can be separate taxes on income from different sources. In particular, a tax rate on capital income, τ_K, and one on income from land, τ_L. Throughout, we will simplify the discussion by assuming that there is no tax on labor income, i.e., the tax on labor, τ_N, is equal to 0. Clearly, the citizens would not like to tax their own incomes, but more generally, in a nondemocratic regime the elite might like to tax the citizens and redistribute to themselves, and to simplify the exposition, we ignore this possibility by restricting attention to the case where $\tau_N = 0$.

How do we model the costs of taxation when there are separate taxes on capital income and land income? The costs of taxation originate, in large part, from the fact that factors are supplied elastically. For example, labor taxation will be "costly" because individuals will take more leisure instead of supplying work to the market. There are two aspects to these costs, both of them relevant for this discussion. First, as less labor is supplied to the market, measured income and therefore tax revenues will decline. This constitutes a cost for those who will use tax revenues, since there are fewer revenues now. Second, there is also a cost of allocative efficiency; without the taxation, labor was being allocated to its best use, market work. Taxation discourages this, and creates a distortion by creating an incentive for time to be re-allocated away from its most efficient uses, forcing it to be used somewhere it is less valuable, in leisure or home production. Capital taxation will be similarly costly, especially as capital can flee to other activities, or even abroad, and avoid taxes. Again, this response of capital will be costly both because there are substantially less revenues from taxation, and also because the allocation of capital between various activities is distorted. More generally, in all cases, distortions from taxation result because in its effort to avoid taxes, each of these factors is not being allocated to its most productive use, and measured market income on which taxes are collected is declining. It is also important to note that both of these costs relate to the "elasticity of the supply" of various factors. When a factor is supplied inelastically, it cannot be withdrawn from market activity very easily, hence measured income will not change and there will be few distortions. Thinking of the supply elasticities as the major factor determining the costs of taxation immediately reveals that taxing capital should be more costly than taxing land. After all, capital can go to other sectors quite easily, but land is set in its place; at best, it can be withdrawn to inactivity.

Motivated by these considerations, we will think that when the tax on capital is τ_K there is a cost of taxation equal to $C_K(\tau_K)\,rK$, and when the tax on the land is τ_L, the cost of taxation is $C_L(\tau_L)\,vL$. As before, we assume that both of these functions are continuous, differentiable and convex. Moreover, we impose the usual boundary condition that $C_L'(0) = C_K'(0) = 0$, and a slightly different boundary condition $C_L'(1) > 1$ and $C_K'(1) > 1$ (the reason for this difference will become clear below). The crucial assumption we make is that

$$C_L'(\tau) < C_K'(\tau) \text{ for all } \tau > 0.$$

This assumption implies that the marginal cost of taxing capital is always higher than the marginal cost of taxing land, which is equivalent to capital being supplied more elastically than land. The important implication of this assumption will be that the citizens would like to impose greater taxes on land than on capital.

To further simplify the discussion, we now depart in one more respect from our baseline model, and as in our targeted transfers model, assume that as well as lump sum transfers, there are also transfers targeted to specific groups, in particular to the citizens, T_p, as well as a lump sum transfer to the elite, T_r.

Given all these pieces, we can write the total post-tax incomes of the elite and the citizens as follows:

$$\begin{aligned} \hat{y}^p &= w + T_p \\ \hat{y}^r &= (1-\tau_K)\frac{rK}{\delta} + (1-\tau_L)\frac{vL}{\delta} + T_r, \end{aligned}$$

which incorporates are assumption above that all capital and land are equally owned by each member of the elite, and there are δ of them.

The government budget constraint can now be written as

$$\delta T_r + (1-\delta)T_p = \tau_K rK - C_K(\tau_K)\,rK + \tau_L vL - C_L(\tau_L)\,vL \tag{20.17}$$

The left hand side of (20.17) is total expenditure on transfers. T_r is the lump sum transfer that members of the elite receive, and is thus multiplied by δ, T_p is the transfer to a citizen, hence it is multiplied by $1-\delta$. The right hand side is total tax revenue from the taxation of capital and land. At the tax rates τ_K, τ_L, capital owners pay a total of $\tau_K rK$ in tax, and landowners pay $\tau_L vL$. From these amounts, we need to subtract the costs of taxation, $C_K(\tau_K)\,rK$ and $C_L(\tau_L)\,vL$.

Given the availability of a targeted transfer to themselves, the citizens would simply redistribute all the income they raise from capital and land using this targeted transfer, hence we will have $T_r = 0$ in democracy.

Next, note that since the citizens are no longer taxing themselves, their most preferred taxes will be those that maximize the net tax receipts, the right hand side of (20.17)—in other words, the citizens would now like to be at the top of the Laffer curve, which relates total tax revenue to tax rate. Therefore, citizens' most preferred taxes can be computed simply by solving the following maximization problem:

$$\max_{\tau_K,\tau_L} \tau_K rK - C_K(\tau_K) rK + \tau_L vL - C_L(\tau_L) vL.$$

The first-order conditions are straightforward and give the most preferred taxes for the poor, τ_K^p, τ_L^p implicitly as:

$$\begin{aligned} C'_K\left(\tau_K^p\right) &= 1 \\ C'_L\left(\tau_L^p\right) &= 1 \end{aligned} \tag{20.18}$$

which maximize their net tax revenues. The assumption that $C'_L(\tau) < C'_K(\tau)$ immediately implies that $\tau_K^p < \tau_L^p$.

Let us next compute the net burden of democratic taxation on the elite. Let us define the burden to be the net amount of redistribution away from the elite. Since they receive no transfers now, this is simply equal to taxes they pay, hence

$$\text{Burden}\left(\tau_K^p, \tau_L^p\right) = \tau_K^p rK + \tau_L^p vL.$$

Using (20.3), we can write this relative to total income and in terms of capital intensity as

$$\mathcal{B} \equiv \frac{\text{Burden}\left(\tau_K^p, \tau_L^p\right)}{Y} = \tau_K^p \frac{k}{k+\sigma} + \tau_L^p \frac{\sigma}{k+\sigma} \tag{20.19}$$

First note that from (20.18) the tax rates τ_K^p and τ_L^p are independent of k. Then equation (20.19) implies that as the economy becomes more capital intensive, the burden of democracy on the elite will decrease. This reflects the fact that capital is less attractive to tax than land. To see this analytically, note that the burden of taxes, $\mathcal{B}$, is decreasing in capital intensity:

$$\frac{d\mathcal{B}}{dk} = \frac{\tau_K^p}{k+\sigma} - \frac{\tau_K^p k + \tau_L^p \sigma}{(k+\sigma)^2} < 0,$$

which follows immediately from the fact that $\tau_K^p < \tau_L^p$. This result implies that elites will be less opposed to democracy for another reason when they are invested more and capital than land; this is because democracy will tax capital less than it will tax land.

There is another interesting interpretation of $\tau_K^p < \tau_L^p$. So far, we have emphasized the different tax rates imposed on incomes generated by land and capital. Another possibility is redistribution of assets. Since asset redistribution has not been explicitly considered in this chapter, we might think that the potential for asset redistribution is also incorporated into these taxes τ_K^p and τ_L^p. Are there any reasons to think that the potential for asset redistribution is different for capital and for land? The answer is yes. While democracy can easily redistribute land via land reform, redistribution of capital is much harder since capital, in the form of factories, is not easily divisible. More important, when these factories are taken away from their owners and given to new parties, they will typically not be very productive. This is because the complex relationships necessary for capitalist production, the specific investments, the know-how, are all in the hands of the original owners, and very difficult, or even impossible to transfer. One could argue that rather than redistribute the capital itself, shares in firms could be redistributed, yet the modern theory of the firm suggests precisely that the incentives of agents within a firm depend on the ownership structure so that capital cannot be arbitrarily redistributed without damaging productivity. Indeed, if capital markets are perfect one would expect the initial ownership structure to be efficient.

Land is much easier to redistribute without creating distortions. When land is taken from big land owners and redistributed to agrarian workers, the loss of efficiency may not be significant, and in fact according to some estimates, there might even be a gain in efficiency since many of the big farms are owned by major landowners who farm more land than is efficient. This suggests that land reform will often be an attractive policy tool for democracies to achieve their fiscal objectives without creating major distortions. Naturally, this implies a greater burden of democracy on landowners than for capital owners. This consideration implies that when land is a more important part of the assets of the rich, they will have more to fear from democracy, and typically they will expect greater redistribution away from them and a greater burden. This could be captured by our result that $\tau_K^p < \tau_L^p$.

Let us now put these two pieces together and analyze the likelihood of coups in a world with different taxes on capital and land. Consider the same political game is above, and also assume that the same fraction of capital and land are destroyed in the process of a coup, i.e., $\varphi_K = \varphi_L$ or that $\xi = 1$.

If the citizens get to set their most preferred taxes and transfers, taxes on capital and land will be given by (20.18), and we will also have $T_r = 0$. This implies that the transfer to each citizen will be given by

$$T_p^p = \frac{\tau_K^p rK - C_K\left(\tau_K^p\right) rK + \tau_L^p vL - C_L\left(\tau_L^p\right) vL}{1-\delta} \tag{20.20}$$

with τ_K^p and τ_L^p given by (20.18) and the superscript p on T_p^p indicates that it is the preferred value of the citizens. Therefore, the corresponding values are those in an unconstrained democracy:

$$\begin{aligned} V^p(D) &= \frac{w+T_p^p}{1-\beta}, \\ V^r(D) &= \frac{\left(1-\tau_K^p\right) rK + \left(1-\tau_L^p\right) vL}{\delta\left(1-\beta\right)}, \end{aligned} \tag{20.21}$$

with factor prices, w, r, and v, given by (20.2), with τ_K^p and τ_L^p given by (20.18) and T_p^p given by (20.20).

Whether the elite mount a coup will depend on the continuation values in democracy and nondemocracy. The citizens again set taxes on capital and labor income, which are potentially different from their most preferred tax rates, τ_K^p and τ_L^p, and we denote them by $\tilde{\tau}_K$ and $\tilde{\tau}_L$. The corresponding redistribution to a citizen is

$$\tilde{T}_p^p = \frac{\tilde{\tau}_K rK - C_K\left(\tilde{\tau}_K\right) rK + \tilde{\tau}_L vL - C_L\left(\tilde{\tau}_L\right) vL}{1-\delta}. \tag{20.22}$$

(That the citizens would decide to cut taxes on capital and land rather than redistribute lump sum to the elite is obvious, since these taxes are distortionary. Note also that if we had allowed labor income to be taxed the citizens could find it optimal to tax themselves and transfers resources to the elite to avoid a coup).

After this, the elite decide whether to undertake the coup. If they do so, society switches to nondemocracy, and the elite set the tax rate. Naturally, they will choose their most preferred tax rates, $\tau_K^N = \tau_L^N = 0$. As a result, the game ends with respective payoffs for the citizens and the elite, $V^p(C,\varphi)$ and $V^r(C,\varphi)$, where

$$\begin{aligned} V^p(C,\varphi) &= \frac{(1-\theta)(1-\varphi)^\theta\left(\frac{K+\sigma L}{1-\delta}\right)^\theta}{(1-\beta)}, \\ V^r(C,\varphi) &= \frac{\theta}{\delta\left(1-\beta\right)}(1-\varphi)^\theta\left(K+\sigma L\right)^\theta\left(1-\delta\right)^{1-\theta} = \frac{\theta}{\delta\left(1-\beta\right)}(1-\varphi)^\theta Y \end{aligned} \tag{20.23}$$

Whether a coup is attractive depends on whether the coup constraint, $V^r(C,\varphi) > V^r(D)$, binds. The answer will be yes when the burden of taxation on the elite is sufficiently high. Using (20.21) and (20.23), the coup constraint can be expressed as

$$(1-\varphi)^\theta > (1-\tau_K^p)\frac{k}{k+\sigma} + (1-\tau_L^p)\frac{\sigma}{k+\sigma}. \tag{20.24}$$

When this constraint does not bind, democracy is fully consolidated.

In contrast, when this constraint binds, democracy is not fully consolidated: if the citizens do not take an action, there will be a coup along the equilibrium path. The action that the citizens can take is to reduce the burden that democracy places upon the elite, by reducing taxes both on capital and land. In particular, the best that the citizens can do is to promise zero taxes on both: $V^r(D, \tau_K^D = 0, \tau_L^D = 0)$ to the elite. As in our previous analysis, we can then define a threshold value for φ, φ^*, such that when $\varphi < \varphi^*$, the promise of limited distribution by the citizens is not sufficient to dissuade the elite from a coup. Therefore, we must have that at φ^*, $V^r(D, \tau_K^D = 0, \tau_L^D = 0) = V^r(C, \varphi^*)$. Solving this equality gives the threshold value φ^* as:

$$\varphi^* = 1 - \left(\left(1 - (1 - \beta(1-q))\, \tau_K^p\right) \frac{k}{k+\sigma} + \left(1 - (1 - \beta(1-q))\, \tau_L^p\right) \frac{\sigma}{k+\sigma} \right)^{\frac{1}{\theta}}. \tag{20.25}$$

PROPOSITION 20.7. *In the game described above, there is a unique Markov perfect equilibrium such that:*

- *If the coup constraint (20.24) does not bind, democracy is fully consolidated and the citizens set their most preferred tax rates on capital and land $\tau_K^p > 0$ and $\tau_L^p > 0$ as given by (20.18).*
- *If the coup constraint (20.24) binds and $\varphi \geq \varphi^*$, then democracy is semi-consolidated. The citizens reduce taxes below τ_K^p and τ_L^p.*
- *If the coup constraint (20.24) binds and $\varphi < \varphi^*$, then democracy is unconsolidated. There is a coup, the elite come to power and set their preferred tax rates, $\tau_K^N = \tau_L^N = 0$.*

Let us again define two threshold levels of capital intensity $\hat{k}$ and k^*, such that as the economy passes these threshold levels it first becomes a semi-consolidated and then a fully consolidated democracy. These threshold values are:

$$k^* = \frac{\left(\left(\left(1 - (1 - \beta(1-q))\, \tau_L^p\right)\right) - (1-\varphi)^\theta\right) \sigma}{(1-\varphi)^\theta - \left(1 - (1 - \beta(1-q))\, \tau_K^p\right)} \tag{20.26}$$

and

$$\hat{k} = \frac{\left((1 - \tau_L^p) - (1-\varphi)^\theta\right) \sigma}{(1-\varphi)^\theta - (1 - \tau_K^p)} \tag{20.27}$$

Then Proposition 20.6 applies exactly as before with k^* and $\hat{k}$ as given by (20.26) and (20.27). The result is therefore very similar to before: as capital and industry become more important relative to land and agriculture, the elite become less averse to democracy and the threat against democracy diminishes. The reason why this happens is somewhat different from before, however. Previously, the burden of democracy was independent of the composition of assets of the elite; their different attitudes towards coups originated from the different costs that the disruption due to a coup would cause. Perhaps more important in practice is that not all segments of the elite will suffer equally in democracy. The present model emphasizes this by constructing a model where land is taxed more heavily (or perhaps redistributed more radically by land reform), and therefore, the elite have more to fear from democracy when land is an important source of income for them. As the degree of capital intensity increases their opposition to democracy declines and consolidation is more likely.

The implications of the analysis here carry over immediately to democratization. Since the burden of democracy falls more heavily on landowners than on capitalists as the capital intensity of the economy increases, repression will become less and less attractive relative to democracy and democratization becomes more likely to arise. Indeed, by analogy to the above analysis there will exist a level of capital intensity which is sufficiently high to ensure that repression will never be attractive to the elite.

20.6. References

(1) Acemoglu, Daron and James Robinson (2006) "Chapter 9: Economic Structure and Democracy" in Acemoglu and Robinson *Economic Origins of Dictatorship and Democracy*, Cambridge University Press.

(2) Moore, Barrington (1966) *The Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*, Beacon Press, Boston MA.

CHAPTER 21

Change and Persistence in Institutions

The models studied so far illustrate how institutions can be chosen in order to influence the future distribution of political power. This provides a workable notion of institutions, but its empirical applications are not without problems. The most important one is that in practice institutions change quite often. For example, colonial rule ends, constitutions get rewritten, democracy collapses. Consequently, a useful model is to help us understand both persistence and change in institutions. The model of switches between dictatorship and democracy goes some way towards this, but does not have micro-founded model of persistence. The only reason why institutions (democracy or dictatorship) persist in that model was that taking actions against any existing regime is costly.

Another problem with the baseline Acemoglu-Robinson (2006a) framework, we presented so far is that the elites role in democracy is not explicitly modeled. But many accounts of early transitions to democracy, especially in less-developed countries, suggest the possibility of "a captured democracy" whereby the elite are able to play a disproportionate role in democracy. Clearly, whether they can do so or not will have a first-order affect on the process of democratization and also on what democracy achieves.

The issue of the persistence of institutions has been looked at much more from the other side, i.e. institutions stay the same but the economic environment changes. The most famous example of this is the typewriter keyboard with its distinctive keyboard arrangement of QWERTY. David (1985) argued that odd organization of letters reflected historical accidents in the 19th century which were then locked into place by network externalities. This organization of keys was determined to *slow down* typing so that mechanical keys did not jam. But then everyone standardized on this keyboard organization which then became locked in place even though the initial technological factors became completely irrelevant. Similar ideas are applied to institutions such as legal systems. The fact that former French colonies burdened with the Civil Code to do switch to supposedly better British legal institutions may be due to such network externalities. A related idea is that people make specific investments in particular sets of institutions (see Coate and Morris, 1999, for an application to policies).

This approach at least does have the advantage of emphasizing what a theory of institutional persistence is about. For example, one reading of Engerman and Sokoloff (1997) is that it is inequality that drives bad institutions and bad institutions lead to inequality. So inequality is a state variable which reproduces itself over time along with bad institutions. This is like the model of Banerjee and Newman (1993) where inequality leads to an inefficient organization of production because of imperfect capital markets which then leads to high inequality etc. As in Banerjee and Newman, the thrust of such an argument is that if you could change inequality in one period through some massive intervention, then institutions would get better. This is not what the persistence of institutions is about however. Interesting institutional persistence says that once created institutions have a tendency to persist even when the economic environment which led to them changes.

Though it is undoubtedly challenging to have nice models of why institutions persist when the environment that led to their creation vanishes, this may be too easy a problem and it may not be the most relevant one. As we noted already, institutions change a lot even if the typewriter keyboard hasn't. (Maybe it hasn't because there is little efficiency incentive to change it - David's claims on this have been severely challenged by Liebowitz and Margolis, 2002). Not only do political institutions change a lot, but so do relevant economic institutions. For example, research by Engerman and Sokoloff (1997) and Acemoglu, Johnson and Robinson (2001) often motivates the bad incentive environment created in Latin America during the colonial period by pointing to specific institutions, such as types of forced labor, compulsory labor services or slavery. Yet these institutions all vanished a long time ago. Cuba and Brazil were the last places to abolish slavery in the Americas in the late 1880s. Thailand abolished it in 1905 and China in 1910. One might wonder how effective all of these acts were, for example Mauritania abolished slavery at least three times in the 20th century, in 1905, 1906 and 1980! (see Engerman, 2006). In terms of forced labor, the Bolivian revolution of 1952 appears to have been the end of unpaid labor services in Latin America (the apparently much hated *pongueaje*). The political science literature on the persistence of institutions, so called "Historical Institutionalism" similarly tends to focus on the origin and persistence of specific institutions.

It is important to note however that even if specific institutions change, the consequences of these specific institutions may not because other institutions may fulfill the same goals. In Acemoglu, Johnson and Robinson (2001) the focus is very much on an institutional outcome, security of property rights. There are many ways that property rights can be made secure, or insecure. The emphasis here is not on specific institutions and many sets of specific institutions can lead to the same outcome with respect to property rights or the institutions that critically condition economic incentives. To take the main historical example developed in Acemoglu and Robinson (2006b) before the Civil War the economy of the US South was dominated by slavery. After the Civil War slavery was abolished but after the 1870s a remarkably similar economy developed using different instruments to repress labor. Here it makes sense to talk about institutional persistence even though specific economic and political institutions changed.

One way of reconciling persistence with change would be a more explicit model of path dependency follows: in the Colonial period the different factor endowments in Latin America and the US led to very different institutions, call these ℓ and u with ℓ being less efficient, though the differences may not be large before the 19th century. In the 19th century industry came along and to have industry you needed to adopt or develop new institutions, such as financial institutions, maybe patent laws etc. There are different ways you can do this, for instance you can have crony capitalistic financial institutions where banks lend to their friends or people with political connections or you can have good financial institutions. Either can be created/adopted. The the idea would be to show that ℓ leads to crony capitalism despite good institutions being feasible, while u leads to good institutions (despite crony capitalism being possible). So far nobody has a model like this.

Here we follows Acemoglu and Robinson (2006b) and develop another idea about how persistence and change can coexist which emphasizes how different sources of political power may offset each other. The main idea is to build in another dimension of asymmetry between citizens and elites, which is that elites, partly because of their established position and partly because of their smaller numbers or social connections, will be more successful in investing in activities that increase their de facto political power (bribery, lobbying, employment of paramilitaries). This observation implies that a change in institutions modifying the distribution of de jure power may lead to a fully or partially offsetting change in de facto political power because of the investments of the elites. In consequence, there may be little change in who determines economic institutions when de jure power changes because it is offset by changes in de facto power.

21.1. Baseline Model

21.1.1. Demographics, Preferences and Production Structure. Consider an infinite-horizon society in discrete time with a unique final good and populated by a continuum 1 of worker/citizens and (a finite) number $M > 1$ of the elites. All agents have the same risk-neutral preferences with discount factor β, given by

$$\sum_{j=0}^{\infty} \beta^j c^i_{t+j} \tag{21.1}$$

at time t, where c^i_{t+j} denotes consumption of agent i at time $t+j$ in terms of the final good. We use the notation $i \in \mathcal{E}$ to denote an elite agent, and $i \in \mathcal{C}$ to denote a citizen.

All workers own one unit of labor, which they supply inelastically. Each member of the elite $i \in \mathcal{E}$ has access to the following production function to produce the unique final good:

$$Y^i_L = \begin{cases} G\left(L^i, N^i_L\right) & \text{if } L^i \leq \frac{L}{M} \\ G\left(\frac{L}{M}, N^i_L\right) & \text{if } L^i > \frac{L}{M} \end{cases} \tag{21.2}$$

where L^i denotes land and N^i_L denotes labor used by this producer, and G exhibits constant returns to scale. This production function implies that there is a maximum land size of L/M after which each producer runs into severe diminishing returns (where the fact that diminishing returns start after land size of L/M is a normalization). There is a total supply of land equal to L in the economy, with no alternative use, and each elite owns L/M units of land (and no labor).

The final good can also be produced with an alternative technology, which can be interpreted as small-scale production by the laborers themselves (or a low productivity proto-industry technology). This alternative technology exhibits constant returns to scale to labor:

$$Y_A = AN_A. \tag{21.3}$$

Clearly, total output of the unique final good in the economy will be $Y = \sum_{i\in\mathcal{E}} Y^i_L + Y_A$, and the market clearing condition for labor is

$$\sum_{i\in\mathcal{E}} N^i_L + N_A \leq 1. \tag{21.4}$$

The main role of the alternative technology, (21.3), will be to restrict how low wages can fall in this economy.

We consider two different *economic institutions*. In the first, labor markets are *competitive*. Given (21.2), each elite will hire $N^i_L = N_L/M$ units of labor, where $N_L = 1 - N_A$, and since G exhibits constant returns to scale, we can write per capita output as:

$$y = G\left(\frac{L}{N_L}, 1\right) = g\left(\frac{L}{N_L}\right). \tag{21.5}$$

When there are competitive labor markets, which we denote by $\tau = 1$, the wage rate (and the wage earnings of each worker), as a function of labor allocated to this sector, N_L, is therefore:

$$w^c\left[N_L\right] \equiv g\left(\frac{L}{N_L}\right) - \frac{L}{N_L} g'\left(\frac{L}{N_L}\right), \tag{21.6}$$

where the superscript c denotes "competitive". The return to landowners with competitive markets is similarly

$$R^c[N_L] \equiv g'\left(\frac{L}{N_L}\right), \tag{21.7}$$

with each landowner receiving $R^c L/M$.

Assumption 1:

$$g(L) - Lg'(L) > A.$$

This assumption implies that even when $N_L = 1$ (i.e., when $L/N_L = L$), the competitive wage in this sector is greater than the marginal product of labor in the alternative technology. Therefore, both the efficient allocation and the competitive equilibrium allocation will have all workers allocated to the land sector, i.e., $N_L = 1$. In light of this, the relevant competitive wage and rental return on land will be

$$w^c \equiv w^c[N_L = 1] \equiv g(L) - Lg'(L), \tag{21.8}$$

and

$$R^c \equiv R^c[N_L = 1] \equiv g'(L). \tag{21.9}$$

Consequently, factor prices at time t as a function of economic institutions are given by $w_t = w(\tau_t = 1) = w^c$ and $R_t = R(\tau_t = 1) = R^c$, with w^c and R^c as defined in (21.8) and (21.9). [More formally, the second welfare theorem combined with preferences in (21.1) implies that a competitive equilibrium is a solution to the following program:

$$\max_{N_L.N_A,\tilde{L}} g\left(\frac{\tilde{L}}{N_L}\right) N_L + AN_A$$

subject to (21.4) and $\tilde{L} \leq L$. Assumption 1 ensures that the solution involves $N_L = 1$ and $\tilde{L} = L$, and the equilibrium factor prices are given by the shadow prices of this program.]

The alternative set of economic institutions are *labor repressive* ($\tau_t = 0$) and allow the landowning elite to use their political power to reduce wages below competitive levels. They cannot, however, force workers to work (i.e., slavery is not allowed), so workers always have access to the alternative small-scale production technology. Consequently, when economic institutions are labored oppressive, the lowest wage that the elite can pay the workers, while still ensuring that $N_L > 0$, is A. This implies that factor prices under these economic institutions are

$$w^r \equiv A, \tag{21.10}$$

and

$$R^r \equiv \frac{f(L) - A}{L}. \tag{21.11}$$

(Recall that the landed elite are paying the wage of A to a total of $N_L = 1$ workers). When economic institutions are labor repressive, then we will have $w_t = w(\tau_t = 0) = w^r$ and $R_t = R(\tau_t = 0) = R^r$. Assumption 1 immediately implies that $R^r > R^c$, since with labor repressive economic institutions wages are kept artificially low, i.e., $w^r < w^c$, so that land owners enjoy greater rents. For future reference, we define

$$\begin{aligned} \Delta R &\equiv R^r - R^c \\ &= \frac{g(L) - A}{L} - g'(L) > 0. \end{aligned} \tag{21.12}$$

One feature to note is that the simple environment outlined here implies that both competitive labor markets and labor repression will generate the same total output, and will differ only in terms of their distributional implications. Naturally, it is possible to introduce additional costs from labor repressive economic institutions, which may include standard monopsony distortions or other costs involved in monitoring and forcing laborers to work at below market-clearing wages (such as wasteful expenditures on monitoring, paramilitaries, or lower efficiency of workers because of the lower payments they receive). Incorporating such costs has no effect on the analysis, and throughout, one may wish to consider the labor repressive institutions as corresponding to "worse economic institutions".

21.1.2. Political Regimes and De Facto Political Power. There are two possible political regimes, denoted by D and N, corresponding to democracy and nondemocracy. The distribution of de jure political power will vary between these two regimes. At any point in time, the "state" of this society will be represented by $s_t \in \{D, N\}$, which designates the political regime that applies at that date. Importantly, irrespective of the political regime (state), the identity of landowners and workers does not change; the same M individuals control the land, and have the potential to exercise additional political power.

Overall political power is determined by the interaction of de facto and de jure political power. Since there is a continuum of citizens, they will have difficulty in solving the collective action problem to exercise de facto political power. Consequently, we treat their de facto power as being exogenous rather than stemming from their own contributions.

In contrast, elites can spend part of their earnings to gather further de facto political power. In particular, suppose that elite $i \in \mathcal{E}$ spends an amount $\theta_t^i \geq 0$ as a contribution to activities increasing their group's de facto power. Then total elite spending on such activities will be $Z_t = \sum_{i \in \mathcal{E}} \theta_t^i$, and we assume that their de facto political power is

$$P_t^E = \phi Z_t, \tag{21.13}$$

where $\phi > 0$. The reason why the elite may choose to spend a positive amount on such activities is that there is a finite number, M, of them, so each of them will take into account that their own contribution to total spending, Z_t, will have an effect on equilibrium outcomes. An important assumption implicit in (21.13) is that the technology for generating de facto political power for the elite is the same in democracy and nondemocracy.

Even though the citizens cannot solve the collective action problem to invest in their de facto political power, since they form the majority in society they always possess some political power. The extent of this power depends on whether the political regime is democratic or nondemocratic. We model the citizens' total political power in a reduced-form manner as follows:

$$P_t^C = \omega_t + \eta I \left(s_t = D\right), \tag{21.14}$$

where ω_t is a random variable drawn independently and identically over time from a given distribution $F(\cdot)$ and measures their de facto power; $I(s_t = D)$ is an indicator function for $s_t = D$, such that $I(s_t = D) = 1$ while $I(s_t = N) = 0$; and η is a strictly positive parameter measuring citizens' de jure power in democracy.

There are two important assumptions embedded in equation (21.14). The first is that the de facto political power of the citizens fluctuates over time, and is hard to predict in advance. The second assumption is that when the political regime is democratic, i.e., $s_t = D$, citizens have greater political power. This represents in a very simple way the fact that democracy allocates de jure political power in favor of the majority. This will be both because of the formal rules of democracy and also because in democratic politics, parties may partly solve the collective action problem of the citizens. Put differently, equation (21.14) implies that in democracy the political power of the citizens shifts to the right in the sense of first-order stochastic dominance. To simplify the discussion, we make the following assumptions on F:

> **Assumption 2:** F is defined over $(\underline{\omega}, \infty)$ for some $\underline{\omega} < 0$, is everywhere strictly increasing and twice continuously differentiable (so that its density f and the derivative of the density, f', exist everywhere). Moreover, $f(\omega)$ is single peaked (in the sense that there exists ω^* such that $f'(\omega) > 0$ for all $\omega < \omega^*$ and $f'(\omega) < 0$ for all $\omega > \omega^*$) and satisfies $\lim_{\omega \to \infty} f(\omega) = 0$.

All of the features embedded in Assumption 2 are for convenience, and how relaxing them affects the equilibrium is discussed below (note that in fact $\lim_{\omega\to\infty} f(\omega) = 0$ follows from single peakedness).

We introduce the variable $\pi_t \in \{0, 1\}$ to denote whether the elite have more (total) political power. In particular, when $P_t^E \geq P_t^C$, we have $\pi_t = 0$ and the elite have more political power and will make the key decisions. In contrast, whenever $P_t^E < P_t^C$, $\pi_t = 1$ and citizens have more political power, and they will make the key decisions.

To complete the description of the environment, it remains to specify what these key decisions are. We assume that the group with greater political power will decide both economic institutions at time t, τ_t, and what the political regime will be in the following period, s_{t+1}.

When the elite have more political power, a representative elite agent makes the key decisions, and when citizens have more political power, a representative citizen does so. Since the political preferences of all elites and all citizens are the same, these representative agents will always make the decisions favored by their group.

21.1.3. Timing of Events. We now briefly recap the timing of events in this basic environment.

At each date t, society starts with a state variable $s_t \in \{D, N\}$. Given this, the following sequence of events take place:

(1) Each elite agent $i \in \mathcal{E}$ simultaneously chooses how much to spend to acquire de facto political power for their group, $\theta_t^i \geq 0$, and P_t^E is determined according to (21.13).
(2) The random variable ω_t is drawn from the distribution F, and P_t^C is determined according to (21.14).
(3) If $P_t^E \geq P_t^C$ (i.e., $\pi_t = 0$), a representative (e.g., randomly chosen) elite agent chooses (τ_t, s_{t+1}), and if $P_t^E < P_t^C$ (i.e., $\pi_t = 1$), a representative citizen chooses (τ_t, s_{t+1}).
(4) Given τ_t, transactions in the labor market take place, R_t and w_t are paid to elites and workers respectively, and consumption takes place.
(5) The following date, $t + 1$, starts with state s_{t+1}.

21.2. Analysis of Baseline Model

We now analyze the baseline model described so far. We first focus on the symmetric Markov Perfect Equilibria (MPE). An MPE imposes the restriction that equilibrium strategies are mappings from payoff-relevant states, which here only include $s \in \{D, N\}$. In particular, in an MPE strategies are not conditioned on the past history of the game over and above the influence of this past history on the payoff-relevant state s. An MPE will consist of contribution functions $\left\{\theta^i(s)\right\}_{i\in\mathcal{E}}$ for each elite agent as a function of the political state, and decision variables $\tau(\pi)$ and $s'(\pi)$ as a function of $\pi \in \{0,1\}$ denoting which side has more political power, and equilibrium factor prices as given by (21.8)-(21.11). Here the function $\tau(\pi)$ determines the equilibrium decision about labor repression conditional on who has power and the function $s'(\pi) \in \{D,N\}$ determines the political state at the start of the next period. Symmetric MPE will in addition impose the condition that contribution functions take the form $\theta(s)$, i.e., do not depend on the identity of the individual elite, i. Symmetry is a natural feature here, and simplifies the analysis. We discuss asymmetric MPE for completeness below. A more formal definition of an MPE is also given below.

The focus on MPE is natural in this context as a way of modeling the potential collective action problem among the elite. Looking at subgame perfect equilibrium (SPE) will allow the elite greater latitude in solving the collective action problem by using implicit punishment strategies. We briefly analyze SPEs below.

21.2.1. Main Results. The MPE can be characterized by backward induction within the stage game at some arbitrary date t, given the state $s \in \{D,N\}$. At the last stage of the game, clearly whenever the elite have political power, i.e., $\pi = 0$, they will choose economic institutions that favor them, i.e., $\tau = 0$, and a political system that gives them more power in the future, i.e., $s' = N$. In contrast, whenever citizens have political power, i.e., $\pi = 1$, they will choose $\tau = 1$ and $s' = D$. This implies that choices over economic institutions and political states are straightforward. Moreover the determination of market prices under different economic institutions has already been specified above (recall equations (21.8)-(21.11)). Thus the only remaining decisions are the contributions of each elite agent to their de facto power, θ^i_t. Therefore, a symmetric MPE can be summarized by a level of contribution as a function of the state $\theta(s)$. It will be convenient to characterize the MPE by writing the payoff to elite agents recursively, and for this reason, we denote the equilibrium value of an elite agent in state s by $V(s)$ (i.e., $V(D)$ for democracy and $V(N)$ for nondemocracy).

Let us begin with nondemocracy. Since we are focusing on symmetric MPE, suppose that all other elite agents, except $i \in \mathcal{E}$, have chosen a level of contribution to de facto power equal to $\theta(N)$. Consequently, when agent $i \in \mathcal{E}$ chooses θ^i, their total power will be

$$P^E\left(\theta^i, \theta(N) \mid N\right) = \phi\left((M-1)\,\theta(N) + \theta^i\right).$$

The elite will have political power if

$$P^E\left(\theta^i,\theta(N)\mid N\right)=\phi\left((M-1)\,\theta(N)+\theta^i\right)\geq\omega_t.$$

Expressed differently, the probability that the elite will have political power in this state is

$$p\left(\theta^i,\theta(N)\mid N\right)=F\left(\phi\left((M-1)\,\theta(N)+\theta^i\right)\right). \tag{21.15}$$

We can then write the net present discounted value of agent $i\in\mathcal{E}$ recursively as

$$\begin{aligned} V(N\mid\theta(N),\theta(D)) &= \max_{\theta^i\geq 0}\left\{-\theta^i+p\left(\theta^i,\theta(N)\mid N\right)\left(\frac{R^rL}{M}+\beta V(N\mid\theta(N),\theta(D))\right)\right. \\ &\quad \left.+\left(1-p\left(\theta^i,\theta(N)\mid N\right)\right)\left(\frac{R^cL}{M}+\beta V(D\mid\theta(N),\theta(D))\right)\right\}, \end{aligned} \tag{21.16}$$

where recall that R^c is the rate of return on land in competitive markets, given by (21.9) and R^r is the rate of return on land under labor repressive economic institutions, given by (21.11). The function $V(N\mid\theta(N),\theta(D))$ recursively defines the value of an elite agent in nondemocracy when all other elite agents choose contributions $\theta(N)$ in nondemocracy and $\theta(D)$ in democracy. Similarly, $V(D\mid\theta(N),\theta(D))$ is the value in democracy under the same circumstances.

The form of the value function in (21.16) is intuitive. It consists of the forgone consumption because of the expenditure θ^i, plus the revenues and the continuation values. In particular, given his contribution θ^i and those of other elite agents in nondemocracy, $\theta(N)$, political power will remain in the hands of the elite with probability $p\left(\theta^i,\theta(N)\mid N\right)$, in which case economic institutions will be labor repressive, and this elite agent receives revenue equal to R^rL/M (rate of return under labor repressive economic institutions, R^r, times his land holdings, L/M) and the discounted continuation value of remaining in nondemocracy, $\beta V(N\mid\theta(N),\theta(D))$. With probability $1-p\left(\theta^i,\theta(N)\mid N\right)$, citizens have greater political power, so they choose $\tau=1$ and labor markets are competitive. In this case a member of the elite receives revenue equal to R^cL/M and continuation value $\beta V(D\mid\theta(N),\theta(D))$, since with power in their hands, the citizens will choose to change the political system to $s_{t+1}=D$.

Agent $i\in\mathcal{E}$ chooses θ^i to maximize his net expected present discounted utility. Let the policy function (correspondence) for the maximization in (21.16) be $\Gamma^N[\theta(N),\theta(D)]$, so that any $\theta^i\in\Gamma^N[\theta(N),\theta(D)]$ is an optimal policy for the value function in (21.16) (in state $s=N$).

Since F is continuously differentiable and everywhere increasing (from Assumption 2), so is $p\left(\theta^i,\theta(N)\mid N\right)$, which implies a particularly simple first-order necessary condition for

(21.16):

$$\phi f\left(\phi\left((M-1)\theta(N)+\theta^i\right)\right)\left(\frac{\Delta RL}{M}+\beta\left(V\left(N\mid\theta(N),\theta(D)\right)-V\left(D\mid\theta(N),\theta(D)\right)\right)\right)\leq 1, \tag{21.17}$$

and $\theta^i \geq 0$, with complementary slackness, where recall that $\Delta R \equiv R^r - R^c$ is defined in (21.12), and f is the density function of the distribution function F. Moreover, it is clear that we need the additional second-order condition that $f'\left(\phi\left((M-1)\theta(N)+\theta^i\right)\right) < 0$. The reason why the maximization problem for individual i in this recursive formulation is so simple is that θ^i does not affect $V\left(N\mid\theta(N),\theta(D)\right)$ or $V\left(D\mid\theta(N),\theta(D)\right)$, so differentiability of the maximand is guaranteed.

Expressed differently, any $\theta^i \in \Gamma^N\left[\theta(N),\theta(D)\right]$ must solve (21.17) and satisfy the corresponding second-order condition. The first-order condition is quite intuitive: the cost of forgone consumption, which is the right hand side of (21.17), must be equal to (or less than) the benefit from this contribution, which is the marginal increase in the probability of the elite having more political power than the citizens, i.e., $\phi f(\cdot)$, and the benefit that the agent will derive from this political power, which is the second term on the left-hand side, consisting of the direct benefit $\Delta RL/M$ plus the benefit in terms of continuation value. Moreover, since we are focusing on a symmetric MPE, $\theta^i > 0$ is equivalent to $\theta(N) > 0$, so if there is any investment in de facto power by the elite, then (21.17) must hold as an equality.

Next, consider the society starting in democracy. With the same argument as above, the elite will have political power if

$$P^D\left(\theta^i,\theta(D)\mid D\right)=\phi\left((M-1)\theta(D)+\theta^i\right)\geq \omega_t+\eta,$$

which only differs from the above expression because with $s_t = D$, the citizens have an additional advantage represented by the positive parameter η. Then the probability that the elite will capture political power in democracy is

$$p\left(\theta^i,\theta(D)\mid D\right)=F\left(\phi\left((M-1)\theta(D)+\theta^i\right)-\eta\right), \tag{21.18}$$

and using the same reasoning as before, the value function for elite agent $i \in \mathcal{E}$ is

$$\begin{aligned} V\left(D\mid\theta(N),\theta(D)\right) &= \max_{\theta^i\geq 0}\left\{-\theta^i+p\left(\theta^i,\theta(D)\mid D\right)\left(\frac{R^rL}{M}+\beta V\left(N\mid\theta(N),\theta(D)\right)\right)\right. \\ &\quad \left. +\left(1-p\left(\theta^i,\theta(D)\mid D\right)\right)\left(\frac{R^cL}{M}+\beta V\left(D\mid\theta(N),\theta(D)\right)\right)\right\} \end{aligned} \tag{21.19}$$

which has first-order necessary condition

$$\phi f\left(\phi\left((M-1)\theta(D)+\theta^i\right)-\eta\right)\left(\frac{\Delta RL}{M}+\beta\left(V\left(N\mid\theta(N),\theta(D)\right)-V\left(D\mid\theta(N),\theta(D)\right)\right)\right)\leq 1, \tag{21.20}$$

and $\theta^i \geq 0$, again with complementary slackness and with second-order condition $f'\left(\phi\left((M-1)\theta(N)+\theta^i\right)-\eta\right)<0$. Denote the policy function (correspondence) implied by the maximization in (21.19) by $\Gamma^D[\theta(N),\theta(D)]$, so that any $\theta^i \in \Gamma^D[\theta(N),\theta(D)]$ solves (21.20).

Consequently, denoting the decision of current economic institutions by $\tau(\pi)$ and future political system by $s'(\pi)$, we can have the following definition of a symmetric MPE:

Definition 21.1. *A symmetric MPE consists of a pair of contribution levels for elite agents $\theta(N)$ and $\theta(D)$, such that $\theta(N) \in \Gamma^N[\theta(N),\theta(D)]$ and similarly $\theta(D) \in \Gamma^D[\theta(N),\theta(D)]$. In addition, economic and political decisions $\tau(\pi)$ and $s'(\pi)$ are such that $\tau(\pi=0)=0$, $s'(\pi=0)=N$, $\tau(\pi=1)=1$ and $s'(\pi=1)=D$, and factor prices are given by (21.8)-(21.11) as a function of $\tau \in \{0,1\}$.*

This definition incorporates the best responses of elites and citizens regarding economic and political institutional, $\tau(\pi)$ and $s'(\pi)$, for convenience. Notice also that we have used the tie-breaking rule that when in different, citizens prefer democracy.

This definition highlights that the main economic actions, in particular, the investments in de facto power, are taken by elite agents, so the characterization of the MPE will involve solving for their optimal behavior.

In a symmetric MPE, θ^i that solves (21.17) must equal $\theta(N)$, thus when strictly positive, $\theta(N)$, must be given by:

$$\phi f(\phi M\theta(N))\left(\frac{\Delta RL}{M}+\beta V(N\mid\theta(N),\theta(D))-\beta V(D\mid\theta(N),\theta(D))\right)=1, \tag{21.21}$$

and similarly the equilibrium condition for $\theta(D)$ (when strictly positive) is

$$\phi f(\phi M\theta(D)-\eta)\left(\frac{\Delta RL}{M}+\beta V(N\mid\theta(N),\theta(D))-\beta V(D\mid\theta(N),\theta(D))\right)=1. \tag{21.22}$$

Given Definition 21.1, these two equations completely characterize symmetric MPEs with $\theta(N)>0$ and $\theta(D)>0$.

Comparison of (21.21) and (21.22) immediately implies that

$$\theta(D)=\theta(N)+\frac{\eta}{\phi M}. \tag{21.23}$$

Moreover inspection of (21.21) and (21.22), combined with the fact that F is continuously differentiable, yields the *invariance* result:

$$p(D)\equiv p(\theta(D),\theta(D)\mid D)=p(\theta(N),\theta(N)\mid N)\equiv p(N), \tag{21.24}$$

which also defines $p(D)$ and $p(N)$ as the respective probabilities of the elite gaining (or maintaining) political power in democracy and nondemocracy.

Intuitively, in democracy the elite invest sufficiently more to increase their de facto political power that they entirely offset the advantage of the citizens coming from their de jure power. A more technical intuition for this result is that the optimal contribution conditions for elite agents both in nondemocracy and democracy equate the marginal cost of contribution, which is always equal to 1, to the marginal benefit. Since the marginal costs are equal, equilibrium benefits in the two regimes also have to be equal. The marginal benefits consist of the immediate gain of economic rents, $\Delta RL/M$, plus the gain in continuation value, which is also independent of current regime. Consequently, marginal costs and benefits can only be equated if $p(D) = p(N)$ as in (21.24).

It is also straightforward to specify when there will be positive investment in de facto power. In particular, the following assumption is sufficient to ensure that the equilibrium will have positive contribution by elite agents to de facto power:

Assumption 3:

$$\min\left\{\phi f(0)\frac{\Delta RL}{M}, \phi f(-\eta)\frac{\Delta RL}{M}\right\} > 1.$$

Since $V(N) - V(D) \geq 0$ (by virtue of the fact that the elite choose nondemocracy), this assumption ensures that in both regimes, an individual would like to make a positive contribution even if nobody else is doing so. [Assumption 3 also implies that $\eta < -\underline{\omega}$ (where recall that $\underline{\omega} < 0$); see condition (21.26) below.] If this assumption is not satisfied, there may also exist equilibria in which the elite make no contribution to increasing their de facto power (see Corollary 21.1).

Proposition 21.1. ***(Invariance)*** Suppose Assumptions 1-3 hold. Then in the baseline model, there exists a unique symmetric MPE. This equilibrium involves $p(D) = p(N) \in (0,1)$, so that the probability distribution over economic institutions is non-degenerate and independent of whether the society is democratic or nondemocratic.

Proof. Assumption 3 ensures that $\theta(D) = 0$ and $\theta(N) = 0$ cannot be part of an equilibrium. Since Assumption 2 implies that $f(\omega)$ is continuous and $\lim_{\omega\to\infty} f(\omega) = 0$, both conditions (21.21) and (21.22) must hold as equalities for some interior values of $\theta(D)$ and $\theta(N)$, establishing existence. The result that $p(D) = p(N) > 0$ then follows immediately from the comparison of these two equalities, which establishes (21.24). The fact that $p(D) = p(N) < 1$ follows from Assumption 2, which imposes that F is strictly increasing throughout its support, so for any interior $\theta(D)$ and $\theta(N)$, $F(\phi M\theta(D) - \eta) = F(\phi M\theta(N)) < 1$. In addition, again from Assumption 2, $f(\omega)$ is single peaked, so only a unique pair of $\theta(D)$ and $\theta(N)$ could satisfy (21.21) and (21.22) with $f'(\phi M\theta(N)) < 0$ and $f'(\phi M\theta(D) - \eta) < 0$

for given $V(N) - V(D)$. The fact that $V(N) - V(D) = \theta(N) - \theta(D) = \eta/(\phi M)$ is uniquely determined (from equation (21.24)) then establishes the uniqueness of the symmetric MPE. □

This proposition is one of the main results of the paper. It shows that there will be equilibrium changes from democracy to nondemocracy and the other way round (this follows from the fact that the equilibrium probability distribution is non-degenerate, i.e., $p(D) = p(N) \in (0, 1)$). Moreover, by assumption these changes in political institutions affect the distribution of de jure power, but they do *not* translate into changes in the law of motion of economic institutions and economic allocations, i.e., we have $p(D) = p(N)$. This is the sense in which there is *invariance* in equilibrium; even when shocks change the political institutions, the probability distribution over equilibrium economic institutions remains unchanged. This result also illustrates how institutional change and persistence can coexist—while political institutions change frequently, the equilibrium process for economic institutions remains unchanged.

REMARK 1. *As will be discussed further below, the invariance result relies on functional form assumptions. Further analysis below will show that when there are differences in the technology of generating de facto power for the elite in democracy and nondemocracy or when economic institutions are costly to change in the short run, de facto power will only offset the change in de jure power partially. Other assumptions implicit in our analysis that are important for the invariance result are: (1) that democracy shifts the power of the citizens additively (rather than ω being drawn from general distributions F_N in nondemocracy and F_D in democracy, with F_D first-order stochastically dominating F_N); (2) that the technology of de facto power for the elite, equation (21.13), is linear. When either of these assumptions are relaxed, we continue to obtain the general insight that endogenous changes in de facto power (at least partially) offset declines in the de jure power of the elite, but do not necessarily get the invariance result.*

REMARK 2. *Assumptions 2 and 3 can be relaxed without affecting the basic conclusions in Proposition 21.1. For example, if we relax the single-peakedness assumption on $f(\omega)$, the conclusions in Proposition 21.1 would continue to apply, except that the symmetric MPE may no longer be unique. Multiple equilibria here are of potential interest, as they correspond to situations in which expectations of future behavior affects current behavior (see, e.g., Hassler et al., 2003). Also, if the parts of Assumption 2 that F is increasing everywhere and $\lim_{\omega \to \infty} f(\omega) = 0$ are relaxed, we may obtain corner solutions, whereby $p(N) = p(D) = 1$,* *and there would be no transitions to democracy from nondemocracy (essentially because returns to individual elites from investing in de facto power may remain high, while the probability of a sufficiently high level of ω becomes 0). Alternatively, if Assumption 3 is relaxed, we can have equilibria with $p(N) = p(D) = 0$. Assumptions 2 and 3 rule out these "corner" equilibria. The following result is interesting in this context.*

Corollary 21.1. *Suppose there exists $\bar{\theta}(N) > 0$ such that*

$$\phi f(\phi M\bar{\theta}(N)) \left(\frac{\Delta RL/M - \beta\bar{\theta}(N)}{1 - \beta F(\phi M\bar{\theta}(N))} \right) = 1, \tag{21.25}$$

that Assumptions 1 and 2 hold, and that

$$\eta > -\underline{\omega}. \tag{21.26}$$

Then in the baseline model, there exists a symmetric MPE in which $p(N) \in (0, 1)$ and $p(D) = 0$.

Proof. Suppose there exists a symmetric MPE with $p(D) = 0$. Then we have $V(D) = R^c L / ((1 - \beta) M)$, while $V(N)$ is still given by (21.16), and the relevant first-order necessary condition for $\theta(N) > 0$ is given by (21.21). Combining this with the expression for $V(D)$, we obtain $\theta(N) = \bar{\theta}(N)$ as in (21.25), and

$$V(N) - V(D) = \frac{F(\phi M\bar{\theta}(N))\Delta RL/M - \bar{\theta}(N)}{1 - \beta F(\phi M\bar{\theta}(N))}.$$

Now using (21.21) and (21.22), we see that (21.25) is sufficient to ensure that positive contribution to de facto power in nondemocracy is optimal for elite agents. Moreover, (21.26) implies that $f(-\eta) = 0$, thus

$$\phi f(-\eta) \left(\frac{\Delta RL/M - \beta\bar{\theta}(N)}{1 - \beta F(\phi M\bar{\theta}(N))} \right) < 1,$$

so that zero contribution in democracy is also optimal for the elite. Moreover, again from (21.26), $F(-\eta) = 0$, which establishes the existence of a symmetric MPE with $p(N) \in (0, 1)$ and $p(D) = 0$. □

Therefore, if we relax part of Assumption 3, symmetric MPEs with democracy as an absorbing state may arise. Clearly, Condition (21.26), which leads to this outcome, is more likely to hold when η is high. This implies that if democracy in fact creates a substantial advantage in favor of the citizens, it may destroy the incentives of the elite to engage in activities that increase their de facto power, and thus change the future distribution of political regimes and economic institutions.

It is also interesting to note that even when Condition (21.26) holds, the equilibrium with $p(D) = p(N) > 0$ characterized in Proposition 21.1 may still exist, leading to a symmetric MPE with $p(D) = p(N)$. Consequently, whether democracy becomes an absorbing state (i.e., fully consolidated), may depend on expectations.

Finally, inspection of the proof of Corollary 21.1 shows that Assumption 3 can be relaxed to:

Assumption 3A: There exists $\bar{\theta}(N) > 0$ satisfying (21.25), and

$$\phi f(-\eta)\left(\frac{\Delta RL/M - \beta\bar{\theta}(N)}{1-\beta F(\phi M\bar{\theta}(N))}\right) > 1.$$

With this modified assumption, all the results continue to hold, though we prefer Assumption 3 since, despite being more restrictive, it is simpler and more transparent.

21.2.2. Non-Symmetric MPE. We now show that the same invariance result obtains without the restriction to symmetric MPE. To do this, we first extend our treatment above and define an MPE more generally. Without symmetry, the power of the elite in nondemocracy as a function of contribution θ^i by agent $i \in \mathcal{E}$ and the distribution of contributions by all other agents, $\theta^{-i}(N) \equiv \left\{\theta^j(N)\right\}_{j\in\mathcal{E}, j\neq i}$, is given by

$$P^E\left(\theta^i, \theta^{-i}(N) \mid N\right) = \phi\left(\sum_{j\in\mathcal{E}, j\neq i}\theta^j(N) + \theta^i\right).$$

Similar to before, in nondemocracy the elite will have political power with probability

$$p\left(\theta^i, \theta^{-i}(N) \mid N\right) = F\left(\phi\left(\sum_{j\in\mathcal{E}, j\neq i}\theta^j(N) + \theta^i\right)\right). \tag{21.27}$$

In democracy, with the same reasoning as before, this probability is given by

$$p\left(\theta^i, \theta^{-i}(D) \mid D\right) = F\left(\phi\left(\sum_{j\in\mathcal{E}, j\neq i}\theta^j(D) + \theta^i\right) - \eta\right). \tag{21.28}$$

The possibility that different individuals will contribute different amounts to the de facto power of the elite implies that value functions can also differ across elite agents, and must also be indexed by i. Therefore, the net present discounted value of agent $i \in \mathcal{E}$ is

$$\begin{aligned} &V^i\left(N \mid \theta^{-i}(N), \theta^{-i}(D)\right) \\ = \;& \max_{\theta^i \geq 0}\left\{-\theta^i + p\left(\theta^i, \theta^{-i}(N) \mid N\right)\left(\frac{R^r L}{M} + \beta V^i\left(N \mid \theta^{-i}(N), \theta^{-i}(D)\right)\right)\right. \\ &\left. + \left(1 - p\left(\theta^i, \theta^{-i}(N) \mid N\right)\right)\left(\frac{R^c L}{M} + \beta V^i\left(D \mid \theta^{-i}(N), \theta^{-i}(D)\right)\right)\right\}. \end{aligned} \tag{21.29}$$

Here $V^i\left(N \mid \theta^{-i}(N), \theta^{-i}(D)\right)$ denotes the value of agent i in nondemocracy when all other elite agents choose contributions $\theta^{-i}(N)$ in nondemocracy and $\theta^{-i}(D)$ in democracy. Similarly, $V^i\left(D \mid \theta^{-i}(N), \theta^{-i}(D)\right)$ is the corresponding value in democracy for agent i. The intuition for this equation is similar to that for (21.16) in the symmetric case.

Agent $i \in \mathcal{E}$ chooses θ^i to maximize his net expected present discounted utility. Let the policy function (correspondence) of agent i that solves the maximization in (21.29) be given by $\Gamma_i^N\left[\theta^{-i}(N), \theta^{-i}(D)\right]$, so that any $\theta^i \in \Gamma_i^N\left[\theta^{-i}(N), \theta^{-i}(D)\right]$ is an optimal policy for the value function in (21.29). Similarly, we have

$$
\begin{aligned}
(21.30) \qquad & V^i\left(D \mid \theta^{-i}(N), \theta^{-i}(D)\right) \\
= & \max_{\theta^i \geq 0}\Bigg\{-\theta^i + p\left(\theta^i, \theta^{-i}(D) \mid D\right)\left(\frac{R^r L}{M} + \beta V^i\left(N \mid \theta^{-i}(N), \theta^{-i}(D)\right)\right) \\
& + \left(1 - p\left(\theta^i, \theta^{-i}(D) \mid D\right)\right)\left(\frac{R^c L}{M} + \beta V^i\left(D \mid \theta^{-i}(N), \theta^{-i}(D)\right)\right)\Bigg\},
\end{aligned}
$$

and let the set of maximizers of this problem be $\Gamma_i^D\left[\theta^{-i}(N), \theta^{-i}(D)\right]$. Then we have the more general definition of MPE as:

Definition 21.2. *An MPE consists of a pair of contribution distributions for elite agents* $\left\{\theta^i(N)\right\}_{i\in\mathcal{E}}$ *and* $\left\{\theta^i(D)\right\}_{i\in\mathcal{E}}$, *such that for all* $i \in E$, $\theta^i(N) \in \Gamma_i^N\left[\theta^{-i}(N), \theta^{-i}(D)\right]$ *and similarly* $\theta^i(D) \in \Gamma_i^D\left[\theta^{-i}(N), \theta^{-i}(D)\right]$. *In addition, economic and political decisions* $\tau(\pi)$ *and* $s'(\pi)$ *are such that* $\tau(\pi = 0) = 0$, $s'(\pi = 0) = N$, $\tau(\pi = 1) = 1$ *and* $s'(\pi = 1) = D$, *and factor prices are given by (21.8)-(21.11) as a function of* $\tau \in \{0, 1\}$.

Proposition 21.2. ***(Non-Symmetric MPE and Invariance)*** *Suppose Assumptions 1-3 hold. Then in the baseline model, any MPE involves* $p(D) = p(N) \in (0, 1)$.

Proof. Let us first define

$$
\Delta V^i\left(\theta^{-i}(N), \theta^{-i}(D)\right) \equiv V^i\left(N \mid \theta^{-i}(N), \theta^{-i}(D)\right) - V^i\left(D \mid \theta^{-i}(N), \theta^{-i}(D)\right).
$$

From the recursive formulations in (21.29) and (21.30), for all $i \in \mathcal{E}$ we have the first-order necessary conditions:

$$
(21.31) \quad \phi f\left(\phi\left(\sum_{j\in\mathcal{E}, j\neq i} \theta^j(N) + \theta^i\right)\right)\left(\frac{\Delta R L}{M} + \beta \Delta V^i\left(\theta^{-i}(N), \theta^{-i}(D)\right)\right) \leq 1 \text{ and } \theta^i \geq 0,
$$

(21.32)

$$
\phi f\left(\phi\left(\sum_{j\in\mathcal{E}, j\neq i} \theta^j(D) + \theta^i - \eta\right)\right)\left(\frac{\Delta R L}{M} + \beta \Delta V^i\left(\theta^{-i}(N), \theta^{-i}(D)\right)\right) \leq 1 \text{ and } \theta^i \geq 0,
$$

both holding with complementary slackness.

The proof proceeds in several steps. First, Assumption 3 implies that $\theta^i(N) = 0$ and $\theta^i(D) = 0$ for all $i \in \mathcal{E}$ cannot be an equilibrium. Therefore there must exist some $i' \in \mathcal{E}$ such that $\theta^{i'}(N) > 0$ and $i'' \in \mathcal{E}$ such that $\theta^{i''}(D) > 0$.

Second, we claim that there must exist some $i \in \mathcal{E}$ for whom both (21.31) and (21.32) hold as equalities. Suppose not. Then it must be the case that for i' and i'' defined in the previous paragraph, we have, respectively, (21.31) and (21.32) holding as equalities and (21.32) and (21.31) are slack. This implies

$$\begin{aligned} 1 &= \phi f\left(\phi\left(\sum_{j\in\mathcal{E},j\neq i'}\theta^j(N)+\theta^{i'}\right)\right)\left(\frac{\Delta RL}{M}+\beta\Delta V^{i'}\left(\theta^{-i'}(N),\theta^{-i'}(D)\right)\right) \\ &> \phi f\left(\phi\left(\sum_{j\in\mathcal{E},j\neq i'}\theta^j(D)+\theta^{i'}-\eta\right)\right)\left(\frac{\Delta RL}{M}+\beta\Delta V^{i'}\left(\theta^{-i'}(N),\theta^{-i'}(D)\right)\right)\end{aligned}$$

or

$$\begin{aligned} f\left(\phi\left(\sum_{j\in\mathcal{E},j\neq i'}\theta^j(N)+\theta^{i'}\right)\right) &= f\left(\phi\left(\sum_{j\in\mathcal{E}}\theta^j(N)\right)\right) \\ &> f\left(\phi\left(\sum_{j\in\mathcal{E}}\theta^j(D)-\eta\right)\right) = f\left(\phi\left(\sum_{j\in\mathcal{E},j\neq i'}\theta^j(D)+\theta^{i'}-\eta\right)\right).\end{aligned}$$

Similarly for i'',

$$\begin{aligned} f\left(\phi\left(\sum_{j\in\mathcal{E},j\neq i''}\theta^j(N)+\theta^{i''}\right)\right) &= f\left(\phi\left(\sum_{j\in\mathcal{E}}\theta^j(N)\right)\right) \\ &< f\left(\phi\left(\sum_{j\in\mathcal{E}}\theta^j(D)-\eta\right)\right) = f\left(\phi\left(\sum_{j\in\mathcal{E},j\neq i''}\theta^j(D)+\theta^{i''}-\eta\right)\right),\end{aligned}$$

yielding a contradiction with the previous inequality.

Third, the fact that there exists some $i \in \mathcal{E}$ for whom both (21.31) and (21.32) hold as equalities implies that

$$f\left(\phi\left(\sum_{j\in\mathcal{E}}\theta^j(N)\right)\right) = f\left(\phi\left(\sum_{j\in\mathcal{E}}\theta^j(D)-\eta\right)\right).$$

This is clearly only possible if

$$\phi\sum_{i\in\mathcal{E}}\theta^i(N) = \phi\sum_{i\in\mathcal{E}}\theta^i(D)-\eta,$$

which implies $p(D) = p(N) > 0$. The fact that $p(D) = p(N) < 1$ again follows from Assumption 2, completing the proof.

Finally, for use in the proof of Corollary 21.2, also note that this argument establishes that for any $i \in \mathcal{E}$ for whom (21.31) or (21.32) holds as equality, both of these equations must hold as equality. $\square$

The only difference between symmetric and non-symmetric MPE is that in symmetric MPE we also know that the total contributions made by the elite will be equally divided among each elite agent. In non-symmetric MPE, this may not be the case, and depending on expectations, some elite agents may be expected to, and consequently do, contribute more than others. This implies that in non-symmetric MPE, different levels of $p(D) = p(N)$ can arise in equilibrium.

Nevertheless, the important conclusion that the probability of the elite dominating political power and imposing their favorite economic institutions is independent of the underlying form of political institutions remains. Given this result, in the rest of the paper we focus on symmetric MPE. Before doing so, however, we can also note the following result:

COROLLARY 21.2. *Among non–symmetric MPEs, the following maximizes $p(N) = p(D)$: for $i' \in \mathcal{E}$, $\theta^{i'}(D) = \theta^{i'}(N) + \eta/\phi$, and for all $i \in \mathcal{E}$ and $i \neq i'$, $\theta^{i}(D) = \theta^{i}(N) = 0$.*

PROOF. The proof of Proposition 21.2 makes it clear that in any equilibrium where condition (21.31) holds as equality for some $i \in \mathcal{E}$, so does (21.32) for the same $i \in \mathcal{E}$, and vice versa. This implies that to maximize $p(N) = p(D)$, we need to maximize $\Delta V^{i}\left(\theta^{-i}(N), \theta^{-i}(D)\right)$ for all $i \in \mathcal{E}$ for whom (21.31) and (21.32) hold as equalities. Clearly, the highest value any such $\Delta V^{i}\left(\theta^{-i}(N), \theta^{-i}(D)\right)$ can take is given by $\Delta V^{i'}\left(\theta^{-i'}(N), \theta^{-i'}(D)\right) = \eta/\phi$ for some i' together with $\theta^{i'}(D) = \theta^{i'}(N) + \eta/\phi$. To see that an equilibrium with $\theta^{i'}(D) = \theta^{i'}(N) + \eta/\phi$ for i' and $\theta^{i}(D) = \theta^{i}(N) = 0$ for all $i \in \mathcal{E}$ and $i \neq i'$ exists, note first that, since $\theta^{i}(D) = \theta^{i}(N) = 0$ for all $i \in \mathcal{E}$ and $i \neq i'$, we have $\Delta V^{i}\left(\theta^{-i}(N), \theta^{-i}(D)\right) = 0$ for all $i \in \mathcal{E}$ and $i \neq i'$. Second, from (21.31) and (21.32), we have that for $i' \in \mathcal{E}$,

$$\phi f\left(\phi\left(\theta^{i'}(N)\right)\right) = \phi f\left(\phi\left(\theta^{i'}(D)\right) - \eta\right) = \left(\frac{\Delta RL}{M} + \beta \Delta V^{i'}\left(\theta^{-i'}(N), \theta^{-i'}(D)\right)\right)^{-1},$$

which, in view of the fact that $\Delta V^{i'}\left(\theta^{-i'}(N), \theta^{-i'}(D)\right) > 0 = V^{i}\left(\theta^{-i}(N), \theta^{-i}(D)\right)$, implies

$$\phi f\left(\phi\left(\theta^{i'}(N)\right)\right) = \phi f\left(\phi\left(\theta^{i'}(D)\right) - \eta\right) < \left(\frac{\Delta RL}{M} + \beta \Delta V^{i}\left(\theta^{-i}(N), \theta^{-i}(D)\right)\right)^{-1},$$

for all other $i \in \mathcal{E}$, establishing that they prefer to make zero contributions. Hence, $\theta^{i'}(D) = \theta^{i'}(N) + \eta/\phi$ for i' and $\theta^{i}(D) = \theta^{i}(N) = 0$ for all $i \in \mathcal{E}$ and $i \neq i'$ is an equilibrium and achieves the highest $p(N) = p(D)$ among all MPEs. □

Intuitively, the equilibrium that makes only one elite agent make all of the investment in de facto power means that this agent has a lot to lose from democracy (because of the higher investment in de facto power involved in this regime), and maximizes investments in de facto power.

21.2.3. Subgame Perfect Equilibria. The analysis so far has focused on MPE. Since the landed elite form a small cohesive group, they may be able to achieve a better equilibrium than the MPE by using threats of punishments against each other. Here, we briefly discuss SPEs (subgame perfect equilibria) of the above game. The main result is that for sufficiently large discount factors, the "best" SPEs also take the same form as the MPE characterized in Propositions 21.1 and 21.2, with the equilibrium probability distribution over economic institutions independent of the political regime.

In characterizing the SPEs, we allow elite agents to use any kind of punishment strategies and coordination, except that in competitive labor markets, they cannot (by law) restrict their labor demand in order to affect factor prices. To define an SPE, let $\theta(s_t, t) \equiv \left\{\theta^j(s_t, t)\right\}_{j \in \mathcal{E}}$ be the vector of contributions by elite agents at time t, when the state is s_t. Let $h^t = (\theta(s_0, 0), \pi_0, \tau_0, s_1, ..., \theta(s_t, t), \pi_t, \tau_t, s_{t+1})$ be the history of contributions, political outcomes and actions up to time t, with $\mathcal{H}^t$ denoting the set of possible histories at time t.

Definition 21.3. *An SPE consists of contribution functions for each elite agent $i \in \mathcal{E}$, $\theta^i : \mathbb{Z}_+ \times \{N, D\} \times \mathcal{H}^{t-1} \to \mathbb{R}_+$ specifying their contribution as a function of time t, state s_t and the history $\mathcal{H}^{t-1}$ up to that point and economic and political decision functions $\tau : \mathbb{Z}_+ \times \{0,1\} \times \{N, D\} \times \mathcal{H}^{t-1} \to \{0,1\}$ and $s' : \mathbb{Z}_+ \times \{0,1\} \times \{N, D\} \times \mathcal{H}^{t-1} \to \{N, D\}$ specifying economic and political institution decisions as a function of time, who has political power, the state and history, such that θ^i is a best response to θ^{-i}, τ and s' for all $i \in \mathcal{E}$ and τ and s' are best responses to $\left\{\theta^i\right\}_{i \in \mathcal{E}}$, and factor prices are given by (21.8)-(21.11) as a function of $\tau \in \{0,1\}$.*

As in most repeated and dynamic games, there exist many subgame perfect equilibria in this model. Our focus is on SPEs that maximize the ex ante—time $t = 0$—utility of the elite. This is natural since we are motivated to look at the SPEs to see how the ability of the elite to coordinate their actions changes the results. We define best or "Pareto optimal" SPEs as SPEs in which, at time $t = 0$, no elite agent can be made better off without some other elite agent being made worse off. In addition, we refer to an SPE as "symmetric Pareto optimal", if it is Pareto optimal and all elite agents use the same equilibrium strategy. The main result is the following:

Proposition 21.3. ***(Subgame Perfect Equilibrium and Invariance)*** *Suppose Assumptions 1-3 hold. Then there exists $\bar{\beta} \in [0,1)$ such that that for all $\beta \geq \bar{\beta}$, the* *symmetric Pareto optimal SPE induces equilibrium probabilities of labor repressive institutions $p(D) = p(N) \in (0,1)$. Moreover, as $\beta \to 1$, any Pareto optimal SPE involves $p(D) = p(N) \in (0,1)$.*

PROOF. We will prove this proposition by showing that for any Pareto optimal distribution of contributions among elite agents, there exists a $\tilde{\beta} \in [0,1)$, such that this distribution can be supported as an SPE and involves equilibrium probabilities $p(D) = p(N) \in (0,1)$. The special case of equal contributions will establish the first part of the proposition, and the fact that as $\beta \to 1$ this is true for any distribution establishes the second part.

First suppose that a single individual controls all contributions by elite agents to de facto power. With the same arguments as above, the problem of this individual can be written recursively as:

$$\tilde{V}(N) = \max_{\theta \geq 0} \left\{ -\theta + F(\phi\theta) \left(\frac{R^r L}{M} + \beta \tilde{V}(N) \right) + (1 - F(\phi\theta)) \left(\frac{R^c L}{M} + \beta \tilde{V}(D) \right) \right\},$$

$$\tilde{V}(D) = \max_{\theta \geq 0} \left\{ -\theta + F(\phi\theta - \eta) \left(\frac{R^r L}{M} + \beta \tilde{V}(N) \right) + (1 - F(\phi\theta - \eta)) \left(\frac{R^c L}{M} + \beta \tilde{V}(D) \right) \right\},$$

where now $\theta = \sum_{i \in \mathcal{E}} \theta^i$ is the total contribution by elite agents, and these expressions only differ from (21.16) and (21.19) because the entire cost of contributions and the entire benefit in terms of land rents are taken into account. Denoting optimal choices in this program by $\theta^*(N)$ and $\theta^*(D)$, the first-order conditions are

$$\phi f(\phi\theta^*(N)) \left(\frac{\Delta RL}{M} + \beta \left(\tilde{V}(N) - \tilde{V}(D) \right) \right) \leq 1 \text{ and } \theta^*(N) \geq 0$$

$$\phi f(\phi\theta^*(D) - \eta) \left(\frac{\Delta RL}{M} + \beta \left(\tilde{V}(N) - \tilde{V}(D) \right) \right) \leq 1 \text{ and } \theta^*(D) \geq 0,$$

holding with complementary slackness. Assumption 3 ensures that $\theta^*(N)$ and $\theta^*(D)$ have to be positive, so the solution to this problem involves

$$\phi\theta^*(N) = \phi\theta^*(D) - \eta \tag{21.33}$$

and thus $p(D) = p(N) > 0$. That $p(D) = p(N) < 1$ again follows from Assumption 2. It is clear that an equilibrium satisfying the above conditions would be Pareto optimal from the ex ante viewpoint of elite agents (for some distribution of the total contributions $\theta^*(N)$ and $\theta^*(D)$ across the elite agents), since no elite agent can be made better off without some other elite agent being made worse off.

Next we show that there exists $\tilde{\beta} \in [0,1)$ such that for $\beta \geq \tilde{\beta}$, any distribution of contributions $\theta^*(N)$ and $\theta^*(D)$ across elite agents can be supported as an SPE. To simplify the notation, consider a stationary distribution (though, with more notation, the argument easily generalizes to any time-varying distribution): $\left\{\theta^{i*}(N)\right\}_{i\in\mathcal{E}}$ and $\left\{\theta^{i*}(D)\right\}_{i\in\mathcal{E}}$. Consider a candidate SPE with this distribution and the feature that following a deviation, the equilibrium reverts back to an MPE. Recall from Proposition 21.2 that in any MPE we have $\tau(\pi=0)=0$ and $s'(\pi=0)=N$, and some contributions denoted by $\left\{\theta^{ip}(N)\right\}_{i\in\mathcal{E}}$, $\left\{\theta^{ip}(D)\right\}_{i\in\mathcal{E}}$ such that $\theta^{p}(N)\equiv\sum_{i\in\mathcal{E}}\theta^{ip}(N)=\theta^{p}(D)-\eta/\phi\equiv\sum_{i\in\mathcal{E}}\theta^{ip}(D)-\eta/\phi$. Given this punishment structure following a deviation, the best deviation for any agent i is to contribute nothing. By comparing the above single-agent maximization problem to Corollary 21.2, we have $\theta^{*}(N)>\theta^{p}(N)$ and $\theta^{*}(D)>\theta^{p}(D)$. Let $h^{t}=\hat{h}^{t}$ denote the history in which in all past periods, all agents have played $\left\{\theta^{i*}(N)\right\}_{i\in\mathcal{E}}$, $\left\{\theta^{i*}(D)\right\}_{i\in\mathcal{E}}$, $\tau(\pi=0)=0$ and $s'(\pi=0)=N$.

Now consider the following strategy profile to support the SPE: if $h^{t}=\hat{h}^{t}$, then the equilibrium specifies that elite agents play $\left\{\theta^{i*}(N)\right\}_{i\in\mathcal{E}}$, $\left\{\theta^{i*}(D)\right\}_{i\in\mathcal{E}}$, $\tau(\pi=0)=0$ and $s'(\pi=0)=N$. If $h^{t}\neq\hat{h}^{t}$ (i.e., if there has been a deviation at some date $t'<t$ from this play), then $\theta^{i}(t,s_{t})=\theta^{ip}(s_{t})$ for all $i\in\mathcal{E}$ and all $s_{t}\in\{N,D\}$, $\tau(\pi=0)=0$ and $s'(\pi=0)=N$.

With this strategy profile, following a deviation, say starting in nondemocracy, elite agent i will obtain:

(21.34)

$$V_{i}^{d}(N)=F\left(\phi\sum_{j\in\mathcal{E},j\neq i}\theta^{j*}(N)\right)\left(\frac{R^{r}L}{M}+\beta\hat{V}_{i}(N)\right)+\left(1-F\left(\phi\sum_{j\in\mathcal{E},j\neq i}\theta^{j*}(N)\right)\right)\left(\frac{R^{c}L}{M}+\beta\hat{V}_{i}(D)\right),$$

where $\phi\sum_{j\in\mathcal{E},j\neq i}\theta^{j*}(N)$ is the total contributions to the de facto power after the deviation, and $\hat{V}_{i}(N)$ and $\hat{V}_{i}(D)$ are the MPE values in the punishment phase following the deviation,

(21.35)

$$\hat{V}_{i}(N)=-\theta^{ip}(N)+F(\phi\theta^{p}(N))\left(\frac{R^{r}L}{M}+\beta\hat{V}_{i}(N)\right)+(1-F(\phi\theta^{p}(N)))\left(\frac{R^{c}L}{M}+\beta\hat{V}_{i}(D)\right),$$

and

(21.36)

$$\hat{V}_{i}(D)=-\theta^{ip}(D)+F(\phi\theta^{p}(D)-\eta)\left(\frac{R^{r}L}{M}+\beta\hat{V}_{i}(N)\right)+(1-F(\phi\theta^{p}(D)-\eta))\left(\frac{R^{c}L}{M}+\beta\hat{V}_{i}(D)\right).$$

The value of deviation in state $s=D$, $V_{i}^{d}(D)$ is defined similarly to (21.34). Since $\theta^{*}(N)=\theta^{*}(D)-\eta/\phi>\theta^{p}(N)=\theta^{p}(D)-\eta/\phi$, we have $F(\phi\theta^{*}(N))=F(\phi\theta^{*}(D)-\eta)>F(\phi\theta^{p}(N))=F(\phi\theta^{p}(D)-\eta)$.

If, on the other hand, this agent follows the SPE strategy of "cooperating", i.e., contributing $\theta^{i*}(N)$ when $s=N$ and $\theta^{i*}(D)$ when $s=D$, he will obtain

(21.37)

$$V_{i}^{c}(N)=-\theta^{i*}(N)+F(\phi\theta^{*}(N))\left(\frac{R^{r}L}{M}+\beta V_{i}^{c}(N)\right)+(1-F(\phi\theta^{*}(N)))\left(\frac{R^{c}L}{M}+\beta V_{i}^{c}(D)\right).$$

Similarly,

(21.38)

$$V_i^c(D) = -\theta^{i*}(D) + F(\phi\theta^*(D) - \eta)\left(\frac{R^r L}{M} + \beta V_i^c(N)\right) + (1 - F(\phi\theta^*(D) - \eta))\left(\frac{R^c L}{M} + \beta V_i^c(D)\right),$$

and also for $V_i^d(D)$ and $\hat{V}_i(D)$. Choose an MPE for the punishment phase such that $\theta^{i*}(N) - \theta^{i*}(D) > \theta^{ip}(N) - \theta^{ip}(D)$ for $s \in \{D, N\}$ for all $i \in \mathcal{E}' \subset \mathcal{E}$ such that $\mathcal{E}' = \{i \in \mathcal{E} : \theta^{i*}(N) > 0$ or $\theta^{i*}(D) > 0\}$. Such an MPE can always be constructed in view of the fact that $\theta^*(N) = \theta^*(D) - \eta/\phi > \theta^p(N) = \theta^p(D) - \eta/\phi$. Now each agent would be happy to follow the SPE strategy as long as

$$V_i^c(D) \geq V_i^d(D) \text{ and } V_i^c(N) \geq V_i^d(N). \tag{21.39}$$

These inequalities are naturally satisfied for all $i \in \mathcal{E} \backslash \mathcal{E}'$ (since any such agent has no incentive to deviate because he is making zero contributions along the equilibrium path), so we only need to check them for $i \in \mathcal{E}'$. Next combining (21.37) and (21.38) and using the fact that, from (21.33), $F(\phi\theta^*(N)) = F(\phi\theta^*(D) - \eta)$, we have

$$V_i^c(N) = -\theta^{i*}(N) + F(\phi\theta^*(N))\left(\frac{\Delta RL}{M} + \beta\left[\theta^{i*}(D) - \theta^{i*}(N)\right]\right), \tag{21.40}$$

and also

$$\hat{V}_i(N) = -\theta^{ip}(N) + F(\phi\theta^p(N))\left(\frac{\Delta RL}{M} + \beta\left(\theta^{ip}(D) - \theta^{ip}(N)\right)\right). \tag{21.41}$$

By virtue of the fact that $F(\phi\theta^*(N)) = F(\phi\theta^*(D) - \eta) > F(\phi\theta^p(N)) = F(\phi\theta^p(D) - \eta)$ and $\theta^{i*}(D) - \theta^{i*}(N) > \theta^{ip}(D) - \theta^{ip}(N)$, the comparison of (21.40) and (21.41) shows that $V_i^c(D)$ and $V_i^c(N)$ increase faster in β than $\hat{V}_i(N)$ and $\hat{V}_i(D)$, and by implication, than $V_i^d(D)$ and $V_i^d(N)$. Moreover, the same observation implies that for any $\left\{\theta^{i*}(N), \theta^{i*}(D)\right\}$ (and associated appropriate punishment MPE, $\left\{\theta^{ip}(N), \theta^{ip}(D)\right\}_{i\in\mathcal{E}}$), we have that $\lim_{\beta\to 1}\left(V_i^c(D)/V_i^d(D)\right) > 1$ and $\lim_{\beta\to 1}\left(V_i^c(N)/V_i^d(N)\right) > 1$, thus there exists $\tilde{\beta}_i \in [0,1)$ such that for all $\beta \geq \tilde{\beta}_i$, player i does not wish to deviate from the SPE. Let $\tilde{\beta} = \max_{i\in\mathcal{E}'}\tilde{\beta}_i$ and since $\mathcal{E}' \subset \mathcal{E}$ is finite, $\tilde{\beta} \in [0,1)$, and we have that for any distribution of contributions $\left\{\theta^{i*}(N), \theta^{i*}(D)\right\}_{i\in\mathcal{E}}$ and any punishment MPE with $\theta^{i*}(N) > \theta^{ip}(N)$ and $\theta^{i*}(D) > \theta^{ip}(D)$ when $s = D$ for all $i \in \mathcal{E}'$, there exist some $\tilde{\beta} \in [0,1)$ such that for all $\beta \geq \tilde{\beta} \in [0,1)$, (21.39) is satisfied for all $i \in \mathcal{E}$. This establishes that there exists $\bar{\beta} < 1$ such that for $\beta \geq \bar{\beta}$, the symmetric Pareto optimal SPE involves $p(D) = p(N) \in (0,1)$, proving the first part of the proposition.

To prove the second part, first note that as $\beta \to 1$, any combination of $\left\{\theta^{i*}(N), \theta^{i*}(D)\right\}_{i\in\mathcal{E}}$ will satisfy (21.39) and $p(D) = p(N) \in (0,1)$, thus we have SPEs that are Pareto optimal with $p(D) = p(N) \in (0,1)$. To complete the proof, we only have to show that there cannot exist any Pareto optimal SPEs that do not have this feature. To obtain a contradiction, suppose that there exists another SPE with contribution levels $\theta(N)$ and $\theta(D)$ different from $\theta^*(N)$ and $\theta^*(D)$. However, as $\beta \to 1$, $\theta^*(N)$ and $\theta^*(D)$ are feasible as shown above, and an SPE with $\{\theta^{i*}(N)\}_{i\in\mathcal{E}}$ and $\{\theta^{i*}(D)\}_{i\in\mathcal{E}}$ that Pareto dominates $\theta(N)$ and $\theta(D)$ can be constructed, showing that no other Pareto optimal SPE can exist as $\beta \to 1$. $\square$

This proposition therefore shows that as long as the discount factor is large enough, the "best" SPEs also give the same invariance result as the MPEs. Intuitively, with a high enough discount factor, the elite act totally cohesively, as a single agent, and the same calculus as in the MPE applies for equating the marginal cost of greater contributions to de facto power to the marginal benefits, again implying $p(D) = p(N)$. When the discount factor is sufficiently small, however, this result may no longer be true, because a different pattern of contributions may be necessary to ensure "incentive compatibility" on the side of the elite agents (i.e., to ensure that certain elite agents are willing to make the contributions they are supposed to make along the equilibrium path).

21.2.4. Comparative Statics. We now return to the symmetric MPE and derive a number of comparative static results. Comparative statics are straightforward in this case, since equations (21.16), (21.19) and (21.23), immediately imply that

$$V(N) - V(D) = \frac{\eta}{\phi M} > 0, \tag{21.42}$$

where we have dropped the conditioning of the value functions on the equilibrium $\theta(D)$ and $\theta(N)$ to simplify the notation. Equation (21.42) is intuitive. In the equilibrium of Proposition 21.1, the only difference between democracy and nondemocracy for the elite is that in democracy they have to spend more in contributions in order to retain the same de facto political power. In particular, the per elite additional spending is equal to $\eta/\phi M$, which is increasing in the de jure political power advantage that democracy creates for the citizens (since, in equilibrium, the elite totally offset this advantage).

Using (21.21) and (21.42) and denoting the equilibrium level of $\theta(N)$ by $\theta^*(N)$, we have:

$$\phi f(\phi M\theta^*(N)) \left(\frac{\Delta RL}{M} + \frac{\beta\eta}{\phi M}\right) = 1. \tag{21.43}$$

Similarly, denoting the equilibrium level of $\theta(D)$ by $\theta^*(D)$, we also have

$$\phi f(\phi M\theta^*(D) - \eta) \left(\frac{\Delta RL}{M} + \frac{\beta\eta}{\phi M}\right) = 1. \tag{21.44}$$

Finally, let us denote the probability that the elite will have political power by $p^* = p(D) = p(N)$, and recall that this probability corresponds both to the probability that the elite will control political power, and also the probability that the society will be nondemocratic and economic institutions will be labor repressive rather than competitive. Thus this probability summarizes most of the economic implications of the model.

PROPOSITION 21.4. ***(Comparative Statics)*** *Suppose that Assumptions 1-3 hold. Then in the baseline model, we have the following comparative static results:*

(1) *An increase in the economic rents that the elite can obtain by controlling political power will increase their contributions and the probability that they control political power, i.e.,*

$$\frac{\partial \theta^*(N)}{\partial \Delta R} > 0, \ \frac{\partial \theta^*(D)}{\partial \Delta R} > 0 \text{ and } \frac{\partial p^*}{\partial \Delta R} > 0.$$

(2) *An increase in the discount factor will increase the elite's contributions and the probability that they control political power, i.e.,*

$$\frac{\partial \theta^*(N)}{\partial \beta} > 0, \ \frac{\partial \theta^*(D)}{\partial \beta} > 0 \text{ and } \frac{\partial p^*}{\partial \beta} > 0.$$

(3) *An increase in the number of the elite will intensify the collective action problem among them, and will reduce their contributions and the probability that they control political power i.e.,*

$$\frac{\partial \theta^*(N)}{\partial M} < 0, \ \frac{\partial \theta^*(D)}{\partial M} < 0, \text{ and } \frac{\partial p^*}{\partial M} < 0.$$

(4) *An increase in the advantage of the citizens in democracy will increase the elite's contributions and the probability that they control political power, i.e.,*

$$\frac{\partial \theta^*(N)}{\partial \eta} > 0, \ \frac{\partial \theta^*(D)}{\partial \eta} > 0, \text{ and } \frac{\partial p^*}{\partial \eta} > 0.$$

(5) *An increase in the effectiveness of the de facto political power technology of the elite has ambiguous effects on their contributions, but increases the probability that they control political power, i.e.,*

$$\frac{\partial p^*}{\partial \phi} > 0.$$

PROOF. All of these comparative static results follow from (21.43) and (21.44) using the Implicit Function Theorem (e.g., Simon and Blume, 1994, Theorem 15.2). We can use the Implicit Function Theorem, since f is differentiable everywhere and moreover, Assumptions 2 and 3 ensure that the equilibrium is always at an interior point. We briefly sketch the argument for some of these results. For example, for $\partial \theta^*(N)/\partial \Delta R$, use the Implicit Function Theorem on (21.43) to obtain

$$\frac{\partial \theta^*(N)}{\partial \Delta R} = -\frac{f(\phi M \theta^*(N)) L}{f'(\phi M \theta^*(N)) M (\phi \Delta R L + \beta \eta)} > 0$$

since $f' < 0$ from the second order condition. Using the Implicit Function Theorem on (21.44) establishes $\partial\theta^*(D)/\partial\Delta R > 0$. To obtain $\partial p^*/\partial\Delta R > 0$, note that $p^* = F(\phi M\theta^*(N))$ and F is everywhere strictly increasing.

The comparative statics in part 2 with respect to β are identical.

Using the Implicit Function Theorem with respect to M also immediately establishes $\partial\theta^*(N)/\partial M < 0$ and $\partial\theta^*(D)/\partial M < 0$ as claimed in part 3. Since $p^* = F(\phi M\theta^*(N))$, the effect on p^* at first appears ambiguous. However, note from (21.43) that as M increases, the second term on the left-hand side declines, so $f(\phi M\theta^*(N))$ has to increase. Since $f' < 0$, this is only possible if $\phi M\theta^*(N)$ declines, so $p^* = F(\phi M\theta^*(N))$ also declines (given the monotonicity of F).

Next, the Implicit Function Theorem also gives the results in part 4, in particular,

$$\frac{\partial\theta^*(N)}{\partial\eta} = -\frac{\beta f(\phi M\theta^*(N))}{\phi f'(\phi M\theta^*(N))\, M\, (\phi\Delta RL + \beta\eta)} > 0,$$

and similarly, $\partial\theta^*(D)/\partial\eta > 0$. The argument for $\partial p^*/\partial\eta > 0$ is again similar. The second term on the left-hand side of (21.43) increases as η increases, so $f(\phi M\theta^*(N))$ has to decline. Since $f' < 0$, this implies that $\phi M\theta^*(N)$ increases, so $p^* = F(\phi M\theta^*(N))$ also increases, establishing $\partial p^*/\partial\eta > 0$.

Finally, it is straightforward to verify that the effect of ϕ on $\theta^*(N)$ and $\theta^*(D)$ is ambiguous. However, writing (21.43) as

$$f(\phi M\theta^*(N))\left(\frac{\phi\Delta RL}{M} + \frac{\beta\eta}{M}\right) = 1,$$

we see that an increase in ϕ increases the second term on the left-hand side, so $f(\phi M\theta^*(N))$ has to decline. Since $f' < 0$, this implies that $\phi M\theta^*(N)$ increases, and $p^* = F(\phi M\theta^*(N))$ must also increase, establishing $\partial p^*/\partial\phi > 0$. $\square$

Many of the comparative statics in Proposition 21.4 are intuitive, and yet quite useful in terms of economic implications. For example, the fact that an increase in ΔR induces both greater contributions by elite agents and also increases the probability that they control political power is very intuitive, since ΔR is a measure of how much they have to gain by controlling political power. The greater is this gain, the more willing is each elite agent to contribute to their collective political power. Since ΔR will be high when A is low, Proposition 21.4 also implies $\partial p^*/\partial A < 0$. In other words, in a society where alternative (e.g., proto-industrial) economic activities are less developed and where, by repressing labor the elite can obtain large rents, political and economic institutions are more likely to be controlled by the elite, and consequently economic institutions are more likely to be tilted towards repressive labor relations rather than competitive labor markets.

The fact that a higher β also increases contributions by the elite and the likelihood of labor repressive institutions is also interesting. In many models, a higher discount factor leads to better allocations. Here, in contrast, a higher discount factor leads to more wasteful activities by the elite and labor repressive economic institutions. The reason is that the main pivotal agents in this model are the elite, which, by virtue of their smaller numbers, take the effect of their contributions on equilibrium allocations into account. Contributing to de facto political power is a form of investment, and some of the returns accrue to the elite in the future (when they secure nondemocracy instead of democracy). Therefore a higher level of β encourages them to invest more in their political power and makes nondemocracy and labor repressive economic institutions more likely.

The third set of comparative statics show that when M increases so that there are more elite agents, the collective action problem among them becomes worse. This highlights the latent free-rider problem in the model. Even though each elite agent contributes to the group's political power, their level of contribution is still suboptimal from the viewpoint of the group, since each elite agent realizes that by contributing more he creates a positive externality on other elites. A greater M increases the extent of this positive externality and intensifies the free-rider problem (the collective action problem). This comparative static therefore suggests that nondemocracy and labor repressive economic institutions are more likely to emerge when there is a relatively small and cohesive group of elite land owners, a pattern consistent with the historical case studies discussed below.

The most surprising comparative static results are those with respect to η. Recall that a higher η corresponds to democracy giving more de jure power to the citizens. We may have therefore expected a greater η to lead to better outcomes for the citizens. In contrast, we find that higher η makes nondemocracy and labor repressive economic institutions more likely (as long as Assumption 3 still holds). The reason for this is that a higher η makes democracy more costly for the elite, so it is in the interest of each elite agent to invest more in the group's political power to avoid democracy. This effect is strong enough to increase the probability that they will maintain political power. However, the overall impact of η on the likelihood of democracy is non-monotonic: if η increases so much that Assumption 3 no longer holds, then Corollary 21.1 applies and democracy may become fully consolidated (i.e., an absorbing state).

Note however that both the results that higher M reduces the political power of the elite and that higher η increases their political power depend on the functional form assumptions already highlighted in Remark 1. Nevertheless, we believe that the baseline model we have is both the simplest and the most natural one, and highlights important first-order effects.

Finally, when ϕ increases, the technology of garnering de facto political power for the elite improves. This may reduce their contributions to the group's de facto power, but it will always increase the equilibrium probability of a nondemocratic regime and labor repressive institutions.

21.3. Generalizations: Markov Regime-Switching Models and State Dependence

The model presented so far yielded stark results, which were partly driven by the assumptions that the elite had the same technology to generate de facto political power in both regimes and were able to change economic institutions immediately after they took control. Relaxing these assumptions leads to a richer form of persistence, in the form of a Markov regime-switching model with state dependence.

Another special feature of the model so far was that it implicitly assumed that changing economic institutions and changing the political system were equally easy (or difficult). An extension in which changing political institutions may require greater political power than influencing economic policies or institutions is also discussed below.

To simplify the discussion, let us from now on focus on symmetric MPE.

21.3.1. Limits on De Facto Power of the Elite. Our first generalization of the above framework assumes that in democracy, because of limits on the activities of the elite, their technology for gathering de facto political power changes to

$$P_t^E(D) = \phi_D Z_t, \tag{21.45}$$

where $\phi_D \in (0, \phi)$ and $Z_t = \sum_{i\in\mathcal{E}} \theta_t^i$. In other words, each unit of the final good spent by the elite for increasing their de facto power is less effective in democracy than in nondemocracy. This is a reasonable assumption, since democratic institutions may prevent the elite from using repression or paramilitaries or from buying politicians as effectively as in a nondemocratic regime. Therefore, in this model democracy has two simultaneous functions; it shifts the distribution of de jure political powers towards the citizens and it limits the exercise of de facto power by the elite.

We now have the probability of the elite controlling the political agenda in democracy as

$$p\left(\theta^{i},\theta\left(D\right)\mid D\right)=F\left(\phi_{D}\left(\left(M-1\right)\theta\left(D\right)+\theta^{i}\right)-\eta\right), \tag{21.46}$$

and the value function in democracy is unchanged and is still given by (21.19). It is clear that Definition 1 still applies to this modified model, and specifies symmetric MPEs.

Assuming interior solutions, then the symmetric MPE is characterized by (21.21) and

$$\phi_{D}f\left(\phi_{D}M\theta\left(D\right)-\eta\right)\left(\frac{\Delta RL}{M}+\beta V\left(N\mid\theta\left(N\right),\theta\left(D\right)\right)-\beta V\left(D\mid\theta\left(N\right),\theta\left(D\right)\right)\right)=1. \tag{21.47}$$

We can impose a variant of Assumption 3 to ensure that the equilibrium is interior:

Assumption 3':

$$\min\left\{\phi f\left(0\right)\frac{\Delta RL}{M},\phi_{D}f\left(-\eta\right)\frac{\Delta RL}{M}\right\}>1.$$

Now recalling that $p\left(N\right)\equiv p\left(\theta\left(N\right),\theta\left(N\right)\mid N\right)$ and $p\left(D\right)\equiv p\left(\theta\left(N\right),\theta\left(N\right)\mid D\right)$, comparison of (21.21) and (21.47) immediately implies that

$$p\left(N\right)>p\left(D\right). \tag{21.48}$$

To see this more explicitly, note that $\phi>\phi_{D}$ implies $f\left(\phi_{D}M\theta\left(D\right)-\eta\right)>f\left(\phi M\theta\left(N\right)\right)$. Since in the neighborhood of equilibrium, $f\left(\cdot\right)$ is decreasing and $F\left(\cdot\right)$ is strictly increasing everywhere, we must have $\phi_{D}M\theta\left(D\right)-\eta<\phi M\theta\left(N\right)$ and $p\left(D\right)=F\left(\phi_{D}M\theta\left(D\right)-\eta\right)<p\left(N\right)=F\left(\phi M\theta\left(N\right)\right)$. Note that $p\left(N\right)$ is the probability of nondemocracy persisting, while $1-p\left(D\right)$ is the probability of democracy persisting. This implies that labor repressive institutions are less likely to arise in democracy than in nondemocracy. Moreover, once the society is democratic, it has a higher probability of remaining democratic than the probability of switching to democracy from nondemocracy. Consequently, in this model there is persistence of both political and economic institutions.

Assumption 3' also implies $p\left(D\right)>0$, so even in democracy, the elite have the potential power to impose their favorite economic institutions, and change the political system back to nondemocracy, and moreover from Assumptions 2 and 3', we have $p\left(N\right)\in\left(0,1\right)$, so the Markov process is ergodic (irreducible and aperiodic). Next, dividing (21.21) by (21.47) yields:

$$\phi_{D}f\left(\phi_{D}M\theta\left(D\right)-\eta\right)=\phi f\left(\phi M\theta\left(N\right)\right), \tag{21.49}$$

which shows that the gap between ϕ and ϕ_{D} will determine the gap between $p\left(D\right)$ and $p\left(N\right)$, thus the extent of persistence of economic and political institutions (e.g., as $\phi_{D}\rightarrow\phi$, $p\left(D\right)\rightarrow p\left(N\right)$). This leads to the following result:

Proposition 21.5. ***(Limits on De Facto Power and State Dependence)*** *Consider the modified model with limits on the elite's de facto power in democracy. Suppose that Assumptions 1, 2 and 3' hold. Then any symmetric MPE leads to a Markov regime switching structure where the society fluctuates between democracy with associated competitive economic institutions* ($\tau = 1$) *and nondemocracy with associated labor repressive economic institutions* ($\tau = 0$)*, with switching probabilities* $p(N) \in (0,1)$ *and* $1 - p(D) \in (0,1)$ *where* $p(D) < p(N)$.

The proof of this proposition is omitted since it is similar to that of Proposition 21.1, and follows straightforwardly from the expressions in the text, in particular, equations, (21.21) and (21.47), and Assumptions of 1, 2 and 3'.

The most important implication of this modified model is that there is now a different type of institutional persistence—which we refer to as state dependence, since the probability distribution over equilibrium political and economic institutions depends on the current state of the system, $s \in \{N, D\}$. While Proposition 21.1 featured invariance in the sense that economic institutions followed the same equilibrium process irrespective of political institutions, it did not lead to persistence in political institutions; the fact that $p(D) = p(N)$ implied that democracy was as likely to follow a democratic regime as it was to follow a non-democratic regime. The results in Proposition 21.5 are different; once in nondemocracy, the society is more likely to remain nondemocratic than it is to switch to nondemocracy from democracy. This is the essence of state dependence.

Also interesting is the fact that the elite still have the ability to solve their collective action problem and gather sufficient de facto power to dominate democratic politics and impose their favorite economic institutions, i.e., $p(D) > 0$ (though here this also corresponds to their ability to also change the political system from democracy to nondemocracy).

21.3.2. Sluggish Economic Institutions. Next we modify the above framework in a different direction, and assume that starting in democracy, the elite cannot impose their favorite economic institutions immediately, for example, democratic politics has already taken some actions that cannot be reversed within the same period. This implies that starting in democracy, economic institutions are "slow-changing" or sluggish. This structure is also formally equivalent to one in which the elite incur a temporary cost when they change economic institutions from competitive to labor repressive.

More specifically, we now allow three different types of economic institutions: $\tau_t = 1$, corresponding to competitive markets, $\tau_t = 0$, corresponding to full labor repression and $\tau_t = 1/2$ corresponding to partial labor repression, in which case, wages are reduced to some level $A \leq \bar{w} < w(\tau_t = 1) \equiv g(L) - Lg'(L)$, and thus returns to land owners with partial labor repression are equal to

$$R^p \equiv R(\tau = 1/2) \equiv \frac{g(L) - \bar{w}}{L}. \tag{21.50}$$

Let us define λ such that

$$\lambda \equiv \frac{R^p - R^c}{\Delta R},$$

with ΔR as defined in (21.12). The fact that $A \leq \bar{w} < w(\tau = 1)$ ensures $\lambda \in (0, 1]$.

The only difference from the baseline model is that starting in $s = D$, even if the elite gain political power, they cannot impose $\tau = 0$, and the best they can do is to set $\tau = 1/2$ (while starting in $s = N$, any $\tau \in \{0, 1/2, 1\}$ is allowed). Given this assumption, the rest of the analysis is similar to before, with the only difference taking place in the value function in democracy, which now takes the form:

$$\begin{aligned} V(D \mid \theta(N), \theta(D)) \quad = \quad & \max_{\theta^i \geq 0} \left\{ -\theta^i + p\left(\theta^i, \theta(D) \mid D\right) \left(\frac{R^p L}{M} + \beta V(N \mid \theta(N), \theta(D)) \right) \right. \\ & \left. + \left(1 - p\left(\theta^i, \theta(D) \mid D\right)\right) \left(\frac{R^c L}{M} + \beta V(D \mid \theta(N), \theta(D)) \right) \right\}. \end{aligned} \tag{21.51}$$

Once again, focusing on interior solutions, this maximization problem implies the first-order condition

$$\phi f\left(\phi\left((M-1)\theta(N) + \theta^i\right) - \eta\right) \left(\frac{\lambda \Delta R L}{M} + \beta\left(V(N \mid \theta(N), \theta(D)) - V(D \mid \theta(N), \theta(D))\right) \right) = 1, \tag{21.52}$$

which only differs from (21.20) because the gain of capturing power is now $\lambda \Delta R$ rather than ΔR. The corresponding second-order condition is $f'\left(\phi\left((M-1)\theta(N) + \theta^i\right) - \eta\right) < 0$. Once again, let the policy correspondence be denoted by $\Gamma^D(\theta(N), \theta(D))$.

The value function in nondemocracy is unchanged and is given by (21.16), and the first-order condition for contributions is given by (21.17), with the policy correspondence given by $\Gamma^N(\theta(N), \theta(D))$.

To define an equilibrium formally, let us also recall that $\pi = 1$ stands for the citizens having political power. Now let $\pi = 0$ stand for the elite having full power, so that they can set $\tau = 0$ if they want to, and let $\pi = 1/2$ denote the elite capturing political power starting in democracy. Thus we have:

Definition 21.4. *A symmetric MPE of the model with sluggish economic institutions consists of a pair of contribution levels for elite agents $\theta(N)$ and $\theta(D)$, such that $\theta(N) \in \Gamma^N[\theta(N), \theta(D)]$ and $\theta(D) \in \Gamma^D[\theta(N), \theta(D)]$. In addition, economic and political decisions $\tau(\pi)$ and $s'(\pi)$ are such that $\tau(\pi = 0) = 0$, $s'(\pi = 0) = N$, $\tau(\pi = 1/2) = 1/2$, $s'(\pi = 1/2) =$*

N, $\tau(\pi=1)=1$ and $s'(\pi=1)=D$, and factor prices are given by (21.8)-(21.11) and $\bar{w}$ and (21.50) when $\tau=1/2$.

Given this definition of a symmetric MPE, the equilibrium condition for $\theta(N)$ is again the same as before, i.e., equation (21.22), while with the same steps as before, the equilibrium condition for $\theta(D)$ is given by:

$$\phi f(\phi M\theta(D)-\eta)\left(\frac{\lambda\Delta RL}{M}+\beta(V(N\mid\theta(N),\theta(D))-V(D\mid\theta(N),\theta(D)))\right)=1. \tag{21.53}$$

Comparison of this condition to (21.22) immediately establishes that as long as $\lambda<1$, i.e., as long as democracy *does put* restrictions on economic institutions that the elite can impose, we have $p(D)<p(N)$.

As before, we impose an assumption to ensure an interior equilibrium:

Assumption 3":

$$\min\left\{\phi f(0)\frac{\Delta RL}{M},\phi f(-\eta)\frac{\lambda\Delta RL}{M}\right\}>1.$$

Proposition 21.6. ***(Sluggish Economic Institutions and State Dependence)*** *Consider the modified model with sluggish economic institutions. Suppose that Assumptions 1, 2 and 3" hold. Then any symmetric MPE leads to a Markov regime switching structure where the society fluctuates between democracy and nondemocracy, with switching probabilities $p(N)\in(0,1)$ and $1-p(D)\in(0,1)$ where $p(D)<p(N)$.*

Proof. This result follows immediately from the comparison of (21.22) with (21.53), which establishes $p(D)<p(N)$. □

21.4. Durable Political Institutions and Captured Democracy

The assumption so far has been that when the elite have more political power than the citizens, they can change both economic institutions and the political system. The historical examples below illustrate a different salient pattern: democracy may emerge and endure, but in a *captured* form, whereby the elite are able to impose their favorite economic institutions (or at the very least, have a disproportionate effect on the choice of economic institutions) in an enduring democracy. It is therefore important to generalize the model so that there can be differences between equilibrium political and economic institutions.

As discussed in previous lectures, in many situations, political institutions are more difficult to change, and may have additional "durability". We now modify the baseline model to incorporate this feature and assume that overthrowing a democratic regime is more difficult than influencing economic institutions. More specifically, the elite require greater political power to force a switch from democracy to nondemocracy than simply influencing economic institutions in democracy. To simplify the discussion, we assume that when they influence economic institutions in democracy, they can choose their favorite economic institutions, labor repression.

Finally, for reasons that will become apparent below, we now assume that the time t preferences of citizens, i.e., those for all $i \in \mathcal{C}$, are given by

$$\sum_{j=0}^{\infty} \beta^j \left(c_{t+j}^i + \nu \left(S_{t+j} \right) \right), \tag{21.54}$$

with $\nu (S = N) = 0$ and $\nu (S = D) > 0$. Therefore, these preferences allow a direct utility for the citizens from democracy (which may in turn be because of ideological reasons or a reduced-form for other benefits provided to the citizens by democracy). Moreover, we will assume that $\nu (S = D)$ is large enough that citizens always prefer and vote for democracy even when this may have economic costs for them.

We model institutional change as follows. We assume that when $s = D$ and $P_t^C + \xi > P_t^E \geq P_t^C$, where $\xi > 0$, the elite can choose economic institutions but cannot change the political system. If, on the other hand, $P_t^E \geq P_t^C + \xi$, the elite can choose both economic institutions and the future political system. Symmetrically when $s = N$ and $P_t^E + \xi > P_t^C \geq P_t^E$, the citizens can choose economic institutions, but cannot change the political system. This formulation builds in the assumption that changing political institutions is more difficult than influencing economic institutions in the most straightforward way.

We again focus on symmetric MPE. Also, to keep the issues separate, here we assume that when the elite have more political power in democracy, they can impose their most preferred economic institutions, $\tau = 0$, as in the baseline model. Finally, to further simplify the discussion we strengthen Assumption 2:

> **Assumption 2':** F is defined over $(\underline{\omega}, \infty)$ for some $\underline{\omega} < 0$, is everywhere strictly increasing and twice continuously differentiable (so that its density f and the derivative of the density, f', exist everywhere), and moreover we have $f'(\omega) < 0$ for all ω and $\lim_{\omega \to \infty} f(\omega) = 0$.

Given these assumptions, the structure of the model is similar to before. The value functions are more complicated, but have similar intuition to those above. In particular, in addition to (21.18), let

$$\hat{p}\left(\theta^i, \theta(D) \mid D\right) = F\left(\phi\left((M-1)\,\theta(D) + \theta^i\right) - \eta - \xi\right), \tag{21.55}$$

so that we have:

$$\begin{aligned} V\left(D \mid \theta\left(N\right),\theta\left(D\right)\right) &= \max_{\theta^i \geq 0}\Bigg\{-\theta^i + p\left(\theta^i,\theta\left(D\right) \mid D\right)\frac{R^r L}{M} + \\ & \left(1 - p\left(\theta^i,\theta\left(D\right) \mid D\right)\right)\frac{R^c L}{M} + \hat{p}\left(\theta^i,\theta\left(D\right) \mid D\right)\beta V\left(N \mid \theta\left(N\right),\theta\left(D\right)\right) \\ & + \left(1 - \hat{p}\left(\theta^i,\theta\left(D\right) \mid D\right)\right)\beta V\left(D \mid \theta\left(N\right),\theta\left(D\right)\right)\Bigg\}, \end{aligned} \tag{21.56}$$

where we have already imposed that when the citizens have sufficient power they will choose democracy.

With similar arguments to before, the maximization in (21.56) implies the following first-order condition

$$\begin{aligned} & \phi f\left(\phi\left(M-1\right)\theta\left(D\right) + \theta^i - \eta\right)\frac{\Delta RL}{M} \\ & + \beta\phi f\left(\phi\left(M-1\right)\theta\left(D\right) + \theta^i - \eta - \xi\right)\left(V\left(N \mid \theta\left(N\right),\theta\left(D\right)\right) - V\left(D \mid \theta\left(N\right),\theta\left(D\right)\right)\right) = 1, \end{aligned} \tag{21.57}$$

which is now sufficient since Assumption 2' ensures that the second-order condition is satisfied.

The main difference of this first-order condition from the one before is that the probability with which the elite gain the economic rent $\Delta RL/M$ is different from the probability with which they secure a change in the political system. For this reason, two different densities appear in (21.57). As before, denote the resulting policy correspondence as $\Gamma^D\left(\theta\left(N\right),\theta\left(D\right)\right)$.

Similarly for nondemocracy, we define

$$\hat{p}\left(\theta^i,\theta\left(N\right) \mid N\right) = F\left(\phi\left(\left(M-1\right)\theta\left(N\right) + \theta^i\right) + \xi\right), \tag{21.58}$$

which leads to a modification of the value function for nondemocracy as

$$\begin{aligned} V\left(N \mid \theta\left(N\right),\theta\left(D\right)\right) &= \max_{\theta^i \geq 0}\Bigg\{-\theta^i + p\left(\theta^i,\theta\left(N\right) \mid N\right)\frac{R^r L}{M} + \\ & \left(1 - p\left(\theta^i,\theta\left(N\right) \mid N\right)\right)\frac{R^c L}{M} + \hat{p}\left(\theta^i,\theta\left(N\right) \mid N\right)\beta V\left(N \mid \theta\left(N\right),\theta\left(D\right)\right) \\ & + \left(1 - \hat{p}\left(\theta^i,\theta\left(N\right) \mid N\right)\right)\beta V\left(D \mid \theta\left(N\right),\theta\left(D\right)\right)\Bigg\}, \end{aligned} \tag{21.59}$$

which also has a similar structure to the value function in democracy in this case, (21.59). Consequently, the first-order (necessary and sufficient given Assumption 2') condition for optimal contribution by an elite agent is also similar:

$$\begin{aligned} & \phi f\left(\phi\left(M-1\right)\theta\left(N\right) + \theta^i\right)\frac{\Delta RL}{M} \\ & + \beta\phi f\left(\phi\left(M-1\right)\theta\left(N\right) + \theta^i + \xi\right)\left(V\left(N \mid \theta\left(N\right),\theta\left(D\right)\right) - V\left(D \mid \theta\left(N\right),\theta\left(D\right)\right)\right) = 1 \end{aligned} \tag{21.60}$$

which again defines the policy correspondence $\Gamma^N\left(\theta\left(N\right),\theta\left(D\right)\right)$.

To define an equilibrium, now introduce the additional notation such that $\pi = (0,0)$ denotes the elite keeping total power in nondemocracy or gaining total power in democracy (i.e., $P_t^E \geq P_t^C$ when $s = N$ or $P_t^E \geq P_t^C + \xi$ when $s = D$); $\pi = (0,1)$ corresponding to the elite keeping control of de jure power but losing control of economic institutions in nondemocracy (i.e., $P_t^E + \xi \geq P_t^C > P_t^E$); $\pi = (1,1)$ means the elite loses power in nondemocracy or fails to gain any power in democracy (i.e., $P_t^C > P_t^E + \xi$ when $s = N$ or $P_t^C > P_t^E$ when $s = D$); and finally, $\pi = (1,0)$ corresponds to the citizens maintaining de jure power in democracy but losing control over economic institutions (i.e., $P_t^C + \xi > P_t^E \geq P_t^C$). Imposing that citizens always prefer democracy to nondemocracy (from preferences in (21.54)), we have:

Definition 21.5. *A symmetric MPE of the model with durable political institutions consists of a pair of contribution levels for elite agents $\theta(N)$ and $\theta(D)$, such that $\theta(N) \in \Gamma^N[\theta(N), \theta(D)]$ and $\theta(D) \in \Gamma^D[\theta(N), \theta(D)]$. In addition, economic and political decisions $\tau(\pi)$ and $s'(\pi)$ are such that $\tau(\pi = (0,0)) = \tau(\pi = (1,0)) = 0$, $s'(\pi = (0,0)) = s'(\pi = (0,1)) = N$, $\tau(\pi = (1,1)) = \tau(\pi = (0,1)) = 1$, $s'(\pi = (1,1)) = s'(\pi = (1,0)) = D$, and factor prices are given by (21.8)-(21.11) as a function of $\tau \in \{0,1\}$.*

Given this definition, a symmetric MPE with $\theta(N) > 0$ and $\theta(D) > 0$ is a solution to the following two equations

(21.61)
$$\phi f(\phi M\theta(D) - \eta)\frac{\Delta RL}{M} + \beta\phi f(\phi M\theta(D) - \eta - \xi)(V(N \mid \theta(N), \theta(D)) - V(D \mid \theta(N), \theta(D))) = 1.$$

(21.62)
$$\phi f(\phi M\theta(N))\frac{\Delta RL}{M} + \beta\phi f(\phi M\theta(N) + \xi)(V(N \mid \theta(N), \theta(D)) - V(D \mid \theta(N), \theta(D))) = 1.$$

It can be verified that Assumption 3 above is sufficient to ensure that zero contributions cannot be equilibria.

The interesting result in this case is that once the society becomes democratic, it will remain so potentially for a long time (i.e., $\hat{p}(D)$ can be small), but the elite will still be able to control the economic institutions (i.e., $p(D)$ could be quite large). This is stated and proved in the next proposition.

Proposition 21.7. ***(Captured Democracy)*** *Consider the modified model with durable political institutions. Suppose that Assumptions 1, 2' and 3 hold. Then we have a Markov regime-switching process with state dependence and $1 > \hat{p}(N) > \hat{p}(D) > 0$. Moreover, democracy is captured in the sense that $0 < p(N) < p(D) < 1$, i.e., democracy will survive but choose economic institutions in line with the elite's interests with even a higher probability than does nondemocracy.*

PROOF. The probability of labor repressive economic institutions under democracy is

$$p(D) = p(\theta(D), \theta(D) \mid D) = F(\phi M \theta(D) - \eta),$$

while it is

$$p(N) = p(\theta(N), \theta(N) \mid N) = F(\phi M \theta(N))$$

in nondemocracy. Suppose, to obtain a contradiction, that $p(D) \leq p(N)$. This is equivalent to

$$\phi M \theta(D) - \eta \leq \phi M \theta(N). \tag{21.63}$$

Since from Assumption 2' f is decreasing everywhere, this implies

$$f(\phi M \theta(D) - \eta) \geq f(\phi M \theta(N)).$$

This equation combined with (21.61) and (21.62) implies that

$$f(\phi M \theta(D) - \eta - \xi) \leq f(\phi M \theta(N) + \xi).$$

Since from Assumption 2' f is decreasing, this is equivalent to

$$\phi M \theta(D) - \eta - \xi \geq \phi M \theta(N) + \xi,$$

which, given $\xi > 0$, contradicts (21.63), establishing that $p(D) > p(N)$, i.e., that democracy is captured.

But, by the same reasoning, $p(D) > p(N)$ implies $f(\phi M \theta(D) - \eta - \xi) > f(\phi M \theta(N) + \xi)$, thus $\phi M \theta(D) - \eta - \xi < \phi M \theta(N) + \xi$. Since F is strictly monotonic, this implies $\hat{p}(N) > \hat{p}(D)$, establishing the Markov regime-switching structure. □

The equilibrium in this proposition is the richest and perhaps the most interesting one we have encountered so far. The equilibrium still takes a Markov regime-switching structure with fluctuations between democracy and nondemocracy; but in democracy, there is no guarantee that economic institutions will be those favored by the citizens. While in all the previous models we studied, the elite were able to impose both their political and economic wishes at the same time, here we have an equilibrium pattern whereby democracy persists, but the elite are able to impose their favorite economic institutions. In fact, the proposition shows that (given Assumption 2') the elite will be able to impose labor repressive economic institutions with a *higher* probability under democracy than in nondemocracy.

The intuition for this (somewhat paradoxical) result is that in democracy there is an additional benefit for the elite to invest in de facto political power, which is to induce a switch from democracy to nondemocracy. Consequently, the elite invest in their de facto power sufficiently more in democracy that they are able to obtain their favorite economic institutions with a greater probability. Nevertheless, the elite are happier in nondemocracy, because the cost of investing in their de facto political power in democracy is significantly higher. In fact, it is precisely because they prefer nondemocracy to democracy that they are willing to invest more in their de facto political power in democracy and obtain the labor repressive economic institutions with a high probability. What about citizens? If there were no additional benefit of democracy, $\nu(S = D) > 0$, then citizens would actually be worse off in democracy than in nondemocracy, because they only care about economic institutions and economic institutions are more likely to be labor repressive in democracy than in nondemocracy. Thus when $\nu(S = D) = 0$, citizens would never choose democracy, and would be happy to remain in nondemocracy (given the limited ability that they have to solve the collective action problem). Therefore, the ideological or other benefits of democracy encapsulated in $\nu(S = D) > 0$ create the possibility of the captured democracy equilibrium, whereby the citizens are willing to vote and defend democracy, but democracy at the end caters to the wishes of the elite.

21.5. Effective Reform

Let us now briefly discuss how institutional persistence can be broken by effective reforms. Our framework shows how the equilibrium path may feature invariance (i.e., labor repressive and generally dysfunctional economic institutions remaining in place despite shocks that change the political organization of society) or state dependence (where dictatorship is more likely to follow dictatorship than it is to follow democracy). Are there any major reforms that could break these various types of persistence?

The comparative static results above suggest potential answers to this question. In particular, the results so far show that a change in political institutions from nondemocracy to democracy is likely to be effective (in terms of leading to equilibrium competitive labor markets and persistent democracy) under two alternative (but complementary) scenarios. First, if democracy creates a substantial advantage for the citizens in the form of a large value of η, then as shown by Corollary 21.1 this will end the cycle of institutional persistence and make the permanent consolidation of democracy an equilibrium.

Second, one of the following reforms is undertaken *simultaneously* with the switch to democracy, then the economy is less likely to switch back to nondemocracy and labor repressive economic institutions: (1) a reduction in ϕ_D in terms of the model allowing for limits on de facto power of the elite in democracy, so that the traditional elites are less able to control politics in a democratic society (for example, preventing local threats of violence or the capture of political parties by the traditional elites would achieve such an outcome). (2) a reduction in ΔR, for example, by means of an increase in A, which will reduce the potential rents that the landed elites can obtain and will discourage them from investing in de facto political power.

This discussion therefore illustrates that while politics as business-as-usual may favor the elite even in democracy, undertaking simultaneous and significant reforms may change the character of the political equilibrium, making democracy and competitive labor markets more likely. An attractive example of simultaneous reform leading to a significant change in the distribution of political power in society is the 1688 Glorious Revolution in England, which not only changed de jure power by dethroning the Stuart monarchy, but also by substantially increasing the role of the Parliament and the allocation of economic resources in society, irreversibly altered the distribution of de facto political power (see, for example, North and Weingast, 1989, Acemoglu, Johnson and Robinson, 2005a). Another interesting example of simultaneous reforms arises in our discussion of how the central aspects of the economic institutions of the U.S. South, having survived the Civil War, finally changed in the 1960s.

21.6. References

(1) Acemoglu, Daron and James Robinson (2006) *Economic Origins of Dictatorship and Democracy*, Cambridge University Press.

(2) Acemoglu, Daron and Robinson, James A. (2006) "Persistence of Power, Elites and Institutions." Unpublished Paper.

(3) Banerjee, Abhijit and Andrew F. Newman (1993) "Occupational Choice and the Process of Development," *Journal of Political Economy*, 101, 274-298.

(4) Brenner, Robert (1976) "Agrarian Class Structure and Economic Development in Pre-Industrial Europe," *Past and Present,* 70, 30-75.

(5) Coate, Stephen and Morris, Stephen E. (1999) "Policy Persistence." *American Economic Review*, 1999, 89, 1327-1336.

(6) David, Paul (1985) "Clio and the Economics of QWERTY" *American Economic Review*, 75, 332-337.

(7) Engerman, Stanley L. (2006) Slavery, Emancipation and Freedom: Comparative Perspectives, Unpublished Book Manuscript.

(8) Engerman, Stanley L. and Sokoloff, Kenneth L. (1997) "Factor Endowments, Institutions, and Differential Growth Paths among New World Economies" in Stephen

Haber ed. *How Latin America Fell Behind*. Stanford: Stanford University Press, 1997.

(9) Liebowitz, Stanley J. and Stephen E. Margolis (2002) The economics of QWERTY : history, theory, and policy, New York; New York University Press.

(10) Thelen, Kathleen (2004) *How institutions evolve: the political economy of skills in Germany, Britain, the United States, and Japan*, New York; Cambridge University Press.

CHAPTER 22

Modeling Constitutions

The analysis so far emphasized the role of de jure political institutions in regulating the future allocation of political power. For example, a constitution may specify how an upper house and the lower house will interact in determining fiscal policy, and this may reduce the power of groups controlling the lower house, as has typically been the case in many bicameral systems especially during the 19th century. However, this modeling approach, and more generally this perspective, presumes that"institutions" or "constitutions" matter in allocating political power. But why do they do so? For example, why don't the players in the game (citizens, elites etc.) ignore the constitution altogether and use simply their "de facto political power" in reaching decisions?

This is a deep and difficult question. One perspective would be that constitutions, or laws in general, are no more than cheap talk in games, so they can not have any effect other than as a selection mechanism among many equilibria (see, for example, Mailath, Morris and Postelwaite, 2001). Essentially, anything that gets done because it is written in a constitution (on a piece of paper) could have been done if people coordinated on it. Although this perspective is theoretically compelling in many ways, in practice we see constitutions and more generally institutions playing a binding role on individuals (though clearly this does depend on expectations). Therefore it is important to understand how they will play such a role and how they will be chosen when individuals recognize such a role of constitutions, laws and institutions. This takes us to the question of how constitutions should be modeled?

You will recognize the parallel between this question and the set of issues that arose in the Acemoglu-Robinson (2006) framework. So what we are doing now is to push this interpretation somewhat further and think of the role of constitutions in the context of procedures of democratic decision-making.

A useful starting point is to think of "self stable" constitutions, which specify a voting protocol that will never lead to its own repeal. This is a relatively weak condition, and we may think that constitutional design should at least satisfy this self stability condition. To see why this is a weak condition, consider the following "stronger condition" that a voting rule should always choose itself over the set of alternative voting rules. Unfortunately, this stronger notion of stability turns out to be not very useful. Koray (2000) shows that, similar to Arrow's impossibility theorem, only voting rules that are dictatorial will satisfy this stronger notion of stability.

However, we will now see that the weaker notion of self-stability provides a more workable model and leads to a range of insights. Moreover, real examples of constitutions suggest that they may often be self-stable at least in the sense that it is often the exercise of de facto power that leads to new constitutions, rather than new constitutions emerging endogenously within the existing rules. A nice example is the discussion by Weingast (1998) of the Missouri Compromise in the US. At the time of writing the US Constitution, the issue of slavery was highly contentious. One issue was representation. Since the number of congressmen a state got depended on population one question was whether or not slaves counted as 'people' when representation was to be determined. The US Constitution struck a compromise (Northerners said no, slaves should not count, Southerners, somewhat ironically, said yes they should), a slave would count as 3/5 of a non-slave! These issues surfaced again with the debate on the admission of Missouri to the Union. Each state had two members in the Senate and after the admission of Alabama in 1819 the North and South were exactly balanced. South was worried however that anti-slave interests would gain a majority in the long run. The issue was would Missouri be a slave state or non-slave state. In the end the compromise was that Missouri would be a slave state (Congress passed a bill forbidding slavery in Missouri but it was vetoed in the Senate) but Maine, a non-slave state was simultaneously admitted into the Union, keeping the balance in the Senate. From then on until the Civil War pairs of states were admitted, one pro-slave one anti-slave.

Other interesting examples come from Latin America and Africa. During the 19th century Latin American countries with a few exceptions (Brazil inherited a monarchy from Portugal which lasted until 1888 and was more stable politically) continually re-wrote their constitutions (see Loveman, 1993, for an overview). Bolivia had 11 constitutions in the 19th century, Colombia had 8, and the Dominican Republic had 15 between 1844 and 1896! Some of these involved cosmetic changes but the notion propagated by some that Latin America countries slavishly adopted the US constitution is nonsense. Constitutions were fought over and some implemented radical changes. For instance Colombia introduced a hyper-federal system in 1863 where the national army was not even allowed to intervene in the local politics of states. In 1886 it moved back to a centralized system. Interesting, the stylized fact is that changes in constitutions were the result of civil wars in all these countries, not endogenous re-writing of the constitutions. It is also interesting that the US Constitution was only amended not completely re-written to benefit the winners after the Civil War. In the case of Africa, countries have constantly re-written constitutions since independence. The first wave of this took place after the creation of one-party states in the late 1960s and early 1970s, again suggesting that de facto power is required. However, the issue of stability is complicated. The 1886 Colombian constitution declared that it was unconstitutional to introduce a new constitution (though it could be amended). Nevertheless, in 1990 the supreme court ruled that it was constitutional to re-write the constitution!

We currently do not have models that capture the richness of the set of issues mentioned above. In addition there have been no really serious empirical studies of the evolution of constitutions over time (plenty of data in Latin America and Africa). But models in which self-stability of constitutions or certain laws are investigated constitute an important step towards a theoretical understanding of the issues that might be central to understanding these facts. We now present a formal model based on Barbera and Jackson (2004), which will be useful not only for introducing the notion of self-stable constitutions, but also an alternative model of political choices when individual preferences are revealed over time, first analyzed by Badger (1972).

22.1. Model and Definitions

Consider the following two-period model.

There is a set of *voters* denoted by $N = \{1, \ldots, n\}$.

The voters will face votes over pairs *alternatives*, a and b. Alternative a is interpreted as the *status-quo*. Alternative b is interpreted as a *change*.

Each voter casts a single vote in $\{a, b\}$. Each individual either prefers a or b, and the utility of the preferred outcome is normalized to 1 and that of the other is normalized to 0. This implies that there is no information about "the strength of preferences" in this context.

A *voting rule* is characterized by a number $s \in \{1, \ldots, n\}$. If at least s voters say "b" then b is elected, and a is elected otherwise.

Some examples of voting rules are as follows.

- If $s = 1$, then b is elected whenever there is at least one voter for change, and so a is elected only when it is unanimously supported.
- If $s = n$, then b is elected if there is unanimous support for change, and a is elected as soon as at least one voter supports it.

- If n is odd and $s = (n+1)/2$ or n is even and $s = n/2 + 1$, then the voting rule is the standard *majority rule*.

As majority rule is referred to at several points in what follows we denote it by s^{maj}. Thus, $s^{\text{maj}} = (n+1)/2$ if n is odd and $s^{\text{maj}} = n/2 + 1$ if n is even.

Voters have preferences over voting rules, as the voting rule will affect the future of the society. Let voter i's preferences over voting rules be represented by the utility function $U_i : \{1, \dots, n\} \to [0, 1]$ where $U_i(s)$ represents voter i's utility for voting rule s. It maps onto $[0, 1]$, since the worst that can happen for the individual is that the outcome he dislikes is chosen with probability 1(which would give him utility 0), and the best is that the outcome he prefers is chosen with probability 1 (which would give him utility 1).

A voting rule s is *self-stable (for society p)* if $\#\{i \mid U_i(s') > U_i(s)\} < s$ for every $s' \neq s$. The property of self-stability ensures that a given voting rule would be robust to change if used for making decisions.

A constitution can specify one voting rule to be used on all issues except for the change of this voting rule, where a different rule may be used. A *constitution* is a pair of voting rules (s, S), where s is to be used in votes over the issues a, b and S is to be used in any votes regarding changes from s to any other rule s'.

A constitution $(s, S))$ is *self-stable* if $\#\{i \mid U_i(s') > U_i(s)\} < S$ for any s'.

Self-stability of a constitution requires that the preferences of voters be such that there does not exist a voting rule s' that would defeat the constitution's prescribed voting rule s to be used for choices over issues, when these two voting rules are compared under the constitution's voting rule S, to be used for choices over rules. So, a self-stable constitution is one that would not be changed once in place. Intuitively, we can think that these are the only rules that will survive in the long run in a society, and so it makes sense to understand what they look like.

22.2. Preferences over Voting Rules

Let us focus on a two period world. In period 2, a vote will be taken over two decisions a and b. At this time, each voter knows his or her preferences over a and b. In period 1, voters do not yet know their preferences over a and b.

A voter can be characterized by a probability $p^i \in (0, 1)$, that he or she will prefer b to a at the time of the vote. The realizations of voters' support for the alternatives are independent. For instance, the probability that voters 1 and 2 support b while voter 3 supports a is $p_1 p_2 (1 - p_3)$.

A voter gets utility 1 if his preferred alternative is chosen in the vote, and utility 0 otherwise.

The *society* of voters is represented by a set of voters N and a vector $p = (p_1, \dots, p_n)$, i.e., (N, p).

In this world, a vote over the alternatives a and b will take place in period 2. There are two different times at which a vote over voting rules could be taken and conceivably be relevant. The first is in period 1 where voters do not yet know their preferences over the alternatives (but know the p_i's). The second is in period 2, just before the vote over the alternatives, at a time where voters know which alternatives they support.

When the initial voting rule requires a simple majority or more, the only votes over voting rules that are of any interest turn out to be in period 1, as votes over voting rules in period 2 are of no consequence. This is easily seen as follows.

Suppose that the voting rule is $s \geq n/2$ at the beginning of period 2. Let x be the number of voters who support a, and $n - x$ be the number of voters who support b. If $n - x \geq s$, then b will pass under voting rule s. In this case, these $n - x$ voters will be happy with the voting rule s and would *not* want to change it to any voting rule that would lead b not to pass. Since there are $n - x \geq s$ such voters, no voting rule that could make a difference could defeat s. Next, consider the case where $n - x < s$. In this case, the $n - x$ voters who prefer b would like to lower the voting quota to some $s' < s$, so they could get b to pass. However, the remaining x voters would prefer to keep s as it is, because they prefer a to b. Thus, these voters would vote against any such change, and again the voting rule would not be changed in any way that could make a difference.

This argument may fail if we start from a submajority voting rule. In that case, voting in the second period would be relevant. For example, imagine a situation in which the voting at the beginning of period 2 requires a submajority to change the constitution, whereas the actual vote over the outcomes is a simple majority rule. In this case, it is clear that a small coalition may initially want to change the voting rule in a way to influence the future outcome. This is again an example of how "manipulating institutions" is useful when the distribution of political power is different between today and tomorrow. Since the trade-offs in this respect are already well understood from our previous analyses, let us focus on the case where votes over voting rules are taken at date 1 before voters know their exact preferences. More specifically, we focus on the preferences and votes over voting rules at period 1, when voters know their p_i's but do not yet know their realized preferences over the alternatives.

Given the likelihood of different patterns of support for a and b, a voter can calculate his or her expected utility (at time period 1) under each voting rule s. Let $U_i(s)$ be the expected utility of voter i if voting rule s is used. This is expressed as follows. For any $k \in \{0, \ldots, n-1\}$, let $P_i(k)$ denote the probability that exactly k of the individuals in $N \setminus \{i\}$ support the change. Then

$$P_i(k) = \sum_{C \subset N \setminus \{i\}: |C| = k} \prod_{j \in C} p_j \prod_{\ell \notin C} (1 - p_\ell). \tag{22.1}$$

This expression follows immediately from the fact that the realizations of preferences across individuals are independent, so, for example, the probability that individuals $j \in C$ each favoring change ex ante with probability p_j will ex post all favor changes given by $\prod_{j \in C} p_j$. Given (22.1), the utility of individual i over voting rule s is given by:

$$U_i(s) \;=\; p_i \sum_{k=s-1}^{n-1} P_i(k) \;+\; (1 - p_i) \sum_{k=0}^{s-1} P_i(k). \tag{22.2}$$

The usual definition of single-peaked preferences requires that all alternatives can be ranked from left to right, that one alternative $\widehat{s}$ is best, and that the alternatives that one encounters by moving leftward (or rightward) away from $\widehat{s}$ are considered worse and worse. The definition here will be slightly weaker, as it allows a voter to have two peaks. In particular, it is possible that $U_i(\widehat{s}) = U_i(\widehat{s} - 1)$. For instance, in a society where n is even and each $p_i = p$ for all i, all individuals will be indifferent between $n/2$ and $n/2 + 1$.

U_i is *single-peaked* if there exists $\widehat{s} \in \{1, \ldots, n\}$ with $U_i(\widehat{s}) \geq U_i(s)$ for all $s \in \{1, \ldots, n\}$ such that $U_i(s) > U_i(s-1)$ for any $\widehat{s} > s > 1$ and $U_i(s-1) > U_i(s)$ for any $n \geq s > \widehat{s}$.

Let $\widehat{s}_i$ denote the *peak* of voter i (In the case where a voter has twin-peaks, the definition above selects the higher of the two peaks as $\widehat{s}_i$; but this feature does not matter in any of the results that follow). The following lemma was first proved by Badger (1972):

LEMMA 22.1. *For any society (profile of p_i's), every voter's preferences over voting rules are single-peaked.*

The following example illustrates this lemma:

Example (Single-Peaked Preferences): Let us consider a simple society where agents can be divided into two different groups, $N^1 = \{1, \ldots, 4\}$ and $N^2 = \{5, \ldots, 10\}$, where the p_i of each voter i in group 1 is $p^1 = .01$ and in group 2 is $p^2 = .99$. In this society, the corresponding peaks of preferences over voting rules are $\widehat{s}^1 = 8$ and $\widehat{s}^2 = 4$. To see why $\widehat{s}^2 = 4$, reason as follows: consider a voter in N^2. Consider a scenario where exactly three voters end up supporting change. Given the extreme values of $p^1 = .01$ and $p^2 = .99$, if there are three voters who end up supporting change, it is very likely that all of those voters are from N^2. Given that there are six voters in N^2 this leads to a probability of nearly 1/2 that a voter in N^2 would assign to supporting change conditional on three voters supporting change. Although this probability is nearly 1/2, it is still less than 1/2 due to the small probability that some of the voters in N^1 will be among those supporting change. So, a voter in N^2 will prefer that society choose the status quo conditional on three voters supporting change. If we consider a scenario where exactly four voters end up supporting change, then the conditional probability that a voter in N^2 would assign to being one of the supporters of change is nearly 2/3. Since it is above 1/2, a voter in N^2 will prefer that society choose change conditional on four voters supporting change. Given these two observations it follows that $\widehat{s}_2 = 4$. Similar reasoning leads to $\widehat{s}_1 = 8$.

Generally we can think of a voter considering each possible scenario of numbers of supporters for each of the alternatives. For each scenario the voter determines which group they are more likely to fall in. The voter's most preferred voting rule ($\widehat{s}_i$) corresponds to the scenario with the smallest sized group supporting change for which the voter finds it more likely that he or she will support change. We can see that if the voting rule is raised or lowered from 4, then there will be some scenarios where the choice will be made in favor of the group that the voter finds it less likely that he or she will fall in. This is the explanation for why we see single-peaked preferences. We can also see why it is rare for a voter to have twin peaks - as that can only happen in a case where the voter assigns probability of exactly 1/2 to each of the two groups in some scenario.

While Lemma 22.1 implies that each voter's preferences over voting rules have the nice property of single-peakedness, the following lemma tells us about how different voters' preferences are related to each other. There are two properties that are useful in noting.

A society of voters has preferences satisfying the *single crossing property* if for any i and j with $p_j \geq p_i$,

$$U_i(s) - U_i(s') \geq U_j(s) - U_j(s')$$

for all $s \geq s'$. The single crossing property is satisfied in this model. The single crossing property allows us to order preferences over voting rules in terms of the p_i's; but more importantly also implies that the preferences are intermediate.

Similarly to the terminology introduced in the notes on *Introduction to the Theory of Voting, Lobbying and Political Agency*, we say that a society of voters has *intermediate preferences* if for any i, j, k with $p_j \geq p_k \geq p_i$:

- $U_i(s) \geq U_i(s')$ and $U_j(s) \geq U_j(s')$ imply that $U_k(s) \geq U_k(s')$, and

- $U_i(s) > U_i(s')$ and $U_j(s) > U_j(s')$ imply that $U_k(s) > U_k(s')$.

Intermediate preferences are usually defined by requiring that there exists some ordering over individuals so that when two individuals have the same ranking over two alternatives, then individuals between them in the ordering have that same ranking. Here the natural ordering over individuals is in terms of their p_i's, the distinguishing characteristic of voters, and so we take the shortcut of defining intermediate preferences directly in terms of that ordering. Hence, a society will have intermediate preferences over voting rules if whenever two voters with p_i and p_j agree on how to rank two rules s and s', then all voters with probabilities p_k between p_i and p_j will also agree on the way to rank these two rules. The simple model we are considering has the following strong features.

Proposition 22.1. *Every society has preferences over voting rules that satisfy the single crossing property and are intermediate.*

Proposition 22.2. *For any society, $\widehat{s}_i \geq \widehat{s}_j$ whenever $p_j \geq p_i$.*

The intuition for why the voters' peaks over voting rules follow an inverse order to the voters p_i's is straightforward, as voters with higher p_i's are more likely to favor change and thus will be in favor of a lower quota than voters who are less likely to favor change.

There are some other facts about the location of the voters' peaks that are worth emphasizing. The *relative* ordering of p_i's is not only important in determining the relative ordering over the $\widehat{s}_i$'s, but it is also critical in determining the actual values of the $\widehat{s}_i$'s. This is seen in the following proposition, which states that *regardless* of p, there is always some voter who has a peak at least as high as the majority rule, s^{maj}, and some other voter who has a peak no higher than s^{maj}.

Proposition 22.3. *For any society there exist i and j such that $\widehat{s}_i \geq s^{\text{maj}} \geq \widehat{s}_j$.*

The intuition (or alternatively the sketch proof) for this proposition is as follows. The unique maximizer of $\sum_i U_i(s)$ is s^{maj}, since s^{maj} chooses the alternative that will result in the largest group of voters who get utility 1 for each realization of preferences over a and b (while the losers get utility 0, by assumption). Thus, if some voter's expected utility is increased by moving to an s that is higher than s^{maj}, then some other voter's expected utility must fall as the result of such a move. The same is true in reverse. So there is at least one voter with a peak at least as high as s^{maj} and at least one voter with a peak no higher than s^{maj}.

Note that by combining Propositions 22.2 and 22.3, we know that the voter who has the highest p_i must have a $\widehat{s}_i$ which is no higher than s^{maj} and the voter who has the lowest p_i must have a $\widehat{s}_i$ that is at least as high as s^{maj}, *and this is true regardless of* p.

The following result, again first derived by Badger (1972), shows that majority rule has special properties.

Proposition 22.4. *For any society (profile of p_i's), the only voting rules that maximize the sum of voters' expected utilities is s^{maj} if n is odd, and s^{maj} and $s^{\mathrm{maj}} - 1$ if n is even.*

Given any realization of voters' preferences at time 2, the choice which maximizes the realized total utility is simply to choose the alternative preferred by a majority. Given that this is the best that one can do realization by realization, it is maximizing in total expectation as well. Any rule other than majority rule (except $n/2$ when n is even) realizes a lower total utility at some realization of preferences as it will select one of the alternatives when a minority supports it, and thus we have the uniqueness claim.

22.3. Self-Stability

We will now see that many societies have some self-stable voting rules, and all societies have self-stable constitutions. Adding a special rule to change rules is therefore a stabilizing factor—at some level, this is trivial; consider the rule that no existing rule can be changed. More interesting than this stabilizing role of two-stage voting procedures is to understand that what voting rules and constitutions are self-stable depends on the parameters of a society. In particular, the attitudes of a society's citizens toward change is a critical factor in determining which voting rules and which constitutions turn out to be self stable. Though the Barbera-Jackson model is rather sparse in terms of the underlying economic and social environment, the issues they focus on are a natural first step. Making these structures richer may help to explain some of the facts we discussed at the start of these notes.

Let us start with the special case where all voters have the same p_i. This is of some interest where this common p is an indicator of the average propensity to favor change of a society's representative voter. The following proposition is simple but at some level also surprising:

Proposition 22.5. *If $p_i = p_j$ for all i and j, then s^{maj} is the unique self-stable voting rule if n is odd. If n is even, then there are two self-stable rules s^{maj} and $s^{\mathrm{maj}} - 1$.*

Thus, majority rule is the unique self-stable voting rule whenever all voters have the same probability of choosing change, irrespective of what this probability might be. One might have guessed that societies where all voters are very likely to want changes would prefer low values of s, that is low barriers to change, and that homogeneously conservative societies would favor high values of s. But the proposition shows that this is not the case; in homogeneous societies, all voters have their peak at $\widehat{s}_i = s^{\text{maj}}$, and thus majority rule is the consensus choice of rule. What actually matters is not the absolute values of the p's but their values relative to those of other voters, because each individual does not know what his or her preferences will be, but evaluates the likelihood of a coalition including himself or herself being the majority at the end.

To illustrate this point, consider a society where $p_i = .01$ for each i and so voters are very conservative and very likely to support the status-quo. In this case, shouldn't it be that voters all prefer a high quota s as they each know they are likely to support the status quo? The answer is no and the reasoning lies in the answer to the following question. Which alternative would a voter prefer society to choose in a generic realization where k voters end up supporting a and $n - k$ voters end up supporting b? That is, the voter can think of the different scenarios possible for numbers of voters supporting a and b, and then ask which side he is most likely to fall on in each scenario. Given the symmetry in p_i's, conditional on this realization of preferences it is most likely that the voter is in the larger of the two groups. So, the voter would like society to choose a in scenarios where $k > n - k$ and society to choose b in scenarios where $k < n - k$, and is indifferent if $k = n - k$. Thus, the voter would like society to choose in favor of the majority as that is where the voter is most likely to be in any realization. Once one understands the above reasoning, then Proposition 22.3 and the importance of relative comparisons becomes clear.

This proposition therefore shows that, under some circumstances, the unique self-stable rule is the efficient majority rule. Unfortunately, the substantial symmetry in a homogeneous society is responsible for the nice conclusion of the result. In more heterogeneous societies, one can lose majority rule as being self-stable, and one can also lose existence of a self-stable voting rule altogether.

Let us next look at an example where there exist self-stable voting rules, but where majority rule is not self-stable. Reconsider the example "single-peak preferences" above. Recall that the society in that example consisted of two groups of voters, $N^1 = \{1, \ldots, 4\}$ with $p^1 = 0.01$, and $N^2 = \{5, \ldots, 10\}$ with $p^2 = 0.99$. Recall that the corresponding peaks of preferences over voting rules were $\widehat{s}^1 = 8$ and $\widehat{s}^2 = 4$. Now it can be verified that $\{7, 8\}$ is the set of self-stable voting rules. It is easy to see that 8 is self-stable as only group N^2 would like to change voting rules if 8 is used, but then they only have 6 members and so are too small to make the change under a rule of 8. The same is true of quota 7, and although in that case group N^1 would like to raise the quota from 7 to 8, it is too small to do so. To see that no other rule is stable, note that 4 is unanimously preferred to any smaller rule, and 8 is unanimously preferred to any larger rule. So the only other candidates for self-stability are the quotas 4, 5, and 6. However, 5 and 6 are not stable because N^2 prefers 4 and has enough voters to move the quota to 4. 4 is not stable since group N^1 would have enough voters to increase the quota.

Example (A Society for which No Rule is Self Stable): Suppose $N = \{1, \ldots, 5\}$. $p_1 = p_2 = p_3 = 1/2$, $p_4 = 3/8$, and $p_5 = 3/16$. Direct calculations lead to $\widehat{s}_1 = \widehat{s}_2 = \widehat{s}_3 = 2$, $\widehat{s}_4 = 3$ and $\widehat{s}_5 = 4$. Let us verify that there is no self-stable voting procedure. All voters want to raise the quota from 1 and lower it from 5. That leaves the quotas of 2, 3, and 4 to be checked as the only possibilities for self-stable voting rules. Voters 1 to 3 would vote to lower it from 3 to 2, voters 1 to 4 would vote to lower it from 4 to 3, and voters 3 and 4 would vote to raise it from 2 to 3. Thus, no voting rule is self-stable.

This example therefore shows that a society may not have a self-stable voting rule. To understand when this happens, let us define some additional terms.

A society (N, p) is *dichotomous* if there exists $N^1 \neq \emptyset$, $p^1 \in (0, 1)$, $N^2 \neq \emptyset$, and $p^2 \in (0, 1)$ such that $N = N^1 \cup N^2$, $p_i = p^1$ for all $i \in N^1$, $p_i = p^2$ for all $i \in N^2$.

A dichotomous society is thus one that can be divided into two groups such that members of the same group have the same p_i's, as in the example above.

Say that a society is *symmetric* if when voters are labeled such that $p_i \geq p_j$ when $i > j$, it follows that $p_i = 1 - p_{n-i}$.

Let $\widehat{s}_{\text{med}}$ denote the median of $(\widehat{s}_1, \ldots, \widehat{s}_n)$, i.e., the median of the peaks of the voters. We then have the following results:

Proposition 22.6. (1) *If $\widehat{s}_{\text{med}} \geq s^{\text{maj}}$, then $\widehat{s}_{\text{med}}$ is self-stable.*

(2) *If there does not exist a self-stable voting rule for a society (N, p), then there exists a self-stable voting rule for the society $(N, \overline{p})$, where $\overline{p}$ is defined by $\overline{p}_i = 1 - p_i$ for each i. Moreover, $\widehat{s}_{\text{med}}$ is self-stable for society $(N, \overline{p})$.*

(3) *A dichotomous society has at least one self-stable voting rule.*

(4) *If a society is symmetric, then s^{maj} is a self-stable voting rule.*

Intuitively, part 1 follows because: if $\widehat{s}_{\text{med}} \geq s^{\text{maj}}$, then at most half of the population would like to lower the rule below the median, and at most half would like to increase it above the median. Since $\widehat{s}_{\text{med}} \geq s^{\text{maj}} \geq (n + 1)/2$, it follows that $\widehat{s}_{\text{med}}$ be self-stable.

The proof of part 2 follows from the observation that the setting we are examining is symmetric in the following way: if in society (N, p) voter i would like society to choose b conditional only on knowing that s voters out of society favor b, then in society $(N, \overline{p})$ voter i would like society to choose a conditional only on knowing that s voters out of society favor a. This implies that if $\widehat{s}_i$ is i's peak under society (N, p), then $n - \widehat{s}_i + 1$ is i's peak under society $(N, \overline{p})$. To establish part 2, note that non-existence of a self-stable voting rule implies that $\widehat{s}_{\text{med}}$ is no larger than $n/2$, as otherwise it would be self-stable. The reasoning above then implies that $\widehat{s}_{\text{med}}$ for society $(N, \overline{p})$ is larger than $n/2$, and so is stable.

Part 4, which asserts the existence of self-stable voting rules for symmetric societies, is an easy corollary of part 2.

The proof of part 3 is more involved: it works by relating the conditional beliefs of the two groups to each other. Let N^1, N^2, n^1, n^2, $\widehat{s}^1$, and $\widehat{s}^2$ be the two groups of voters, the cardinalities of these groups, and their peaks, respectively. The main case that has to be ruled out to establish existence is where $n^2 \geq \widehat{s}^1$ and $n^1 \geq \widehat{s}^2$, when $\widehat{s}^1 \neq \widehat{s}^2$. If the beliefs of N^1 are such that $n^2 \geq \widehat{s}^1$, this means that the voters in N^1 have *relatively* high beliefs that they will be among the supporters of b. This implies that the voters in N^2 have *relatively* low beliefs that they will be among the supporters of b, and so $\widehat{s}^2$ will be high enough to be larger than n^1. The challenge in the proof is to show that these relative statements translate into absolute statements about the relationship between $\widehat{s}^1$ and $\widehat{s}^2$ and their comparison to n^1 and n^2.

This proposition, especially part 2, has some powerful implications. It implies that non-existence is a problem for less than "half" of the potential societies, in terms of the p's. This implies that while non-existence can occur for open sets of societies, it still is a problem that is not completely pervasive.

It is also useful to note that when self-stable voting rules exist, there may be a number of them. Moreover, the set of self-stable voting rules need not be an interval, nor need it include s^{maj}. These points are illustrated in the following example.

Example (A Society with Multiple and Non-Adjacent Self Stable Rules): The society (N, p) is dichotomous. $N^1 = \{1, \ldots, 5\}$ and $N^2 = \{6, \ldots, 16\}$ with $p^1 = .01$ and $p^2 = .99$. Here $\widehat{s}^1 = 14$ and $\widehat{s}^2 = 6$. It follows that $\{6, 12, 13, 14\}$ is the set of self-stable voting rules. It is clear that the set of self-stable voting rules will consist of a set of intervals, each of which includes at least one $\widehat{s}_i$. This puts an upper bound on the number of disjoint intervals that can be included, at the number of distinct p_i's that are present in the society.

It is also interesting to note that rules with $s < s^{\text{maj}}$ can be problematic in the following sense. Consider a situation where a and b are each supported by half of the population. A vote under s will result in b becoming the new status quo. But then, with b as the new status quo, the other half of the voters would support (and could ensure) a change back to a if it is proposed for a vote against b. Thus, there is the potential to continuously cycle back and forth between a and b as the status quo. This, of course, is only a potential problem of sub-majority rules.

Suppose that a society somehow precludes itself from ever selecting a sub-majority rule. If this is the case, then the existence of self-stable voting rules is ensured. To see this, consider such a society. The preferences of voters over the restricted set of s's ($s \geq s^{\text{maj}}$) are still single peaked. Voters whose unrestricted peaks were at least s^{maj} have the same peak on the restricted set, while voters whose peaks were below s^{maj} now have s^{maj} as a peak. The median of the restricted peaks will be self-stable over the restricted set of voting rules. This leads to the following result

PROPOSITION 22.7. *For any society where only $s \geq s^{\text{maj}}$ are admissible voting rules, $\widehat{s}_{\text{med}}$ (defined relative to restricted preferences) is a self-stable voting rule.*

This proposition follows from part (1) of Proposition 22.6. Hence, for societies that exclude s's below s^{maj} a priori, self-stability would favor median voting rules in practice.

We have so far not looked at whether self-stable pairs of constitutions and voting rules always exist. The following proposition answers this in the affirmative:

PROPOSITION 22.8. *For any society, the constitutions (s^{maj}, n) and $(\widehat{s}_{\text{med}}, S)$ for any $S \geq s^{\text{maj}}$ are self-stable.*

This result follows in a trivial manner if we allow $S = n$, so that only unanimous choice can change the voting rule. This would be a trivial example of a self-stable pair of constitutions and voting rules.

More interesting ones also exist, however. To see this, we simply need to use the fact that preferences are intermediate preferences (Proposition 22.1) and relative positioning of voter's peaks (Proposition 22.3). The idea is as follows. The self-stability of (s^{maj}, n) follows from the observation that by Proposition's 22.3 and 22.1 there is always at least one voter who will wish to keep the voting rule over issues no higher than s^{maj} and at least one who will wish to keep the voting rule no lower than s^{maj}. Thus, there is no unanimous consent to raise or lower the voting rule from s^{maj}. The self-stability of $(\widehat{s}_{\text{med}}, S)$ with $S \geq s^{\text{maj}}$ follows from Proposition 22.1 and the definition of $\widehat{s}_{\text{med}}$, as by intermediate preferences fewer than $n/2$ voters will prefer to raise the voting rule from $\widehat{s}_{\text{med}}$, and similarly fewer than $n/2$ voters will prefer to lower the voting rule from $\widehat{s}_{\text{med}}$.

The self-stability of constitutions using majority rule as a voting procedure is of particular interest because of the prominence of majority rule in actual constitutions and its special properties including overall efficiency (Proposition 22.4). For example, the particular constitution (s^{maj}, n) is self-stable for any society.

We can next explore the conditions on the distribution of p_i's that are sufficient for other constitutions (s^{maj}, S) to be self-stable for values of $S < n$.

Let $z_i = p_i/(1-p_i)$. Thus, z_i represents the ratio of the probability that i supports change compared to the probability that i supports the status quo. Any positive number is a potential z_i.

PROPOSITION 22.9. *For any society with even n the constitution (s^{maj}, S) is self-stable if*

$$S > \#\left\{i : \sum_{C\subset N, |C|=n/2, i\in C} \left(\prod_{j\in C} z_j\right) \geq \sum_{C\subset N, |C|=n/2, i\notin C} \left(\prod_{j\in C} z_j\right)\right\} > n - S. \tag{22.3}$$

Note that (22.3) can be rewritten as

$$S > \#\left\{i : z_i \geq \sum_{k\neq i} \lambda_k^i z_k\right\} > n - S, \tag{22.4}$$

where

$$\lambda_k^i = (2/n)\frac{\sum_{|C|=n/2-1; i,k\notin C}\left(\prod_{j\in C} z_j\right)}{\sum_{|C|=n/2-1; i\notin C,}\left(\prod_{j\in C} z_j\right)}.$$

Here, the λ_k^i are weights such that $\sum_{k\neq i}\lambda_k^i = 1$, and so $\sum_{k\neq i}\lambda_k^i z_k$ is a weighted average of z_k's over k's other than i. Thus, (22.4) says roughly that the number of voters with above average z_i's is not too high and not too low. It can be shown that this is also equivalent to having the number of voters with below average z_i's not be too high or too low.

To see the implications of Proposition 22.9, let us consider the constitution where $s' = 2n/3$. That constitution is stable, provided there are at least 1/3 of the voters who do not wish to raise the voting rule from s^{maj} and at least 1/3 of the voters who do not wish to lower it from s^{maj}. The proof of the proposition involves showing that these are equivalent to the inequalities relating the z_i's. The requirements of the proposition are then that at least 1/3 and no more than 2/3 of the voters have a z_i that is bigger than the weighted average of the other voters' z_i's. This is in effect a limitation on the skewness of the distribution of the z_i's (or, in effect, the p_i's). If the distribution of z_i's is not too skewed, then $(s^{\text{maj}}, 2n/3)$ will be self-stable.

More generally, Proposition 22.9 provides the reasoning behind why a super-majority will be required for rules changes in a constitution where majority rule is used for ordinary decisions.

22.4. Stability and Reform of Political Institutions

Let us next look have a brief look at a recent paper by Roger Lagunoff (2006), which generalizes the insights of the models of franchise extension to a more general setting. In many ways, this can be thought of as another model of the stability of constitutions.

Consider the following setting: we have an infinite-horizon dynamic game in which the state variable is $x_t = (\theta_t, \omega_t)$ where $\theta_t \in \Theta$ is a variable capturing the political rule/state, i.e., the method of making collective choices in society. In the Acemoglu-Robinson (2006) model this is dictatorship versus democracy, while in the Roberts (1999) model, this is the size of the franchise. The other variable is $\omega_t \in W$ and includes other state variables, such as shocks. In the Acemoglu-Robinson (2006), the ability of different groups to solve their collective action problem would be captured by this variable.

In a Markov Perfect Equilibrium of this dynamic game, strategies of the players define a mapping determining some policy choice $p_t \in P$ as well as next period's political state θ_{t+1}. Therefore, a stationary equilibrium can be represented as a mapping

$$\phi : \Theta \times W \to \Theta \times P.$$

We say that there exists reform (endogenous change) in political institutions if there exists θ' and a positive measure subset of W, B, such that

$$\phi\left(\theta', B\right) \neq \left(\theta', \cdot\right).$$

Or more simply, we can decompose the mapping ϕ, which determines both policies and future institutions into two parts:

$$\mu : \Theta \times W \to \Theta$$

determines political transitions, while

$$\psi : \Theta \times W \to P$$

determines policies. We can then refer to the pair (μ, ψ) as a pair of political strategies. There will be endogenous reform if

$$\mu\left(\theta', B\right) \neq \theta'$$

for a positive measure subset B of W.

The use of this framework is that it is sufficiently general to address and nest a range of models/questions. This generality comes at the cost of being specific about how decisions are made etc.. Nevertheless, it is a useful generality.

For example, Lagunoff (2006) makes further progress by assuming that different institutional states can be represented by "social welfare functions". This avoids problems of non-existence of collective choices or voting cycles. Given this simplifying (reduced-form) assumption, Lagunoff is able to establish an interesting result.

To prepare for this result, recall that μ determines the transition of the political state. Let P denote the policy space with generic element p. Recall that $\psi : \Theta \times W \to P$ is the policy function, mapping from political and other state variables into the current policy. Finally, $u(p, \omega)$ is an instantaneous utility function, defined in the usual way and δ is the discount factor.

Lagunoff considers political environments in which preferences can be represented by a social welfare function of the form $F(v(p,\theta), \omega, \theta)$ where $v(p,\theta)$ is the value of the decisive voter under political system θ when the policy is $p \in P$. Finally, $q(\omega' | \, \omega, p))$ is the distribution function of the state ω as a function of past state and policies.

For example, $\theta \in \Theta$ may correspond to a dictatorial rule, in which case $v(p,\theta)$ would correspond to the utility of the dictator, and the social welfare function F would put all the weight on this utility. In general, existence of equilibria in such environments is non-trivial (see Lagunoff, 2004), but we will not focus on these issues here.

We say that given the social welfare function F, a rule θ is *dynamically consistent* if for every pair of mappings (ψ, μ) and every $\omega \in W$, we have that

$$
\begin{aligned}
(22.5) \qquad & \arg\max_{\theta'} \; F\left((1-\delta)u(\omega, \psi(\omega,\theta)\,) \; + \; \delta \int V(\omega', \theta';\; \psi, \mu) dq(\omega' | \, \omega, \psi(\omega, \theta)), \; \omega, \theta\right) \\
= \quad & \arg\max_{\theta'} \delta \int F\left(V(\omega', \theta';\; \psi, \mu), \; \omega', \theta\right) dq(\omega' | \, \omega, \psi(\omega, \theta))
\end{aligned}
$$

whenever the two sets of maximizers are nonempty.

This definition implies that the social welfare function F would choose the same political state for tomorrow from today's perspective (fixing the current utility profile), as it would do tomorrow. This is similar to dynamic consistency for an individual, but note that here dynamic inconsistency can arise when all individuals have dynamically consistent preferences themselves, but political power shifts over time as in the models we have seen so far.

There are natural candidates for dynamically consistent social preferences, including that of the fictitious social planner with geometrically discounted payoffs:

$$F\left((1-\delta)u(\omega,p) \ + \ \delta \int V(\omega',\theta';\ \psi,\mu)dq(\omega'|\,\omega,p),\ s\right) \tag{22.6}$$
$$= \ (1-\delta)F\left(u(\omega,p),\ \theta\right) \ + \ \delta \int_{\omega'} F\left(V(\omega',\theta';\ \psi,\mu),\ \theta\right) dq(\omega'|\,\omega,p)$$

A central result of Lagunoff (2006) is the following:

Proposition 22.10. *Let F possess a single maximizer and (ψ,μ) be any politically feasible pair. Then a political rule θ is stable in μ if and only if it is dynamically consistent.*

Proof. The necessity part of the proof —the fact that dynamically consistent θ implies stability—is immediate.

Sufficiency: we need to show that whenever θ is stable in μ, then θ is dynamically consistent.

By stability, $\mu(\omega,\theta)=\theta$ a.e. ω. Because the rule is single valued, political feasibility of (ψ,μ) then implies

$$\theta = \arg\max_{\tilde{\theta}} F\left((1-\delta)u(\omega,\psi(\omega,\theta)\ ) + \delta\int V(\omega',\tilde{\theta};\psi,\mu)dq(\omega'|\,\omega,\psi(\omega,\theta)\ ),\omega,\theta\right)$$

Suppose to derive a contradiction, that θ is not dynamically consistent. Then

$$\theta \notin \arg\max_{\tilde{\theta}} \int F\left(V(\omega',\tilde{\theta};\psi,\mu),\omega',\theta\right) dq(\omega'|\,\omega,\psi(\omega,\theta)\ ) \tag{22.7}$$

where the set of maximizers on the right-hand side of (22.7) is nonempty (since otherwise, θ is trivially dynamically consistent). Let $\hat{\theta}$ denote a maximizer of (22.7). Since $\hat{\theta}\neq\theta$, there is some set of ω' with positive measure such that

$$F\left(V(\omega',\hat{\theta};\psi,\mu),\omega',\theta\right) \ > \ F\left(V(\omega',\theta;\psi,\mu),\omega',\theta\right) \tag{22.8}$$

However, by the definition of political feasibility, for all ω', we have that

$$\begin{aligned} & F\left(V(\omega',\theta;\psi,\mu),\omega',\theta\right) \\ = \ & \max_{\tilde{p},\tilde{\theta}} \ F\left((1-\delta)u(\omega',\tilde{p})\ ) + \delta\int V(\omega'',\tilde{\theta};\psi,\mu)dq(\omega''|\,\omega',\tilde{p}\ ),\omega',\theta\right) \\ \geq \ & F\left((1-\delta)u(\omega',\psi(\omega',\hat{\theta})\ ) + \delta\textstyle\int V(\omega'',\mu(\omega',\hat{\theta});\psi,\mu)dq(\omega''|\,\omega',\psi(\omega',\hat{\theta})\ ),\omega',\theta\right) \\ = \ & F\left(V(\omega',\hat{\theta};\psi,\mu),\omega',\theta\right). \end{aligned} \tag{22.9}$$

This implies $F\left(V(\omega',\theta;\psi,\mu),\omega',\theta\right) \geq F\left(V(\omega',\hat{\theta};\psi,\mu),\omega',\theta\right)$, which contradicts (22.8) and establishes that θ is dynamically consistent, completing the proof. □

Lagunoff (2006) shows that the same results generalize when there are private decisions (in addition to the policies $p\in P$), but private decisions (such as investment or insurrection decisions) do not interact with public policies. Yet when private and public decisions do interact, even richer dynamics can arise when private decisions interact with public policies. Lagunoff (2006) illustrates that a special case of such interaction is the Acemoglu-Robinson (2006) model of democratization.

22.5. Dynamics and Stability of Constitutions, Coalitions and Clubs

Let us now discuss a model that is rich enough to nest a variety of approaches presented above. The model here is based on Acemoglu, Egorov and Sonin (2000 it) and provides a general inflexible framework for thinking of the stability of constitutions and political arrangements.

22.5.1. Environment. Let us first introduce the general environment. There is a finite set of players $\mathcal{I}$. Time is discrete and infinite, indexed by t ($t \geq 1$). There is a finite set of *states* which we denote by $\mathcal{S}$. We We denote the number of elements of the sets $\mathcal{I}$ and $\mathcal{S}$ by $|\mathcal{I}|$ and $|\mathcal{S}|$, respectively. Recall that these states may represent different institutions simply affecting payoffs, or constitutions that may affect both payoffs and the procedures for decision-making (e.g., the ruling coalition in power, the degree of supermajority, the weights or powers of different agents). Although our game is one of non-transferable utility, a limited amount of transfers can also be incorporated, by allowing multiple (but a finite set of) states with the same procedure for decision-making but with a reallocation of payoffs across players.

The initial state of the world is denoted by $s_0 \in \mathcal{S}$ and is taken as given. For any $t \geq 1$, the state $s_t \in \mathcal{S}$ is endogenously determined. A non-empty set $X \subset \mathcal{I}$ is called *coalition*, and we denote the set of coalitions by $\mathcal{C}$ (that is, $\mathcal{C}$ is the set of nonempty subsets of $\mathcal{I}$). Each state $s \in \mathcal{S}$ is characterized by a pair $(w_s(\cdot), \mathcal{W}_s)$. Here, for each fixed state $s \in \mathcal{S}$,

$$w_s : \mathcal{I} \to \mathbb{R}_{++}$$

is a mapping assigning a positive stage payoff $w_s(i)$ to each individual $i \in \mathcal{I}$ ($w_s(i) > 0$ is useful as a normalization that makes zero payoff the worst outcome); $\mathcal{W}_s$ is a (possibly empty) subset of $\mathcal{C}$ representing the set of *winning coalitions* for state s. Throughout the paper, we maintain the following assumption.

Assumption 22.1. ***(Winning Coalitions)*** *For any state $s \in \mathcal{S}$, $\mathcal{W}_s \subset \mathcal{C}$ satisfies two properties:*

(a) If $X, Y \in \mathcal{C}$, $X \subset Y$, and $X \in \mathcal{W}_s$ then $Y \in \mathcal{W}_s$.

(b) If $X, Y \in \mathcal{W}_s$, then $X \cap Y \neq \varnothing$.

Part (a) simply states that if some coalition X is winning for state s, then increasing the size of the coalition would not reverse this. This is a natural assumption for almost any decision rule. Part (b) rules out the possibility that two disjoint coalitions could be winning for the same state, thus imposing a form of (possibly weighted) majority or supermajority rule. Notice that $\mathcal{W}_s = \varnothing$ is not ruled out by this assumption. If $\mathcal{W}_s = \varnothing$, then state s is *exogenously stable*. For each state $s \in \mathcal{S}$ we define the set of *blocking* coalitions by $\mathcal{B}_s = \{X \in \mathcal{C} \mid \mathcal{I} \setminus X \notin \mathcal{W}_s\}$. Clearly, $\mathcal{B}_s \subset \mathcal{W}_s$, meaning that winning coalitions are also blocking (but not necessarily vice versa).

We define the following binary relations on the set of states $\mathcal{S}$. For $x, y \in \mathcal{S}$, we write

$$x \sim y \iff \forall i \in \mathcal{I} : w_x(i) = w_y(i)\,. \tag{22.10}$$

In this case we call states x and y *payoff-equivalent*, or simply, *equivalent*. For any $z \in \mathcal{S}$, the binary relation $\succeq_z$ is defined by

$$y \succeq_z x \iff \{i \in \mathcal{I} : w_y(i) \geq w_x(i)\} \in \mathcal{W}_z. \tag{22.11}$$

If (22.11) holds, we say that y is *weakly preferred* to x in z. Relation $\succ_z$ is defined by

$$y \succ_z x \iff \{i \in \mathcal{I} : w_y(i) > w_x(i)\} \in \mathcal{W}_z. \tag{22.12}$$

If (22.12) holds, we say that y is *strictly preferred* to x in z. Relation $\sim$ clearly defines an equivalence class, in that if $x \sim y$ and $y \sim z$, then $x \sim z$. In contrast, the binary relations $\succeq_z$ and $\succ_z$ may not even be transitive. Nevertheless, for any $x, z \in \mathcal{S}$ we have $x \not\succ_z x$, and whenever $\mathcal{W}_z$ is not empty (i.e., z is not exogenously stable), $x \succeq_z x$. Also, for any $x, y, z \in \mathcal{S}$, $y \succ_z x$ implies $x \not\succeq_z y$, and similarly $y \succeq_z x$ implies $x \not\succ_z y$; these implications follow from Assumption 22.1.

The following assumption introduces some basic properties of payoff functions.

ASSUMPTION 22.2. ***(Payoffs)*** *Payoff functions $\{w_i(\cdot)\}_{i \in \mathcal{I}}$ satisfy the following properties:*

(a) For any nonempty collection of states $\mathcal{Q} \subset \mathcal{S}$, there exists state $z \in \mathcal{Q}$ such that for any $x \in \mathcal{Q}$, $x \not\succ_z z$.

(b) For any $s \in \mathcal{S}$, let $\mathcal{Q} = \{y \in \mathcal{S} : y \succ_s s\}$. If $\mathcal{Q}$ is non-empty, then there exists $z \in \mathcal{Q}$ such that for any $x \in \mathcal{Q}$, $x \not\succ_s z$.

Part (a) of Assumption 22.2 requires that within any collection of states there exists a state z such that the set of players that prefer another state is not sufficiently large (not winning in z). Part (b) of Assumption 22.2, on the other hand, requires the same for winning coalitions within a given state s (naturally this part of the assumption is trivially satisfied if $\mathcal{Q}$ only has two elements). Assumption 22.2 will play a major role in our analysis and ensures "acyclicity". In particular, note that part (a) rules out cycles of the form $y \succ_z z$, $x \succ_y y$, $z \succ_x z$, while part (b) rules out cycles of the form $y \succ_s z$, $x \succ_s y$, $z \succ_s z$. It is easy to construct examples to show that neither of the two parts of Assumption 22.2 follows from the other.

In addition to Assumptions 22.1 and 22.2, which are natural in this context and are assumed to hold throughout the paper, we will also sometimes impose:

Assumption 22.3. ***(Comparability)*** *For $x, y, z \in \mathcal{S}$ such that $x \succ_z z$, $y \succ_z z$, and $x \nsim y$, either $y \succ_z x$ or $x \succ_z y$.*

This assumption states that if two states y and z are weakly preferred to x (in x), then y and z are $\succ_x$-comparable. It turns out that this condition is precisely the one necessary to guarantee uniqueness of equilibria of dynamically stable states. This assumption is not necessary for a range of our results, and for this reason, some of our main results are stated without imposing it.

At each date, each individual maximizes her discounted expected utility:

$$U_t(i) = (1-\beta)\sum\nolimits_{\tau=t}^{\infty} \beta^{\tau} u_{\tau}(i), \tag{22.13}$$

where $\beta \in (0,1)$ is a common discount factor and for now we can think of $u_t(i)$ as given by the payoff function $w_i(\cdot)$ introduced in Assumption 22.2. Throughout, we will consider situations in which β is large.

We define $s^{\infty} \in \mathcal{S}$ as a *dynamically stable state* if after some time $T < \infty$, the state is s^{∞} (i.e., $s_t = s^{\infty}$ for all $t \geq T$), provided that the game started in this state. In particular, our objective is to determine when such dynamically stable states exist and to characterize which (dynamically stable) state will be reached starting from an (arbitrary) *initial state* $s_0 \in \mathcal{S}$.

22.5.2. Axiomatic Characterization. It is useful to start with an axiomatic characterization before moving to the extensive-form game, specifying how voting decisions and transitions take place, will be presented in the next section. Before presenting this extensive-form game and its analysis, the essential economic forces can be succinctly captured by an axiomatic characterization of the dynamically stable states. The key economic insight enabling this axiomatic characterization is the following: *when individuals are patient, an individual should not wish to transition to a state that will ultimately lead to another state giving her lower utility.* This basic insight enables a tight characterization of dynamically stable states.

Our axiomatic characterization will determine a mapping

$$\phi : \mathcal{S} \to \mathcal{S}$$

that will assign to each initial state $s_0 \in \mathcal{S}$ a dynamically stable state $s^\infty \in \mathcal{S}$. This axiomatic characterization will therefore bypass the analysis of the dynamics leading to this stable state, but simply determine the dynamically stable states given the initial state $s_0 \in \mathcal{S}$.

In the spirit of the discussion in the previous two paragraphs, we impose the following three axioms on this mapping.

Axiom 5. ***(Desirability)*** *If $x, y \in \mathcal{S}$ are such that $y = \phi(x)$, then either $y = x$ or $y \succ_x x$.*

Axiom 6. ***(Rationality)*** *If $x, y, z \in \mathcal{S}$ are such that $z \succ_x x$, $z = \phi(z)$, and $z \succ_x y$, then $y \neq \phi(x)$.*

Axiom 7. ***(Stability)*** *If $x, y \in \mathcal{S}$ are such that $y = \phi(x)$, then $y = \phi(y)$.*

All three axioms are natural in light of what we have discussed above. Axiom 5 essentially says that the population will not move to another state unless there is a winning coalition that supports this transition (for example, depending on the specification, $y \succ_x x$ might mean that starting with state x and majority rule, the majority of the population will vote for a reform towards y). Axiom 6 imposes the idea that if there exists some state z preferred to y by the group of decisive individuals starting in state x, then ϕ should not pick y ahead of z starting in x. Finally and most importantly, Axiom 7 encapsulates the stability notion discussed above—that an individual should not prefer a state that will ultimately lead to another, less preferred state. This notion is economically captured by the statement that if mapping ϕ will pick state y starting from state y, then it should also pick y starting from y (otherwise, y would lead to another state z, and as stated by Axiom 6, if this state z were indeed preferred to y, then ϕ would have picked z in the first instance).

Notice also that all of these notions apply to individual preferences. Since collective decision-making aggregates individual preferences, these axioms then indirectly apply to the mapping ϕ (e.g., ϕ might aggregate individual preferences according to majority rule or weighted supermajority rule, etc.).

The next definition reiterates the meaning of dynamically stable states and its relationship to mapping ϕ.

Definition 22.1. ***(Dynamically Stable States)*** *For any $\phi : \mathcal{S} \to \mathcal{S}$ that satisfies Axioms* 1–3*, a state $s^{\infty} \in \mathcal{S}$ is* dynamically stable *if $\phi(s^{\infty}) = s^{\infty}$. The set of dynamically stable states is $\mathcal{D} = \{s \in \mathcal{S}\colon \phi(s) = s\}$.*

The next theorem is one of our main results. It establishes the existence of dynamically stable states and provides a recursive characterization of such states.

Theorem 22.1. ***(Axiomatic Characterization of Dynamically Stable States)*** *Suppose Assumptions 22.1 and 22.2 hold. Then:*

(1) *There exists mapping ϕ satisfying Axioms* 1–3.
(2) *Any ϕ that satisfies Axioms* 1–3 *can be recursively computed as follows. Let $\mu_1 \in \mathcal{S}$ be such that $\phi(\mu_1) = \mu_1$. Then, construct the sequence of states $\left\{\mu_1, ..., \mu_{|\mathcal{S}|}\right\}$ with the property that if for any $l \in (j, |\mathcal{S}|]$, $\mu_l \not\succ_{\mu_j} \mu_j$. Let*

$$\mathcal{M}_k = \left\{s \in \left\{\mu_1, \ldots, \mu_{k-1}\right\} : s \succ_{\mu_k} \mu_k \text{ and } \phi(s) = s\right\}.$$

Then, for each $k = 2, ..., |\mathcal{S}|$,

$$\phi(\mu_k) = \begin{cases} \mu_k & \text{if } \mathcal{M}_k = \varnothing \\ s \in \mathcal{M}_k\colon \nexists z \in \mathcal{M}_k \text{ with } z \succ_{\mu_k} s & \text{if } \mathcal{M}_k \neq \varnothing \end{cases}.$$

(If there exist more than one $s \in \mathcal{M}_k$: $\nexists z \in \mathcal{M}_k$ with $z \succ_{\mu_k} s$, we pick any of these; this corresponds to multiple ϕ functions).
(3) *For any two mappings ϕ_1 and ϕ_2 that satisfy Axioms* 1–3 *the stable states of these mappings coincide.*
(4) *If, in addition, Assumption 22.3 holds, then the mapping that satisfies Axioms* 1–3 *is "payoff-unique" in the sense that for any two mappings ϕ_1 and ϕ_2 that satisfy Axioms* 1–3 *and for any $s \in \mathcal{S}$, $\phi_1(s) \sim \phi_2(s)$.*

Proof. **(Part 1)** To prove existence, we first construct the sequence of states $\left\{\mu_1, ..., \mu_{|\mathcal{S}|}\right\}$ such that

$$\text{if } 1 \le j < l \le |\mathcal{S}|, \text{ then } \mu_l \not\succ_{\mu_j} \mu_j, \tag{22.14}$$

The construction is by induction. Suppose we have defined μ_j for all $j \le k-1$, where $k \le |\mathcal{S}|$. Then applying Assumption 22.2(a) to the collection of states $\mathcal{S} \setminus \left\{\mu_1, \ldots, \mu_{k-1}\right\}$, we conclude that there exists μ_k satisfying (22.14). By construction, μ is a bijection that satisfies (22.14).

The second step is to construct a candidate mapping $\phi : \mathcal{S} \to \mathcal{S}$. This is again by induction. For $k = 1$, let $\phi(\mu_k) = \mu_k$. Suppose we have defined μ_j for all $j \le k-1$ where

$k \leq |\mathcal{S}|$. Define the collection of states

$$\mathcal{M}_k = \left\{ s \in \left\{ \mu_1, \ldots, \mu_{k-1} \right\} : s \succ_{\mu_k} \mu_k \text{ and } \phi(s) = s \right\}. \tag{22.15}$$

$\mathcal{M}_k$ is the subset of states where ϕ has already been defined and satisfies $\phi(s) = s$ and which are preferred to μ_k within μ_k. If $\mathcal{M}_k$ is empty, then we define $\phi(\mu_k) = \mu_k$. If $\mathcal{M}_k$ is non-empty, then take $\phi(\mu_k)$ such that

$$s \not\succ_{\mu_k} \phi(\mu_k) \text{ for any } s \in \mathcal{M}_k \tag{22.16}$$

(such state $\phi(\mu_k)$ exists because we can apply Assumption 22.2(b) to $\mathcal{M}_k$). Proceeding likewise for all $2 \leq k \leq |\mathcal{S}|$ we construct mapping ϕ.

To complete the proof, we need to verify that mapping ϕ satisfies Axioms 1–3. This is straightforward for Axioms 5 and 7. In particular, by construction, either $\phi(\mu_k) = \mu_k$ (in that case these axioms trivially hold), or $\phi(\mu_k)$ is an element of $\mathcal{M}_k$; in that case $\phi(\mu_k) \succ_{\mu_k} \mu_k$ and $\phi(\phi(\mu_k)) = \phi(\mu_k)$ by (22.15). To check Axiom 6, suppose that for some state μ_k there exists z such that $z \succ_{\mu_k} \mu_k$, $z = \phi(z)$, and $z \succ_{\mu_k} \phi(\mu_k)$. Then $z \succ_{\mu_k} \mu_k$, combined with condition (22.14), implies that $z \in \left\{ \mu_1, \ldots, \mu_{k-1} \right\}$. But the last condition, $z \succ_{\mu_k} \phi(\mu_k)$, now contradicts (22.16). This means that such z does not exist, and therefore Axiom 6 is satisfied.

(Part 2) Suppose that ϕ_1 and ϕ_2 are two mappings satisfying Axioms 1–3. Consider the sequence of states $\{\mu_k\}_{k=1}^{|\mathcal{S}|}$. If the sets of stable states of ϕ_1 and ϕ_2 do not coincide, there exists some k such that $\phi_1(\mu_j) = \mu_j \iff \phi_2(\mu_j) = \mu_j$ for $j < k$, but either $\phi_1(\mu_k) = \mu_k$ and $\phi_2(\mu_k) \neq \mu_k$ or $\phi_1(\mu_k) \neq \mu_k$ and $\phi_2(\mu_k) = \mu_k$. Without loss of generality assume the former is the case (the argument for the latter case is identical). By Axiom 5, $\phi_2(\mu_k) \succ_{\mu_k} \mu_k$, so by (22.14) $\phi_2(\mu_k) = \mu_l$ for some $l < k$. Therefore, we have $\mu_l \succ_{\mu_k} \mu_k$ and because $l < k$ and $\phi_1(\mu_l) = \phi_2(\mu_l) = \mu_l$, $\phi_1(\phi_1(\mu_l)) = \phi_1(\phi_2(\mu_l)) = \phi_1(\mu_l)$. But then Axiom 6 implies that $\phi_1(\mu_k)$ cannot equal y which satisfies $\mu_l \succ_{\mu_k} y$, in particular, $\phi_1(\mu_k) \neq \mu_k$, yielding a contradiction to the hypothesis that $\phi_1(\mu_k) = \mu_k$ and $\phi_2(\mu_k) \neq \mu_k$. Consequently, $\phi_1(s) = s$ if and only if $\phi_2(s) = s$.

(Part 3) Suppose Assumption 22.3 holds. Suppose, to obtain a contradiction, that ϕ_1 and ϕ_2 are two distinct mappings that satisfy Axioms 1–3. Then there exists some state s such that $\phi_1(s) \nsim \phi_2(s)$. Part 2 of this Theorem implies that $\phi_1(s) = s$ if and only if $\phi_2(s) = s$; since $\phi_1(s) \nsim \phi_2(s)$, we obtain that $\phi_1(\mu_{ks}) \neq s \neq \phi_2(s)$. Now Axiom 5 implies $\phi_1(s) \succ_s s$, $\phi_2(s) \succ_s s$, and Assumption 22.3 implies that either $\phi_1(s) \succ_s \phi_2(s)$ or $\phi_2(s) \succ_s \phi_1(s)$; without loss of generality assume the former. Then for $y = \phi_2(s)$ there exists $\phi_1(s)$ such that $\phi_1(s) \succ_s y$, $\phi_1(s) \succ_s s$, and $\phi_2(\phi_1(s)) = \phi_1(s)$ (the latter because $\phi_1(s)$ is a ϕ_1-stable state by Axiom 7, and by Part 2 of this Theorem, it is also a ϕ_2-stable state). But then we can apply Axiom 6 to mapping ϕ_2 and derive the conclusion that $\phi_2(s)$ cannot equal y. This contradiction completes the proof. □

Theorem 22.1 shows that a mapping that satisfies Axioms 1–3 necessarily exists and provides a sufficient condition for uniqueness. Even when the uniqueness condition, Assumption 22.3, does not hold, we know that payoff-relevant dynamically stable states coincide for any two mappings ϕ_1 and ϕ_2 that satisfy Axioms 1–3.

Theorem 22.1 also provides a simple recursive characterization of the mapping ϕ. Intuitively, Assumption 22.2(a) ensures that there exists some state $\mu_1 \in \mathcal{S}$, such that there does not exist another state $s \in \mathcal{S}$ with $s \succ_{\mu_1} \mu_1$. Taking μ_1 as base, we recursively construct the set of states $\mathcal{M}_k \subset \mathcal{C}$, $k = 1, ..., |\mathcal{S}|$, that includes dynamically stable states that are preferred to state μ_k (that is, $\phi(s) = s$ and $s \succ_{\mu_k} \mu_k$). When the set $\mathcal{M}_k$ is empty, then no dynamically stable state is part of a winning coalition starting in state μ_k, and therefore we must have $\phi(\mu_k) = \mu_k$. When this set is nonempty, then we can pick a dynamically stable state that will arise starting from state μ_k. In addition to its recursive (and thus easy-to-construct) nature, this characterization is useful because it highlights a fundamental property of dynamically stable states: *a state μ_k is made dynamically stable precisely by the absence winning coalitions in μ_k favoring a transition to another dynamically stable state.* We will see that this insight plays an important role in the applications below.

We have motivated the analysis leading up to Theorem 22.1 with the argument that, when agents are sufficiently forward-looking, only dynamically stable states should be observed (at least in the "long run"). The analysis of the extensive-form game presented next will substantiate this interpretation further.

22.5.3. Noncooperative Foundations of Dynamically Stable States. We now describe an extensive-form game meant to capture the basic economic interactions emphasized so far in the simplest possible way. The main result here will establish the equivalence between the Markov Perfect Equilibria (MPE) of this extensive-form game and the axiomatic characterization of Theorem 22.1.

The essential features of the extensive-form game are a protocol for a sequence of agenda-setters and proposals at each date, and a protocol for voting over proposals. We do this by introducing a natural number K_s and a mapping

$$\pi_s : \{1, \ldots, K_s\} \to \mathcal{I} \cup \mathcal{S}$$

for each state $s \in \mathcal{S}$. This mapping specifies a finite sequence of elements from $\mathcal{I} \cup \mathcal{S}$, where K_s ($\leq |\mathcal{I}| + |\mathcal{S}|$) is the length of sequence for state s and determines the sequence of agenda-setters and proposals. In particular, if $\pi_s(\tau) \in \mathcal{I}$, then it denotes an agenda-setter who will make a proposal from the set of states $\mathcal{S}$. Alternatively, if $\pi_s(\tau) \in \mathcal{S}$, then it directly corresponds to an exogenously-specified proposal over which individuals vote. Therefore, the extensive-form game is general enough to include both proposals for a change to a new state initiated by agenda-setters (a subset of the players that may depend on the state) or those that are exogenously placed on the table (as is the case in standard voting models where alternatives are voted over in pairwise contests). We impose the following mild requirements on $\pi_s(\cdot)$:

Assumption 22.4. ***(Agenda-Setting and Proposals)*** *For every state $s \in \mathcal{S}$, $\pi_s(K_s) = s \in \mathcal{S}$, and one (or both) of the following two conditions is satisfied:*

(a) For any state $q \in \mathcal{S}$, there is an element $k : 1 \leq k \leq K_s$ of sequence π_s such that $\pi_s(k) = q$.

(b) For any player $i \in \mathcal{I}$ there is an element $k : 1 \leq k \leq K_s$ of sequence π_s such that $\pi_s(k) = i$.

This assumption implies that either sequence π_s contains all possible states or it allows all possible agenda-setters to eventually make a proposal. In either case, state s itself is the last element of this sequence.

The game starts at time $t = 1$ with agenda-setting period and with state s_0 exogenously given. Each period begins with state $s = s_{t-1}$ and current alternative is $A_0 = s$. After that, for each $k \in \{1, \ldots, K_s\}$, the following procedure takes place. First, proposal $P_{k,t} \in \mathcal{S}$ is determined: if $\pi_s(k) \in \mathcal{S}$, then $P_{k,t} = \pi_s(k)$ automatically; if $\pi_s(k) \in \mathcal{I}$, then $P_{k,t}$ is proposed by agenda-setter $\pi_s(k)$. After this stage, proposal $P_{k,t}$ is voted against current alternative $A_{k-1,t}$; voting is sequential according to some predetermined sequence (we will show that this sequence does not matter for our results) and each player can say either *yes* or *no*. Let $Y_{k,t}$ be the set of those who vote *yes* for proposal $P_{k,t}$ (at time t and stage t). The new alternative $A_{k,t}$ is chosen to be $P_{k,t}$ in the following two cases: (a) $P_{k,t} \neq s$ and $Y_{k,t} \in \mathcal{W}_s$ or (b) $P_{k,t} = s$ and $Y_{k,t} \in \mathcal{B}_s$; otherwise, $A_{k,t} = A_{k-1,t}$. Intuitively, it takes a winning coalition to move to a new proposal, but a blocking coalition is sufficient to reintroduce the status-quo s. When all k votings are over, the final alternative $A_{K_s,t}$ is well-defined, and the transition to state $s_t = A_{K_s,t}$ takes place. We also assume that state $s_t = s_{t-1}$, then voting process comes to an end and players receive $w_s(i)$ in each period thereafter. This assumption does not imply that the game is finite, since $s_t = s_{t-1}$ may never happen. Nevertheless, it is important for our analysis, since, together with the assumption on "transaction costs" that players incur when $s_t \neq s_{t-1}$ (see below), it enables us to prove the existence of a pure strategy Markov Perfect Equilibrium. Although the existence of pure strategy equilibria can be guaranteed using other assumptions on the extensive-form game, none of these are as simple, or as natural, as the one we are using here.[1]

Payoffs in this dynamic game are given by (22.13), with

$$u_t(i) = \begin{cases} w_s(i) & \text{if } s_t = s_{t-1} = s \\ 0 & \text{if } s_t \neq s_{t-1} \end{cases}$$

for each $i \in \mathcal{I}$. In other words, in the first period after a transition (that is, if the current state is not the same as the state in the previous period), each individual receives zero payoff. In all other periods, each individual receives the payoff as specified in Assumption 22.2. The period of zero payoff can be interpreted as representing a "transaction cost" associated with the change in the state. Since the game is infinitely-repeated and β is high, this one period of zero payoff should not be of major consequence. In particular, once (and if) a dynamically stable state is reached), individuals will receive $w_s(i)$ at each date thereafter.

We next define a Markov Perfect Equilibrium (MPE). For this purpose, consider a general n-person infinite-stage game, where each individual can take an action at every stage. Let the action profile of each individual be

$$a^i = \left(a^i_1, a^i_2, \ldots\right) \text{ for } i = 1, \ldots, n,$$

with $a^i_t \in A^i_t$ and $a^i \in A^i = \prod_{t=1}^{\infty} A^i_t$. Let $h_t = (a_1, \ldots, a_t)$ be the history of play up to stage t (not including stage t), where $a_s = \left(a^1_s, \ldots, a^n_s\right)$, so h_0 is the history at the beginning of the game, and let H_t be the set of histories h_t for $t : 0 \leq t \leq T-1$. We denote the set of all potential histories up to date t by $H^t = \bigcup_{s=0}^{t} H_s$. Let t-continuation action profiles be

$$a^{i,t} = \left(a^i_t, a^i_{t+1}, \ldots\right) \text{ for } i = 1, \ldots, n,$$

with the set of continuation action profiles for player i denoted by $A^{i,t}$. Symmetrically, define t-truncated action profiles as

$$a^{i,-t} = \left(a^i_1, a^i_2, \ldots, a^i_{t-1}\right) \text{ for } i = 1, \ldots, n,$$

[1] In future versions we may switch to one of these alternative excessive-forms. One possibility is to assume that voting continues at each date (so that $s_t = s_{t-1}$ does not cause the game to end), but allow individuals to abstain in voting (and also assume simultaneous voting). In this case, we can have that for a proposal $P_{k,t} = s_t$ starting with state s_{t-1} at time t, $Y_{k,t} \notin \mathcal{W}_{s_{t-1}}$ and $N_{k,t} \notin \mathcal{B}_{s_{t-1}}$. If this is the case, all players receive payoff $w_i(s_{t-1})$, and next period $P_{t+1} = s_t$ again (i.e., the incomplete voting from the previous round continues) and only those who have not cast their votes already vote at this point. Using this alternative game, we can establish the equivalents of Theorems 22.2 and 22.4 below.

with the set of t-truncated action profiles for player i denoted by $A^{i,-t}$. We also use the standard notation a^i and a^{-i} to denote the action profiles for player i and the action profiles of all other players (similarly, A^i and A^{-i}). The payoff functions for the players depend only on actions, i.e., player i's payoff is given by

$$u^i\left(a^1, \ldots, a^n\right).$$

A (possibly mixed) strategy for player i is

$$\sigma^i : H^\infty \to \Delta\left(A^i\right),$$

where $\Delta(X)$ denotes the set of probability distributions defined over the set X, and for any $h \in H^\infty$ actions in the support of $\sigma^i(h)$ are feasible. A t-continuation strategy for player i (corresponding to strategy σ^i) specifies plays only after time t (including time t), i.e.,

$$\sigma^{i,t} : H^\infty \setminus H^{t-2} \to \Delta\left(A^{i,t}\right),$$

where $H^\infty \setminus H^{t-2}$ is the set of histories starting at time t.

A continuation strategy $\sigma^{i,t}$ is *Markovian* if

$$\sigma^{i,t}\left(h^{t-1}\right) = \sigma^{i,t}\left(\tilde{h}^{\tau-1}\right)$$

for all $\tau \geq t$, $h^{t-1}, \tilde{h}^{\tau-1} \in H^\infty$ such that for any $a^{i,t}, \tilde{a}^{i,\tau} \in A^{i,t}$ and any $a^{-i,t} \in A^{-i,t}$ we have

$$u^i\left(a^{i,t}, a^{-i,t} \mid h^{t-1}\right) \geq u^i\left(\tilde{a}^{i,\tau}, a^{-i,t} \mid h^{\tau-1}\right)$$

implies that

$$u^i\left(a^{i,t}, a^{-i,t} \mid \tilde{h}^{t-1}\right) \geq u^i\left(\tilde{a}^{i,\tau}, a^{-i,t} \mid \tilde{h}^{\tau-1}\right).$$

We say that a strategy profile $\left(\hat{\sigma}^1, ..., \hat{\sigma}^n\right)$ is *Markov Perfect Equilibrium (MPE)* if each strategy $\hat{\sigma}^1$ is Markovian and

$$u^i\left(\hat{\sigma}^i, \hat{\sigma}^{-i}\right) \geq u^i\left(\sigma^i, \hat{\sigma}^{-i}\right) \text{ for all } \sigma^i \in \Sigma^i \text{ and for all } i = 1, \ldots, n.$$

A pure strategy MPE is defined as a MPE in which only pure strategies are used.

The main result is given in the following theorem and establishes a close correspondence between the MPEs of the game described here and the outcomes picked by mapping ϕ described in Theorem 22.1. The proof is somewhat long and is omitted, but the idea is simple, especially in light of the axiomatic characterization above.

Theorem 22.2. ***(Noncooperative Foundations of Dynamically Stable States)*** *Suppose Assumptions 22.1 and 22.2 hold, β is sufficiently close to 1, and the initial state is $s_0 \in \mathcal{S}$. Then:*

(1) *For any mapping ϕ satisfying Axioms* 1–3 *there is a set of sequences $\{\pi_s(\cdot)\}_{s\in\mathcal{S}}$ and a MPE σ of the game such that $s_t = \phi(s_0)$ for any $t \geq 1$. In other words, the game reaches $\phi(s_0)$ in a finite number of steps (after one period) and stays in this state thereafter.*

(2) *Conversely, for any set of sequences $\{\pi_s(\cdot)\}_{s\in\mathcal{S}}$, any MPE in pure strategies σ has the property that it reaches a certain state, s^∞, in a finite number of periods (with one transition): for $t \geq 1$, $s_t = s^\infty$. Moreover, there exists mapping $\phi : \mathcal{S} \to \mathcal{S}$ that satisfies Axioms* 1–3 *such that $s^\infty = \phi(s_0)$.*

(3) *If, in addition, Assumption 22.3 holds, then the MPE is essentially unique in the sense that for any set of sequences $\{\pi_s(\cdot)\}_{s\in\mathcal{S}}$, any MPE strategy profile σ induces $s_t \sim \phi(s_0)$ for all $t \geq 1$, where ϕ satisfies Axioms* 1–3.

22.5.4. Limited State Transitions. We have so far assumed that any transition (from any state into any other state) is possible. In many interesting applications, there will be certain transitions that are not possible. However, in the model of voting over coalitions, coalition formation, and nondemocracies (Acemoglu, Egorov, and Sonin, 2007) discussed above, only current members of the ruling coalition can be part of future ruling coalitions and thus transitions to states that include individuals previously eliminated are ruled out. let us know reformulate the Assumptions and the Axioms above to incorporate the feature that only certain state transitions are allowed and generalize the results in Theorems 22.1 and 22.2.

The key to the analysis in this part is the binary relation $\rightsquigarrow$ on the set of states $\mathcal{S}$. For any $x, y \in \mathcal{S}$, we write $x \rightsquigarrow y$ to denote that a transition from x to y is possible and $x \rightsquigarrow \mathcal{Q}$ for some $\mathcal{Q} \subset \mathcal{S}$ to denote that the transition to any state z in $\mathcal{Q}$ is possible (provided that these positions are supported by a winning coalition in x). Our analysis so far thus corresponds to the special case where $x \rightsquigarrow \mathcal{S}$ for any $x \in \mathcal{S}$. We adopt the following natural assumption on the transition relation.

Assumption 22.5. ***(Feasible Transitions)*** *Relation $\rightsquigarrow$ satisfies the following properties:*

(a) (reflexivity) $\forall x \in \mathcal{S} : x \rightsquigarrow x$;

(a) (transitivity) $\forall x, y, z \in \mathcal{S} : x \rightsquigarrow y$ and $y \rightsquigarrow z$ imply $x \rightsquigarrow z$.

Part (b) Assumption 22.5 requires that if some indirect transition from x to z is feasible, so is a direct transition between the states. Without requiring transitivity, there would be additional technical details to take care of, because, for instance, if transition from x to z

is possible through y only, then it is only possible if both a winning coalition in x prefers z to x and a winning coalition in y prefers z to y. Nevertheless, this assumption can be dispensed with, and we could assume instead that whenever $x \rightsquigarrow y$ and $y \rightsquigarrow z$ but $x \not\rightsquigarrow z$, then $\mathcal{W}_x = \mathcal{W}_y$ (or a weaker version of this assumption).

We next consider slightly weaker versions of Assumption 22.2 and Assumption 22.3, incorporating the fact that only certain transitions are feasible (since when some transitions are not feasible, it becomes easier to rule out cycles).

Assumption $2'$**:** ***(Payoffs with Limited Transitions)*** *Payoff functions* $\{w_i(\cdot)\}_{i\in\mathcal{I}}$ *satisfy the following properties:*

(a) For any nonempty $\mathcal{Q} \subset \mathcal{S}$ *such that* $x \rightsquigarrow y$ *for any* $x, y \in \mathcal{S}$*, there exists state* $z \in \mathcal{S}$ *such that for any* $x \in \mathcal{S}$*, we have* $x \not\succ_z z$*;*

(b) For any $s \in \mathcal{S}$*, let* $\mathcal{S} = \{y \in \mathcal{S} : s \rightsquigarrow y \text{ and } y \succ_s s\}$*. If* $\mathcal{S}$ *is non-empty, then there exists* $z \in \mathcal{S}$ *such that for any* $x \in \mathcal{S}$ *we have* $x \not\succ_s z$*.*

Assumption $3'$**:** ***(Comparability with Limited Transitions)*** *For* $x, y, z \in \mathcal{S}$ *such that* $z \rightsquigarrow x$*,* $z \rightsquigarrow y$*,* $x \succ_z z$*,* $y \succ_z z$*, and* $x \nsim y$*, either* $y \succ_z x$ *or* $x \succ_z y$*.*

Finally, let us reformulate Axioms 1–3 for this slightly modified set up (note that Axiom 3 is unchanged, though we state it again for completeness).

Axiom $1'$**:** ***(Desirability)*** *If* $x, y \in S$ *are such that* $y = \phi(x)$*, then either* $y = x$ *or* $x \rightsquigarrow y$ *and* $y \succ_x x$*.*

Axiom $2'$**:** ***(Rationality)*** *If* $x, y, z \in S$ *are such that* $x \rightsquigarrow z$*,* $z \succ_x x$*,* $z = \phi(z)$*, and* $z \succ_x y$*, then* $y \neq \phi(x)$*.*

Axiom $3'$**:** ***(Stability)*** *If* $x, y \in S$ *are such that* $y = \phi(x)$*, then* $y = \phi(y)$*.*

With this new set of Axioms, a slightly modified version of Theorem 22.1 holds:

Theorem 22.3. ***(Dynamically Stable States with Limited Transitions)*** *Suppose that binary relation* $\rightsquigarrow$ *satisfies Assumption 22.5, and that Assumptions 22.1 and* $2'$ *hold. Then:*

(1) *There exists mapping* ϕ *satisfying Axioms* $1'$–$3'$*.*

(2) *Any mapping* ϕ *that satisfies Axioms* $1'$–$3'$ *this characterize recursively as follows. Construct sequence* $\left\{\mu_1, ..., \mu_{|\mathcal{S}|}\right\}$ *with the property that if for any* $l \in (j, |\mathcal{S}|]$*, either* $\mu_j \not\rightsquigarrow \mu_l$ *or* $\mu_l \not\succ_{\mu_j} \mu_j$*. Let* $\phi(\mu_1) = \mu_1$*. For any* $k : 2 \leq k \leq |\mathcal{S}|$*, define*

$$\mathcal{M}_k = \left\{s \in \left\{\mu_1, \ldots, \mu_{k-1}\right\} : \mu_k \rightsquigarrow s,\ s \succ_{\mu_k} \mu_k, \text{ and } \phi(s) = s\right\}$$

and let

$$\phi(\mu_k) = \begin{cases} \mu_k & \text{if } \mathcal{M}_k = \varnothing \\ s \in \mathcal{M}_k : \nexists z \in \mathcal{M}_k \text{ with } \mu_k \rightsquigarrow z \text{ and } z \succ_{\mu_k} s & \text{if } \mathcal{M}_k \neq \varnothing \end{cases}.$$

(3) *For any two mappings ϕ_1 and ϕ_2 that satisfy Axioms $1'$–$3'$ the stable states of these mappings coincide.*

(4) *If, in addition, Assumption $3'$ holds, then the mapping that satisfies Axioms $1'$–$3'$ is "payoff-unique" in the sense that for any two mappings ϕ_1 and ϕ_2 that satisfy Axioms $1'$–$3'$ and for any $s \in \mathcal{S}$, $\phi_1(s) \sim \phi_2(s)$.*

Proof. The proof is an extension of that of Theorem 22.1. The idea of the proof is to construct a mapping (sequence) $\mu : \{1, \ldots, |\mathcal{S}|\} \leftrightarrow \mathcal{S}$ such that for any $1 \leq k < |\mathcal{S}|$ we have that

$$\text{if } 1 \leq j < l \leq |\mathcal{S}|, \text{ then } \mu_j \not\rightsquigarrow \mu_l \text{ or } \mu_l \not\succ_{\mu_j} \mu_j. \tag{22.17}$$

To construct mapping μ, for each $x \in \mathcal{S}$ we consider its equivalence class $\mathcal{E}_x$ defined by

$$\mathcal{E}_x = \{y \in \mathcal{S} : x \rightsquigarrow y \text{ and } y \rightsquigarrow x\}.$$

Assumption 22.5 guarantees that $\{\mathcal{E}_x \mid x \in \mathcal{S}\}$ indeed defines an equivalence relation with different classes either coinciding or not intersecting. The binary relation $\rightsquigarrow$ on elements of $\mathcal{S}$ induces relation $\rightsquigarrow$ in equivalence classes by letting $\mathcal{E}_x \rightsquigarrow \mathcal{E}_y$ if and only if $x \rightsquigarrow y$; note that this relation is well-defined in the sense that it does not depend on the elements x and y picked from $\mathcal{E}_x$ and $\mathcal{E}_y$, respectively. Furthermore, this relation is acyclical in the sense that there do not exist distinct classes $\mathcal{E}_1, \ldots, \mathcal{E}_l$ such that $\mathcal{E}^j \rightsquigarrow \mathcal{E}^{j+1}$ for $1 \leq j < l$ and $\mathcal{E}^l \rightsquigarrow \mathcal{E}^1$. Consequently, we can form a sequence of all equivalence classes $\mathcal{E}^1, \ldots, \mathcal{E}^m$ (where m is the number of classes) such that $\mathcal{E}^j \not\rightsquigarrow \mathcal{E}^k$ for any $1 \leq j < k \leq m$. Now, within each class $\mathcal{E}^k$, we enumerate its elements as $\mu_1^k, \ldots, \mu_{|\mathcal{E}^k|}^k$ so that $\mu_l^k \not\succ_{\mu_j^k} \mu_j^k$ for $1 \leq j < l \leq |\mathcal{E}^k|$ (this is feasible due to Assumption $2'$(a)). Next, construct the sequence μ as follows: we give members of class $\mathcal{E}_1$ numbers 1 to $|\mathcal{E}_1|$ in the order they are listed in the sequence $\mu^1 \equiv \left(\mu_1^1, \ldots, \mu_{|\mathcal{E}^1|}^1\right)$, then we take members of class $\mathcal{E}_2$ as they are listed in the sequence μ^2, and so on. It is easy to show that the sequence μ constructed in this way satisfies (22.17). The rest of the proof closely follows the one of Theorem 22.1 and is omitted. □

Similarly, an equivalent of Theorem 22.2 again applies.

Theorem 22.4. ***(Noncooperative Foundations of Dynamically Stable States with Limited Transitions)*** *Suppose that binary relation $\rightsquigarrow$ satisfies Assumption 22.5, that*

Assumptions 22.1 and $2'$ *hold, that* β *is sufficiently close to* 1*, and that the initial state is* $s_0 \in \mathcal{S}$*. Then:*

(1) *For any mapping* ϕ *satisfying Axioms* $1'$–$3'$ *there is a set of sequences* $\{\pi_s(\cdot)\}_{s\in\mathcal{S}}$ *and a MPE* σ *of the game such that* $s_t = \phi(s_0)$ *for any* $t \geq 1$*. In other words, the game reaches* $\phi(s_0)$ *in a finite number of steps (after one period) and stays in this state thereafter.*

(2) *Conversely, for any set of sequences* $\{\pi_s(\cdot)\}_{s\in\mathcal{S}}$*, any MPE in pure strategies* σ *has the property that it reaches a certain state,* s_∞*, in a finite number of periods (with one transition): for* $t \geq 1$*,* $s_t = s^\infty$*. Moreover, there exists mapping* $\phi : \mathcal{S} \to \mathcal{S}$ *that satisfies Axioms* $1'$–$3'$ *such that* $s^\infty = \phi(s_0)$*.*

(3) *If, in addition, Assumption 22.5.4 holds, then the MPE is essentially unique in the sense that for any set of sequences* $\{\pi_s(\cdot)\}_{s\in\mathcal{S}}$*, any MPE strategy profile* σ *induces* $s_t \sim \phi(s_0)$ *for all* $t \geq 1$*, where* ϕ *satisfies Axioms* $1'$–$3'$*.*

These theorems therefore show that the essential results of Theorems 22.1 and 22.2 generalize to an environment with limited transitions. The intuition for these results and the recursive characterization of dynamically stable states are essentially identical to those in Theorems 22.1 and 22.2.

We now revisit the examples discussed in the Introduction, as well as a number of new examples, and show how the theory developed above can be applied in these cases to derive predictions about dynamically stable states. In some of the applications, we will allow for $w_i(s) = 0$ for some i and s. This is to simplify notation, and setting $w_i(s) = \varepsilon$ for $\varepsilon > 0$ and small would not change any of the results or interpretations.

22.5.5. Inefficient Inertia and Lack of Reform. Consider a society consisting of N individuals and a set of finite states $\mathcal{S}$. We start with $s_0 = a$ corresponding to absolutist monarchy, where individual E holds power. More formally, $\mathcal{W}_a = \{E\}$. Suppose that for all $x \in \mathcal{S}\setminus\{a\}$, we have that $\{E\} \notin \mathcal{B}_x$, that is, E is not a block in coalition (and thus not a winning coalition). Moreover, there exists a state, "democracy," $d \in \mathcal{S}$ such that $\phi(x) = d$ for all $x \in \mathcal{S}\setminus\{a\}$. The other words, starting with any regime other that absolutist monarchy, we will eventually end up with democracy. Suppose also that there exists $y \in \mathcal{S}$ such that

$$w_y(i) > w_a(i),$$

meaning that all individuals are better off in state y than in absolutist monarchy, a. In fact, the gap between the payoffs in state y and those in a could be arbitrarily large. Is straightforward to verify that Assumptions 22.1–22.3 are satisfied in this game.

To understand economic interactions in the most straightforward manner, consider the extensive-form game. Inspection of this extensive-form game makes it clear that for β sufficiently large, E will not accept any reforms away from a. In particular, since, given our specification, the game will reach state d in a finite number of periods, for any gap between $\max_{x\in\mathcal{S}}\{w_x(E)\}$ and $w_a(E)$, there exists a sufficiently large β that E is better off to remain in state a.

This example illustrates the potential (and potentially large) inefficiencies that can arise in games of dynamic collective decision-making and emphasizes that commitment problems are at the heart of these inefficiencies. If the society could collectively commit to stay in some state y, then these inefficiencies could be partially avoided. And yet such a commitment is not possible, since once state y is reached, E is no longer a blocking coalition and the rest of the society wishes to progress towards d.

22.5.6. Middle Class and Democratization. Let us consider a variation on this game. Suppose again that the initial state is $s_0 = a$, , where $\mathcal{W}_a = \{E\}$. To start with, suppose that there is only one other agent, P, representing the poor, and two other states, $d1$, democracy with limited redistribution, and $d2$, democracy with extensive redistribution. Suppose that $\mathcal{W}_{d1} = \mathcal{W}_{d2} = \{P\}$. As before,

$$w_{d2}(E) < w_a(E) < w_{d1}(E),$$

and

$$w_a(P) < w_{d1}(P) < w_{d2}(P),$$

so that P prefers extensive redistribution. Given the fact that $\mathcal{W}_{d1} = \mathcal{W}_{d2} = \{P\}$, once democracy is established, the poor can implement extensive a distribution. Anticipating this, E will resist democratization.

Now imagine that an additional social group emerges, M, representing the middle class, and the middle class is sufficiently numerous so that

$$\mathcal{W}_{d1} = \mathcal{W}_{d2} = \{M, P\}.$$

Their preferences are also opposed to extensive redistribution, so

$$w_a(M) < w_{d2}(M) < w_{d1}(M).$$

This implies that once state $d1$ emerges, there no longer exists a winning coalition to force extensive redistribution. Now anticipating this, E will be happy to establish democracy (extend the franchise). Therefore, this example illustrates how the presence of an additional player can have moderating effect (see Acemoglu and Robinson, 2006a, for examples in which the middle class may have played such a role in the process of democratization).

22.5.7. Voting in Clubs. Let us now return to the dynamic game of voting in clubs considered previously. The society consists of N individuals, so that $\mathcal{I} = \{1, \ldots, N\}$. Following Roberts (1999), suppose that there are N states, of the form $s_k = \{1, \ldots, k\}$, $1 \leq k \leq N$. Roberts (1999) imposes the following strict single crossing property:

$$\text{for all } l > k \text{ and } j > i, \;\; w_{s_l}(j) - w_{s_k}(j) > w_{s_l}(i) - w_{s_k}(i).$$

He then considers two voting schemes: majority voting within a club (where in club s_k one needs more than $k/2$ votes for a change in club size to be implemented) or median voter voting (where the agreement of individual $(k+1)/2$ if k is odd or $k/2$ and $k/2+1$ if k is even are needed). Roberts proves that under either rule there are no cycles, and the same set of stable clubs emerges.

To show how Roberts's model is a special case of the analysis here, let us adopt the following simplifying assumption

$$\text{for any } i \in \mathcal{I} \text{ and } k \neq l, \; w_{s_k}(i) \neq w_{s_l}(i). \tag{22.18}$$

Though not necessary, this assumption simplifies the analysis, in particular, avoiding certain complications that arise when k is even. Majority and median voting rules imply the following structure of winning coalitions,

$$\mathcal{W}_{s_k}^{maj} = \{\mathcal{C} : |\mathcal{C} \cap s_k| > k/2\}$$

and

$$\mathcal{W}_{s_k}^{med} = \begin{cases} \{\mathcal{C} : (k+1)/2 \in \mathcal{C}\} & \text{if } k \text{ is odd;} \\ \{\mathcal{C} : \{k/2, k/2+1\} \subset \mathcal{C}\} & \text{if } k \text{ is even.} \end{cases}$$

In addition, let us also refer to a *Modified Roberts model*, in which only odd-sized clubs are allowed. It is straightforward to verify that Roberts's original proof of existence of equilibria apply without any change to this modified model.

The next proposition establishes that the analysis of this paper is applicable to Roberts's model, and also establishes the link between the mapping ϕ that satisfies Axioms 1–3 and Roberts's *Markov Voting Equilibrium* when attention is restricted to odd-sized clubs. We also provide a definition of Markov Voting Equilibrium for completeness.

Definition 22.2. ***(Markov Voting Equilibrium)*** *A transition rule* $y^*(\cdot)$ *is a mapping that corresponds a probability distribution (a lottery over the next states) to each state* s_k*. Define continuation value of individual* i *if the current state is* s_k *and transition rule is* $y(\cdot)$ *by* $V_i(s_k, y(\cdot))$*. For each state* s_k *consider set* $Y(s_k)$ *such that for any* $y \in Y(s_k)$ *and any state* z*, the number of players among players* $1, \ldots, k$ *such that* $V_i(y, y^*(\cdot)) > V_i(z, y^*(\cdot))$ *is at least as large as the number of those with* $V_i(y, y^*(\cdot)) < V_i(z, y^*(\cdot))$*. Transition rule* $y^*(\cdot)$ *is a* Markov Voting Equilibrium *if the support of* $y^*(s_k)$ *is a subset of* $Y^*(s_k)$ *for any state* s_k*.*

Note that according to this definition, a majority that strictly prefers a transition is not necessary for a transition to take place, which contrasts with the approach taken in our paper. Nevertheless, the following proposition shows that Roberts's (1999) model and results follow directly from the general framework developed in this paper.

Proposition 22.11.

1. *Assumptions 22.1 and 22.2 are satisfied in original Roberts model and in the Modified Roberts model.*
2. *In the Modified Roberts model, Assumption 22.3 is satisfied and there exists a unique mapping* ϕ *that satisfies Axioms* 1–3.
3. *Suppose that* β *is close to 1 and that a pure strategy Markov Voting Equilibrium exists in the Modified Roberts model. Then a steady state reached starting with club* s *in this equilibrium coincides with the dynamically stable state* $\phi(s)$*.*

22.5.8. Gradual Franchise Extension. Let us now use a variant of Roberts's model, together with our extension with limited transitions studied above, to analyze gradual franchise extension. The society consists of N individuals (or social groups), and there are N states of the form $s_k = \{1, \ldots, k\}$, $1 \leq k \leq N$ as in the previous subsection. The transition rules are such that state s_k can only transition to s_{k-1} or to s_{k+1}. This implies that the transitivity the part, part (b), of Assumption 22.5 needs to be relaxed. It can be verified that in most cases (including the one described here) the results hold without this additional assumption (essentially because the same winning coalitions are sufficient to induce the requisite sequence of transitions). Suppose also that Assumptions 22.1 and 22.2 are satisfied.

Given the structure, it can be shown that when $\phi(s_j) = s_k$ where $k > j+1$, the dynamic game of franchise extension starting with state s_j will proceed gradually; there will first be an extension to a franchise of $\{1, \ldots, j+1\}$, then to $\{1, \ldots, j+2\}$, etc. until we reach $\{1, \ldots, k\}$. It is also possible to construct other examples of gradual franchise extension using this framework.

22.5.9. Stable Voting Rules and Constitutions. Let us now return to the question of self-stable coalitions posed in Barbera and Jackson discussed earlier. The society takes the form of $\mathcal{I} = \{1, \ldots, N\}$, and each state now directly corresponds to a "constitution" represented by a pair (a, b), where a and b are natural numbers between 1 and N. In Barbera and Jackson's interpretation, a votes are needed to implement a change in some policy variable away from status quo, while b votes are needed to change the current state. Barbera and Jackson consider cases both with $a, b \leq N/2$ and with $a, b > N/2$ (though they note that the former could lead to non-existence of equilibria). Let us simplify the discussion by focusing on the case where $a, b > N/2$. States with $a = b$ correspond to *voting rules*, and those with $a < b$ correspond to *constitutions*, where modifying the current state is more difficult than changing the policy.

Players differ in their ex-ante probability of favoring the change; assume that this probability for player i is $p(i)$, which induces preferences over states $s = (a, b)$. Without loss of generality, assume that $p(i)$ is nondecreasing in i. Barbera and Jackson show that this utility, $w_{(a,b)}(i)$, is "single-set-peaked," in that if there is more than one peak, these must be two neighboring peaks and that $w_{(a,b)}(i)$ satisfies a single-crossing condition. Moreover, it can be verified that generically (in terms of perturbations of p's), there is only one peak and the single-crossing condition holds strictly. To simplify the exposition let us suppose that there is a single peak and strict single crossing holds.

In Barbera and Jackson, voting rules and constitutions are assumed to have only one-step dynamics (i.e., any deviation ends the game). Consequently, it is possible that a rule is unstable in the sense of Barbera and Jackson, but players will not deviate from it if they play an infinite-horizon game (the converse, however, is true: a stable point in the sense of Barbera and Jackson is necessarily stable point in this game). Nevertheless, we may slightly alter the setup to replicate Barbera and Jackson's predictions about the stability on voting rules. In particular, let us augment any voting rule α by adding a state that yields the same payoff as α but cannot be changed (or, alternatively, requires unanimous voting to be changed). In that case, players would always be able to switch to this "additional" state which is impossible to move away from. cab with these assumptions, the following result can be established.

PROPOSITION 22.12. *In the modified version of the model of Barbera and Jackson described above:*

(1) *Assumptions 22.1 and 22.2 are satisfied.*

(2) *There exist mappings* ϕ_v *for the case of voting rules* ($a = b$) *and* ϕ_c *for the case of constitutions* ($a \leq b$) *that satisfy Axioms* 1–3.

(3) *Any stable voting rule* a *satisfies* $\phi_v(a) = a$.

(4) *Any stable constitution* b *satisfies* $\phi_c(b) = b$ *and moreover any* b *such that* $\phi_c(b) = b$ *is a stable constitution.*

22.5.10. Other Examples. let us briefly discuss a range of other examples. This discussion is not meant to provide a full analysis of the corresponding economic problems or even of these examples. Instead, it is simply meant to illustrate the variety of different economic interactions that can be incorporated into this framework.

Here are some additional examples.

EXAMPLE 22.1. **(Coalition Formation in Democracy)** *Suppose that there are three parties in the parliament, 1, 2, 3, and any two of them would be sufficient to form a government. Suppose that party 1 has more seats than party 2, which in turn has more seats than party 3. The initial state is* $\varnothing$, *and all coalitions are possible states. Since any two parties are sufficient to form a government, we have that* $W_\varnothing = \{\{1,2\},\{1,3\},\{2,3\},\{1,2,3\}\}$. *First suppose that all governments are equally strong and a party with a greater share of seats in the parliament will be more influential in the coalition government. Consequently,* $w_{\{1,2\}}(3) = 0 < w_{\{1,2,3\}}(3) < w_{\{1,3\}}(3) < w_{\{2,3\}}(3)$. *Other payoffs are defined similarly. In this case, it can be verified that* $\phi(\varnothing) = \{2,3\}$. *Next suppose that governments that have a greater number of seats in the parliament are stronger. In this case, we can have* $w_{\{1,2\}}(3) = 0 < w_{\{1,2,3\}}(3) < w_{\{2,3\}}(3) < w_{\{1,3\}}(3)$ *and likewise, so that the strongest two-party coalition may form, i.e.,* $\phi(\varnothing) = \{1,2\}$.

This model becomes more interesting if we enrich it further. In particular, suppose that after the coalition between parties 1 and 2 forms, party 1, by virtue of its greater number of seats, can sideline party 2 and rule by itself, i.e., $\{1,2\} \rightsquigarrow \{1\}$ *and* $\mathcal{W}_{\{1,2\}} = \{1\}$. *Suppose that party 2 can also do the same starting from the coalition with party 3, i.e.,* $\{2,3\} \rightsquigarrow \{2\}$ *and* $\mathcal{W}_{\{2,3\}} = \{2\}$. *However, once party 2 starts ruling by itself, then party 1 can regain power, that is,* $\{2\} \rightsquigarrow \{1\}$ *and* $W_{\{2\}} = \{1\}$, *because it has more seats than party 1. In this case, it can be shown that* $\phi(\varnothing) = \{2,3\}$.

What makes $\{2,3\}$ *dynamically stable in this case is the fact that* $\{2\}$ *is not dynamically stable itself. This example therefore reiterates the fundamental principle discussed after Theorem 22.1 and also in the context of coalition formation in nondemocracies in the previous subsection: the instability of states that can be reached from a state* s *contributes to the stability of state* s.

Example 22.2. **(Concessions in Civil War)** *Suppose that a government, G, is engaged in a civil war with a rebel group, R. The civil war state is denoted by c. The government can initiate peace and transition to state p, so that $\mathcal{W}_c = \{G\}$. However, $p \rightsquigarrow r$, where r denotes a state in which the rebel group becomes very strong and dominant in domestic politics. Moreover, $\mathcal{W}_p = \{R\}$, and naturally, $w_r(R) > w_p(R)$. If $w_r(G) < w_c(G)$, there will be no peace and $\phi(c) = c$ despite the fact that we may also have $w_p(G) > w_c(G)$. This illustrates the same forces as in the example on resistance to efficiency-enhancing reforms or to beneficial institutional changes discussed above.*

As an interesting modification, suppose next that the rebel group R can first disarm partially, in particular, $c \rightsquigarrow d$, where d denotes the state of partial disarmament. Moreover, $d \rightsquigarrow dp$, where the state dp involves peace with the rebels that have partially disarmed. Suppose that $\mathcal{W}_{dp} = \{G, R\}$, meaning that once they have partially disarmed, the rebels can no longer become dominant in domestic politics. In this case, provided that $w_{dp}(G) > w_d(G)$, we have $\phi(c) = dp$. Therefore, the ability of the rebel group to make a concession changes the set of dynamically stable states. This example therefore shows how the role of concessions can also be introduced into this framework in a natural way.

Although we have been primarily motivated by political economy environments, the next two examples show that the general idea developed here can be applied to a variety of other problems.

Example 22.3. **(Holdup)** *This example shows how the canonical model of holdup can be captured in a slightly modified version of our model. Imagine that there are two players, 1 and 2. The economic situation is one in which player 1 has to decide whether to make an investment and increase the quality of a good that he will supply to player 2. Player 2 then has to decide how much to pay to player 1. Let the states be $\{s_0, i_0, ni_0, ih, il, nih, nil\}$, where s_0 is the initial state, i_0 represents investments by player 1, and ni_0 represents no investments. We have $s_0 \rightsquigarrow i_0$ and $s_0 \rightsquigarrow ni_0$, and $\mathcal{W}_{s_0} = \{1\}$. This captures the fact that player 1 can decide unilaterally whether to invest or not. ih corresponds to investment by player 1 rewarded by high payment by player 2, whereas il involves investment by player 1 and low payment by player 2. nih and nil are defined similarly. Clearly, $i_0 \rightsquigarrow \{ih, il\}$ and $ni_0 \rightsquigarrow \{nih, nil\}$. However, we do not allow $s_0 \rightsquigarrow \{ih, il\}$ or $s_0 \rightsquigarrow \{nih, nil\}$, so that the transitivity the part, part (b), of Assumption 22.5 is relaxed. The characterization results provided in Theorems 22.3 and 22.4 continue to hold in this case. Since player 2 will unilaterally decide whether to make high or low payment, $\mathcal{W}_{i_0} = \mathcal{W}_{ni_0} = \{2\}$, and because the game ends after the payments, $\mathcal{W}_{ih} = \mathcal{W}_{il} = \mathcal{W}_{nih} = \mathcal{W}_{nil} = \emptyset$. With the usual assumptions on payoffs, in* *particular,* $w_{nih}(1) > w_{nil}(1)$*,* $w_{ih}(1) > w_{il}(1)$*, and* $w_{nih}(2) < w_{nil}(2)$*,* $w_{ih}(2) < w_{nil}(2)$*, and also assuming that states* s_o*,* i_0*, and* ni_0 *have the lowest payoffs, so that no player wishes to stay in these states, we can verify that this is a special case of our general model and* $\phi(s_0) = nil$*, even though nih may involve higher rewards for both players.*

22.6. Conclusions

What have we learned? The material here presents some new ideas and reiterates one of the main themes of the course. The main theme throughout is that we have to think of institutions (and constitutions as a special form of political institutions) as a way of influencing the future process of collective decision-making. In the baseline Acemoglu-Robinson (2006) framework, this is the role played by democracy versus dictatorship (and perhaps by different forms of democratic institutions). Here, constitutions play the same role by determining what type of majorities are necessary to change the status quo.

The new lessons are as follows:

(1) Models in which constitutional/institutional rules are chosen before preferences are fully realized (but heterogeneity of views or interests is still present) provide a nice and simple framework for analyzing institutional design. The issue of whether constitutions are self stable is a particular example of this type of institutional design question. There will been many more institutional design type questions in the future of political economy research.

(2) Despite our ability to write interesting models of institutional design and the effect of constitutions, we are still far from an understanding of why constitutions matter. Why aren't constitutions, laws and institutions simply "cheap talk"? (Perhaps they are...). This question is present in the background of many of our analyses, and we do not have a satisfactory answer yet.

(3) Most importantly, the process of institutional and constitutional reform will not necessarily lead to efficient outcomes. Instead, there will be strong forces leading towards *stability*, that is, to constitutions or institutional structures which are themselves stable, so that players or social groups are not afraid of further changes making them worse off once the process of reform has started.

22.7. References

(1) Acemoglu, Daron, Georgy Egorov and Konstantin Sonin (2008) "Dynamics and Stability of Constitutions, Coalitions and Clubs" mimeo.

(2) Acemoglu, Daron and James Robinson (2006) *Economic Origins of Dictatorship and Democracy*, Cambridge University Press.

(3) Badger, Wade (1972) "Political Individualism, Positional Preferences And Optimal Decision Rules" in *Probability Models of Collective Decision-Making* edited by R.G. Niemi and H.F. Weisberg, Columbus, Ohio Merrill Publishing.

(4) Barbera, Salvador and Matthew Jackson (2004) "Choosing How to Choose: Self-Stable Majority Rules and Constitutions," *Quarterly Journal of Economics*, 119, 1011-1048.

(5) Koray, Semih (2000) "Self-Selective Social Choice Functions Verify Arrow and Gibbard- Satterthwaite Theorems," *Econometrica*, 68, 981-995.

(6) Lagunoff, Roger (2004) "Markov Equilibrium in Models of Dynamic Endogenous Political Institutions" http://www.georgetown.edu/faculty/lagunoff/

(7) Lagunoff, Roger (2006) "Dynamic Stability and reform of Political Institutions," http://www.georgetown.edu/faculty/lagunoff/

(8) Loveman, Brian (1993) The constitution of tyranny: regimes of exception in Spanish America, Pittsburgh: University of Pittsburgh Press.

(9) Mailath, George J., Stephen Morris and Andrew Postelwaite (2001) "Laws and Authority," //www.econ.yale.edu/~sm326/authority.pdf

(10) Weingast, Barry R. (1998) "Political Stability and Civil War: Institutions, Commitment, and American Democracy," in Robert Bates, Avner Greif, Margaret Levi, Jean-Laurent Rosenthal and Barry R. Weingast eds. *Analytic Narratives.* Princeton; Princeton University Press.

CHAPTER 23

Dynamics of Political Compromise

The main question addressed by Dixit, Grossman and Gul (2000) is how to sustain political compromise in a dynamic democratic system, without the ability to write such compromises in the constitution. Essentially, this is the same as the "incentive-compatible promises" studied previously. While the current model is motivated by democratic politics, it is also potentially useful for understanding compromises in non-democratic politics.

The main results of the Dixit, Grossman and Gul (2000) model will be as follows:

- Political compromise related to satisfying "incentive compatibility" supported by the threat of future punishments as in repeated games.
- Subgame perfect equilibria have a relatively simple "quasi-Markovian" structure, whereby allocations depend only on the majority of the party in power and whether this party was even stronger in the past.
- Supermajority type rules designed to protect minorities can have counterintuitive results and reduce compromise (because they make future punishments more difficult)

23.1. Modeling Political Compromise

23.1.1. Model. There are two political "parties", party 1 and party 2, sharing a pie of size 1. The need for compromise is modeled by assuming that both parties are risk-averse, each with utility function $U^i(\cdot)$, which has the party's share of the pie in the period as its argument. U^i is continuous, increasing, and strictly concave, with $U^i(0) = 0$, and $U^i(1) = 1$.

Time is discrete and infinite $t = 0, 1, 2, \ldots, \infty$, and each party maximizes its discounted expected utility with discount factor δ.

The distribution of political power is captured by the state of the political system. There are K possible states of the political system. Let $S = \{1, 2, \ldots, K\}$ denote the set of states, with X_t denoting the state at the time t. Let $\{S^1, S^2\}$ be a partition of S, such that when $X_t \in S^1$, party 1 is "in power" and gets to decide the division of the pie. Similarly, if $X_t \in S^2$, then party 2 is in power, and this party decides how to split the pie in period t.

Suppose that X_0 is degenerate, with $\Pr[X_0 = x_0] = 1$ for some $x_0 \in S$, which means that equilibrium outcomes will be indexed by this initial state. After this initial date, a Markov process ϕ determines the transitions of the political system. Let the resulting transition probabilities in this Markov process be:

$$p_{kl} = \Pr[X_{t+1} = l \mid X_t = k]$$

The important assumption here, which makes the model tractable, is that policies, in particular the division of the pie, does not affect transitions. This, for example, rules out situations in which greater resources for a given group increases their power in the future. It also rules out the types of institutional commitments analyzed above in the context of the Acemoglu-Robinson (2006) framework. But in other respects, the current model is considerably richer than those we have studied so far.

A history $h_t = (x_0, x_1, \ldots, x_t)$ refers to a possible realization of the political states X_0 to X_t in the first $t+1$ periods. Let H_t denote the set of all $t+1$ period histories having positive probability; that is, $H_t = \{h_t \mid \Pr[(X_0, X_1, \ldots, X_t) = h_t] > 0\}$.

In any period, the stage game involves simply the choice by the party in power (knowing the exact state) regarding the division of the pie. Then the state next period is realized etc.

This implies that a strategy for the in-power party, then, is a mapping from the history and the current state to a number in $[0, 1]$. The out-of-power party takes no action until it regains power.

The relevant concept of equilibrium is subgame perfect equilibrium, so that strategies are best responses given all histories.

The key will be an "incentive compatibility" constraint for the party in power. The best deviation for the party in power is to grab the whole pie, and from then on, there will be punishments forever where all future parties will grab the whole pie.

The subgame perfect equilibrium will be characterized as a *division processes*, $\rho = \{Y_0, Y_1, Y_2 \ldots\}$ such that Y_t is a random variable representing the allocation to party 1 at time t, and each Y_t is known given the realization of the states from X_0 through X_t, we will know what allocation should emerge in period t. Since X_0 is degenerate, so too will be the random variable Y_0 describing the allocation in the initial state. Denote the realizations of Y_t by $y_t(h_t)$ and the initial allocation by $y(x_0)$.

For any division process ρ, define the *payoffs* for the two parties:

$$V_t^1(\rho, h_t) = \mathbb{E}\left[\sum_{\tau \geq t} \delta^{\tau - t} U^1(Y_\tau) \mid (X_0, X_1, \ldots, X_t) = h_t\right] \,,$$

and

$$V_t^2(\rho, h_t) = \mathbb{E}\left[\sum_{\tau \geq t} \delta^{\tau - t} U^2(1 - Y_\tau) \mid (X_0, X_1, \ldots, X_t) = h_t\right] .$$

V_t^i is the expected value for party i, discounted to time t, of pursuing the process ρ from t onward, if the history until t has been h_t.

Let $\bar{\rho} = \{\bar{Y}_0, \bar{Y}_1, \ldots, \}$ denote the division process associated with lack of cooperation (which is also the Markovian equilibrium). Thus $\bar{Y}_t = 1$ if $X_t \in S^1$ and $\bar{Y}_t = 0$ if $X_t \in S^2$. It is clear that this is also the min max for both players, and thus will act as the punishment strategy.

Incentive compatibility therefore requires that the payoff to the party in power from continuing along the path ρ be at least as great as the payoff from $\bar{\rho}$. In other words, *incentive compatibility* requires

$$V_t^i(\rho, h_t) \geq V_t^i(\bar{\rho}, h_t)$$

for all $t \geq 0$ and all $h_t \in H_t$ such that $x_t \in S^i$.

Starting at time 0 in state x_0, one can trace out the whole Pareto frontier of payoffs, namely the set $(v^1, v^2) \in R^2$ such that

$$v^2 = \max_{\rho \in \mathcal{F}} V_0^2(\rho, x_0) \qquad \text{subject to} \qquad V_0^1(\rho, x_0) \geq v^1 , \tag{23.1}$$

where $\mathcal{F}$ is the set of all feasible incentive-compatible division processes; i.e., the processes for which $V_t^i(\rho, h_t) \geq V_t^i(\bar{\rho}, h_t)$ for all $t \geq 0$ and for all $h_t \in H_t$ such that $x_t \in S^i$.

23.1.2. Pareto Optimal Equilibria. Let us now focus on the solution to the program (23.1).

First, note that if the incentive compatibility constraints were not present, "policy smoothing" would dictate a constant policy path, since the utility function of both parties are concave.

Next, let us first define a crucial object, $y^*(x_0)$, which will play a key role in the analysis.

Let $y^*(x_0)$ be the initial allocation in an equilibrium, which gives the maximum to the party not in power consistent with the incentive compatibility constraint of the party in power. In other words, let $\rho^* = \{y^*(x_0), Y_1^*, Y_2^*, \ldots\}$ be the equilibrium division process, then we must have $V_t^i(\rho^*, h_t) = V_t^i(\bar{\rho}, h_t)$ for the party in power.

Then we have:

Proposition 23.1. *The division process $\rho = \{y(x_0), Y_1, Y_2, \ldots\}$ is efficient if and only if* *(i)* $y(x_0) \begin{cases} \geq y^*(x_0) & \text{if } x_0 \in S^1 \\ \leq y^*(x_0) & \text{if } x_0 \in S^2 \end{cases}$

and

(ii) $Y_t = \begin{cases} \max[y^*(X_t), Y_{t-1}] & \text{if } X_t \in S^1 \\ \min[y^*(X_t), Y_{t-1}] & \text{if } X_t \in S^2 \end{cases},$

for all $t \geq 1$.

To interpret this proposition, suppose that party 1 is in power. Then the proposition states that the share of party 1 is constant over time until one of two things happens:

- Party 1 is still in power but the state changes to X_t such that had X_t be in the initial state, party 1 would have received more, i.e., $y^*(X_t) > y^*(x_0)$. In this case, the share of party 1 increases.
- Party 2 loses power and the new state X_t is such that $y^*(X_t) < y^*(x_0)$.

The equilibrium then continues starting with this new division.

The reason why there are these changes in the division process is because the incentive compatibility constraint of one of the two parties binds at certain points. As long as the incentive compatibility constraints of both parties are slack, there is no change in the division, because of the policy smoothing motive. Whenever an incentive compatibility constraint binds, the share of party 1 goes up or down.

The mathematical intuition is as follows: let $X_t \in S^1$. First it must $y^*(x_0)$ initial division is monotonic in the present value of expected utility of a party. Thus it acts like a "sufficient statistic".

Next, suppose party 1is in power at time t. Then it can also be shown that for all $t \geq 1$, $Y_t \geq Y_{t-1}$ as long as party 1 remains in power. Suppose to the contrary that there exists a history h_t with $x_t \in S^1$ such that $y(h_t) < y(h_{t-1})$ (where with a slight abuse of notation $y(h_t)$ is the share of party 1 after history h_t). Then it would be possible to raise $y(h_t)$ and lower $y(h_{t-1})$ to some common intermediate value in such a way that expected utility at $t-1$ rises for both parties while that for party 1 also rises at t, and the incentive compatibility constraints would be satisfied. This would contradict the assumed efficiency of ρ

Finally, it can be shown that for $t \geq 1$, $Y_t \leq \max[Y_{t-1}, y^*(X_t)]$. Again suppose to the contrary that there exists a history h_t such that $y(h_t) > \max[y(h_{t-1}), y^*(x_t)]$. The fact that $y(h_t) > y^*(x_t)$ implies that the incentive compatibility of party 1 (which is in power) is slack. Then we can reduce $y(h_t)$ and increase $y(h_{t-1})$ to increase V^i_{t-1} and V^j_{t-1} (from concavity) without violating feasibility.

Taken together these three statements imply $Y_t = \max[Y_{t-1}, y^*(X_t)]$ when $X_t \in S^1$. Similarly, $Y_t = \min[Y_{t-1}, y^*(X_t)]$ for $X_t \in S^2$ follows.

Now we provide the details of the proof of Proposition 23.1.

Proof of Proposition 23.1:

Let us start with the "only if" part of the proof. Consider the problem of maximizing the expected present value of the party not in power at date 0, subject to giving the one in power a payoff at least equal to an arbitrarily specified value α (which need not equal the payoff under the punishment division process $\bar{\rho}$).

Since the values of period utilities U^i are bounded in [0,1], and the payoff functions V_t^i are defined as the expected present values of utilities, the values of V_t^i are also bounded, and lie in the range $[0, 1/(1-\delta)]$. Therefore, for any $\alpha \in [0, 1/(1-\delta)]$, any initial state where party i is in power ($x_0 \in S^i$), and $j \neq i$, we can define

$$\mathcal{L}(\alpha) = \max_{\rho \in \mathcal{F}} \; V_0^j(\rho, x_0) \quad \text{subject to} \quad V_0^i(\rho) \geq \alpha \,.$$

A division process that solves this problem for some α will be labelled *efficient*.

Next by a series of lemmas, existence and uniqueness of efficient division processes will be established.

Lemma 23.1. *For any $x_0 \in S^i$ and $\alpha \in [0,1]$, either $\mathcal{L}(\alpha)$ has a unique solution, or $V_0^i(\rho, x_0) < \alpha$ for all $\rho \in \mathcal{F}$.*

Proof. Obviously, if $V_0^i(\rho, x_0) < \alpha$ for all $\rho \in \mathcal{F}$, then the maximization problem has no solution. So consider the opposite case, and define

$$M(\alpha) = \sup_{\rho \in \mathcal{F}} \; V_0^j(\rho, x_0) \quad \text{subject to} \quad V_0^i(\rho, x_0) \geq \alpha \,.$$

Specifying a division process is equivalent to specifying a sequence of real numbers, one for each t-period history of the Markov process ϕ. Therefore each ρ can be identified with an element of $[0,1]^\infty$. Endow this space with the product topology. Then the payoff functions V_0^1, V_0^2 are continuous, and the feasible set $\mathcal{F} = \{\, \rho \mid V_t^i(\rho, h_t) \geq V_t^i(\bar{\rho}, h_t)$ for all $t \geq 0$ and for all $h_t = (x_0, x_1, \ldots x_t)$ such that $x_t \in S^i\}$, is compact. Consequently, there exists a subsequence $\rho^{(n)} \in \mathcal{F}$ such that $\rho^{(n)} \to \rho \in \mathcal{F}$ and $V^j(\rho, x_0) = M(\alpha)$. This is then a solution to $\mathcal{L}(\alpha)$. Uniqueness follows immediately from the strict concavity of V^i and V^j and the convexity of $\mathcal{F}$. □

Now if $\alpha = V_0^i(\bar{\rho}, x_0)$, the constraint on the maximization problem is the incentive compatibility constraint of the party initially in power. This is automatically satisfied by the division process $\rho^* = \{y^*(x_0), Y_1^*, Y_2^*, \ \ldots\}$, so $\mathcal{L}(V_0^i(\bar{\rho}, x_0))$ has a unique solution, ρ^*.

Lemma 23.2. *Let α_1, α_2 be any two numbers in $[0,1]$. For $m = 1$, 2, denote the unique solutions to $\mathcal{L}(\alpha_m)$ by $\rho^m = \{y^m(x_0), Y_1^m, Y_2^m, \ \ldots\}$. Then $\alpha_1 > \alpha_2$ implies*

$$y^1(x_0) \geq y^2(x_0) \text{ if } i = 1 \qquad \text{and} \qquad y^1(x_0) \leq y^2(x_0) \text{ if } i = 2\,.$$

This lemma says that if the constraint on the player in power is tighter, then he or she has two start off with a larger share of the surplus.

Proof. Consider the case $i = 1$ and $j = 2$ and define

$$W^2(\beta) = \max_{\rho \in \mathcal{F}} \ \mathbb{E}_{X_1}\left[V_1^2(\rho, (x_0, X_1))\right] \quad \text{subject to} \quad \mathbb{E}_{X_1}\left[V_1^1(\rho, (x_0, X_1))\right] \geq \beta\,.$$

The argument used in the proof of Proposition 23.1 establishes that if the feasible set of this maximization problem is non-empty, then it has a unique solution; that is, $W^2(\beta)$ is attained for some ρ. Then we can re-write the maximization problem $\mathcal{L}(\alpha)$ as

$$\max_z \left\{ U^2(1-z) + \delta\ W^2\left(\frac{\alpha - U^1(z)}{\delta}\right)\right\}.$$

Since ρ^1 is the unique solution to $\mathcal{L}(\alpha_1)$, we have

$$U^2(1-y^1(x_0)\,) + \delta\ W^2\left(\frac{\alpha_1 - U^1(y^1(x_0)\,)}{\delta}\right) > U^2(1-y^2(x_0)\,) + \delta\ W^2\left(\frac{\alpha_1 - U^1(y^2(x_0)\,)}{\delta}\right).$$

Rearranging yields

$$\begin{aligned}&\delta\ \left[W^2\left(\frac{\alpha_1 - U^1(y^1(x_0)\,)}{\delta}\right) - W^2\left(\frac{\alpha_1 - U^1(y^2(x_0)\,)}{\delta}\right)\right]\\ > \ &U^2(1-y^2(x_0)\,) - U^2(1-y^1(x_0)\,)\,.\end{aligned}$$

Similarly, since ρ^2 is the unique solution of $\mathcal{L}(\alpha^2)$, we have

$$\begin{aligned}&\delta\ \left[W^2\left(\frac{\alpha_2 - U^1(y^2(x_0)\,)}{\delta}\right) - W^2\left(\frac{\alpha_2 - U^1(y^1(x_0)\,)}{\delta}\right)\right]\\ > \ &U^2(1-y^1(x_0)\,) - U^2(1-y^2(x_0)\,)\,.\end{aligned}$$

Combining these two, we get

$$\begin{aligned}&W^2\left(\frac{\alpha_1 - U^1(y^1(x_0)\,)}{\delta}\right) - W^2\left(\frac{\alpha_1 - U^1(y^2(x_0)\,)}{\delta}\right)\\ > \ &W^2\left(\frac{\alpha_2 - U^1(y^1(x_0)\,)}{\delta}\right) - W^2\left(\frac{\alpha_2 - U^2(y^2(x_0)\,)}{\delta}\right)\end{aligned} \tag{23.2}$$

Since the utility functions U^i are concave, the set $\mathcal{F}$ of feasible division processes is convex, and the function W^2 is concave. Therefore

$$W^2(a+\Delta) - W^2(a) > W^2(b+\Delta) - W^2(b) \text{ for } a > b \text{ implies } \Delta < 0.$$

Using this in (23.2), we observe that

$$\frac{1}{\delta}\;[\;U^1(y^2(x_0)\,) - U^1(y^1(x_0)\,)\;] < 0\,.$$

Since U^1 is increasing, this implies $y^1(x_0) > y^2(x_0)$. □

The following lemma formalizes the notion that policy smoothing is beneficial.

Lemma 23.3. *Let $h_t = \{x_0, x_1, \dots x_t\} \in H_t$, and $p = p_{lm}$ for $l = x_{t-1}$ and $m = x_t$. Let $\rho = \{y_0, Y_1, Y_2, \dots\}$ be a division process. If $p > 0$ and $y(h_t) \neq y(h_{t-1})$, then there exists another division process $\hat{\rho}$ such that $V_0^j(\hat{\rho}, x_0) > V_0^j(\rho, x_0)$ for $j = 1,\ 2$.*

Proof. Define

$$\gamma \equiv [\delta\, p\, y(h_t) + y(h_{t-1})]/(\delta p + 1)\,,$$

and define $\hat{\rho}$ by $\hat{\rho}(h_{t-1}) = \hat{\rho}(h_t) = \gamma$ and $\hat{\rho} = \rho$ for all other dates and/or histories. The result follows from the strict concavity of the V^i. □

Now returning to the proof of Proposition 23.1, consider the case $X_t \in S^1$; the proof for the other case is symmetric. To prove

$$Y_t = \max[y^*(X_t), Y_{t-1}]\,,$$

we need to establish three properties:

(1) For all $t \geq 0$, $Y_t \geq y^*(X_t)$:

Proof. Suppose for any t and a history $h_t \in H_t$ we have $y(h_t) < y^*(x_t)$. Let α be the value of $V_t^1(\rho, h_t)$, and α^* the value of $V_t^1(\bar{\rho}, h_t)$.

The continuation of ρ from (t, h_t) must be efficient; therefore it solves $\mathcal{L}(\alpha)$. And, since the law of motion of the state variable X_t, ϕ, is Markovian, $\alpha^* = V_0^1(\bar{\rho}, x_t)$, that is, regarding the problem as if it were starting afresh with the initial state x_t. Our assumption that $y(h_t) < y^*(x_t)$, using Lemma 23.2, implies $\alpha \leq \alpha^*$, and uniqueness of efficient processes (Lemma 23.1) strengthens this to $\alpha < \alpha^*$. But then $V_t^1(\rho, h_t) < V_t^1(\bar{\rho}, h_t)$, so $\rho \notin \mathcal{F}$, contradicting feasibility. □

(2) For all $t \geq 1$, $Y_t \geq Y_{t-1}$:

Proof. If for some history h_t we have $y(h_t) < y(h_{t-1})$, then define γ as in Lemma 23.3. Obviously $p > 0$. By concavity of U^1,

$$U^1(h_{t-1}) + \delta\, p\, U^1(h_t) < U^1(\gamma) + \delta\, p\, U^1(\gamma)\,,$$

and the values on all other realizations are equal. Therefore the $\hat{\rho}$ constructed in Lemma 23.3 is feasible, and there it was shown to give higher expected utilities for both players. This contradicts the efficiency of ρ. □

(3) For all $t \geq 1$, $Y_t \leq \max[y^*(X_t), Y_{t-1}]$:

Proof. Suppose for some t, and a history $h_t \in H_t$ we have $y(h_t) > y^*(x_t)$. Defining α and α^* as in Part I, above, here we have $\alpha > \alpha^*$.

Suppose we simultaneously have $y(h_t) > y(h_{t-1})$, and define $\hat{\rho}$ as in Part II above. For some $\lambda \in [0, 1]$, define $\rho' = \lambda\,\rho + (1-\lambda)\,\hat{\rho}$. Then $V^1_{t-1}(\rho', h_{t-1}) > V^1_{t-1}(\rho, h_{t-1})$ because of the benefit from smoothing. Starting at (t, h_t), we have $V^1_t(\rho', h_t) < V^1_t(\rho, h_t)$, but because U^1 is continuous the difference can be made arbitrarily small by choosing λ sufficiently close to 1. Also, under the provisional assumptions of this part, $V^1_t(\rho, h_t) > \alpha^* = V^1_t(\bar{\rho}, h_t)$. Therefore $V^1_t(\rho', h_t) \geq V^1_t(\bar{\rho}, h_t)$. For all other dates and histories, ρ' and ρ give equal values. Therefore ρ' is feasible.

Then by the strict concavity of U^1 and U^2, we get $V^j_0(\rho', x_0) > V^j_0(\rho, x_0)$ for $j = 1$, 2. This contradicts the efficiency of ρ. □

These three properties together establish necessity and complete the proof of the "only if" part.

Finally, to prove the "if" part (efficiency), namely that every division process of this form is efficient, consider any process $\rho = \{y_0, Y_1, Y_2, \dots\}$ satisfying the conditions (i) and (ii) in the statement of the proposition. For ease of notation suppose $x_0 \in S^1$; the argument for $x_0 \in S^2$ is analogous. Let $\alpha = V^1(\rho, x_0)$ and suppose $\hat{\rho} = \{\hat{y}_0, \hat{Y}_1, \hat{Y}_2, \dots\}$ solves the problem $\mathcal{L}(\alpha)$. By the "only if" part of the proposition proved above, $\hat{\rho}$ satisfies the conditions (i) and (ii).

Both ρ and $\hat{\rho}$ satisfy (ii), which ensures that along any path, the process with the higher initial share of party 1, y_0 or $\hat{y}_0$, will also have a uniformly higher share, y_t or $\hat{y}_t$, up to some time t (possibly $t = \infty$) and the two will coincide thereafter. Therefore

$$V^1_0(\hat{\rho}, x_0) \geq V^1_0(\rho, x_0) \text{ if and only if } \hat{y}_0 \geq y_0$$

and

$$V^2_0(\hat{\rho}, x_0) \geq V^2_0(\rho, x_0) \text{ if and only if } \hat{y}_0 \leq y_0\,.$$

In the present context, the constraint in $\mathcal{L}(\alpha)$ gives us $V^1_0(\hat{\rho}, x_0) \geq V^1_0(\rho, x_0)$, while the fact that $\hat{\rho}$ is efficient gives us $V^2_0(\hat{\rho}, x_0) \geq V^2_0(\rho, x_0)$. The two together imply $\hat{y}_0 = y_0$. Then $\hat{\rho} = \rho$, establishing the efficiency of ρ. This completes the proof of Proposition 23.1.■

An important feature is that the structure of the division process is independent of the Markov process describing power switches. It is, however, possible to derive more results by specifying this Markov process.

Consider two states k, $\ell \in S^1$, and suppose that party 1 will surely hold power longer starting in ℓ than in k. This implies that its incentive constraint will be more severe in ℓ than in k. Then, presumably, it should be given more in ℓ than in k to stave off defection, that is, $y^*(\ell) > y^*(k)$.

Let k be some state in S^1. Define the set of precursors of k, $P(k)$, as the set of states $\ell \in S^1$ such that every positive-probability path from ℓ to S^2 goes through k. From any state ℓ in $P(k)$, party 1 can fall from power only if it first reaches the state k. This means that loss of power appears unambiguously more distant as seen from state ℓ as compared to state k. Then

Proposition 23.2. *For all $\ell \in P(k)$, $y^*(\ell) \geq y^*(k)$.*

This proposition defines a partial order P over states, and captures the intuitive notion that when the party in power is stronger in some states it should receive more in those states.

In fact, by specifying a tighter structure on the Markov process, it is possible to define a complete order as well, but we will not get into this in these notes.

23.1.3. Supermajority Rules. Let us now use this model to investigate the implications of "supermajority rules" in this setup. Supermajority rules have become very popular in recent decades, and their justification is often the protection of minorities (ideological, ethnic or income-wise), by making it harder for the party in power to legislate changes. Given the simplicity of the model here, supermajority rules can be in a straightforward manner.

Suppose, in particular, that there exists some states in S^0, where neither party 1 nor party 2 can they legislate (i.e., make) changes without the consent of the other party. In particular, in these states the *status quo* continues unless both parties agree to a change.

Formally, the stage game in period $t \geq 1$ involves announcements by each party of a proposed allocation for the period. If $X_t \in S^i$ for $i = 1$ or 2, then the proposal by party i is implemented. If $X_t \in S^0$ and the proposals coincide, the common proposal is put into effect. Otherwise, $Y_t = Y_{t-1}$. We restrict attention to processes that have an initial realization x_0 in S^1 or S^2.

Similar to before, define the equilibrium division process as $\bar{\rho}^s = \{\bar{Y}_0^s, \bar{Y}_1^s, \ldots\}$, now applying for the supermajority regime. At any time t, if party 1 has the current supermajority or had the last supermajority, then set $\bar{Y}_t^s = 1$. Whereas if party 2 has the current supermajority or had the last supermajority, then set $\bar{Y}_t^s = 0$. This process differs from the corresponding process for a regular majority regime in as much as the allocation changes only when a new party captures supermajority support.

Incentive-compatibility now requires that for $i = 1, 2$, we require $V_t^i(\rho, h_t) \geq V_t^i(\bar{\rho}^s, h_t)$ for all histories $h_t \in H_t$ such that $x_t \in S^i$. This means that a party in a supermajority state must have no unilateral incentive to deviate, considering the punishment that would (eventually) ensue. For histories $h_t \in H_t$ such that $x_t \in S^0$ we have two cases to consider. If the process ρ has $y(h_t) \neq y(h_{t-1})$ then we must have $V_t^1(\rho, h_t) \geq V_t^1(\bar{\rho}^s, h_t)$ and $V_t^2(\rho, h_t) \geq V_t^2(\bar{\rho}^s, h_t)$; *both* parties must accede to any change in the allocation, and so neither can have any incentive to deviate. Alternatively, if the process ρ has $y(h_t) = y(h_{t-1})$, then no incentive constraint applies. Since constancy of shares is the default action, it does not require ratification by either party.

Analogous the to before, define $y^{*s}(x_0)$ as the first element in $\rho^{*s} \equiv \{y^{*s}(x_0), Y_1^{*s}, Y_2^{*s}, \dots\}$, where ρ^{*s} solves

$$\max_{\rho \in \mathcal{F}^s} V_0^j(\rho, x_0)$$

for $x_0 \in S^i$ and $j \neq i$. Here $\mathcal{F}^s$ is the set of feasible division processes; i.e., those that satisfy all of the incentive-compatibility constraints of the supermajority regime. Then we have:

Proposition 23.3. *The division process $\rho = \{y(x_0), Y_1, Y_2, \ldots\}$ is efficient under supermajority rule if and only if*

(i) $y(x_0) \begin{cases} \geq y^{*s}(x_0) & \text{if } x_0 \in S^1 \\ \leq y^{*s}(x_0) & \text{if } x_0 \in S^2 \end{cases}$

and

(ii) $Y_t = \begin{cases} \max[y^{*s}(X_t), Y_{t-1}] & \text{if } X_t \in S^1 \\ \min[y^{*s}(X_t), Y_{t-1}] & \text{if } X_t \in S^2 \\ Y_{t-1} & \text{if } X_t \in S^0 \end{cases}$

for all $t \geq 1$.

As before, an efficient process begins by giving a party with an initial supermajority enough to deter its immediate defection at the start of the game. Thereafter, policy changes occur only when some party enjoys a supermajority state. Even then, the *status quo* persists unless the previous split of the pie would not have been enough to deter the strong party from grabbing the pie, had the current state been the starting point.

The form of an efficient process under supermajority rule resembles that under simple majority rule. The new feature is the addition of states in which neither party has authority to change policy unilaterally. The proposition shows that in these states no policy change takes place.

The interesting question is whether an equilibrium with supermajority rule is better than one with majority rule. It turns out that generally supermajority is not beneficial since its reduces the punishments against the party in power. This makes compromise more difficult and reduces policy smoothness.

23.2. Political Stability and Political Compromises

An intuition going back to Olson some suggests that greater political stability, and more durable dictatorships, are good for economic growth, because they imply that the group in power will internalize the future implications of their actions. This intuition is not correct however. We will now see this using a model of dynamic political compromise including production.

23.2.1. Demographics, Preferences and Technology. We consider an infinite horizon economy in discrete time with a unique final good. The economy consists of N parties (groups). Each party j has utility at time $t = 0$ given by

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u_j(c_{j,t}, l_{j,t}), \tag{23.3}$$

where $c_{j,t}$ is consumption, $l_{j,t}$ is labor supply (or other types of productive effort), and $\mathbb{E}_0$ denotes the expectations operator at time $t = 0$. To simplify the analysis without loss of any economic insights, we assume that labor supply belongs to the closed interval $\left[0, \bar{l}\right]$ for each party. We also impose the following assumption on utility functions.

Assumption 23.1. ***(utility function)*** *The instantaneous utility function*

$$u_j : \mathbb{R}_+ \times \left[0, \bar{l}\right] \to \mathbb{R},$$

for $j = 1, ..., N$ is uniformly continuous, twice continuously differentiable in the interior of its domain, strictly increasing in c, strictly decreasing in l and jointly strictly concave in c and l, with $u_j(0, 0) = 0$ and satisfies the following Inada conditions:

$$\lim_{c \to 0} \frac{\partial u_j(c, l)}{\partial c} = \infty \text{ and } \lim_{c \to \infty} \frac{\partial u_j(c, l)}{\partial c} = 0 \text{ for all } l \in \left[0, \bar{l}\right],$$

$$\frac{\partial u_j(c, 0)}{\partial l} = 0 \text{ and } \lim_{l \to \bar{l}} \frac{\partial u_j(c, l)}{\partial l} = -\infty \text{ for all } c \in \mathbb{R}_+.$$

The differentiability assumptions enable us to work with first-order conditions. The Inada conditions ensure that consumption and labor supply levels are not at corners. The concavity assumptions are also standard, but important for our results, since they create a desire for consumption and labor supply smoothing over time.

The economy also has access to a linear aggregate production function given by

$$Y_t = \sum_{j=1}^{N} l_{j,t}. \tag{23.4}$$

23.2.2. Efficient Allocation without Political Economy. As a benchmark, we start with the efficient allocation without political economy constraints. This is an allocation that maximizes a weighted average of different groups' utilities, with *Pareto weights* vector denoted by $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)$, where $\alpha_j \geq 0$ for $j = 1, ..., N$ denotes the weight given to party j. We adopt the normalization $\sum_{j=1}^{N} \alpha_j = 1$. The program for the (unconstrained) efficient allocation can be written as:

$$\max_{\left\{[c_{j,t}, l_{j,t}]_{j=1}^{N}\right\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \left[\sum_{j=1}^{N} \alpha_j u_j(c_{j,t}, l_{j,t}) \right] \tag{23.5}$$

subject to the resource constraint

$$\sum_{j=1}^{N} c_{j,t} \leq \sum_{j=1}^{N} l_{j,t} \text{ for all } t. \tag{23.6}$$

Standard arguments imply that the *first-best allocation*, $\left\{ \left[c_{j,t}^{fb}, l_{j,t}^{fb} \right]_{j=1}^{N} \right\}_{t=0}^{\infty}$, which is a solution to the solution to program (23.5), satisfies the following conditions:

$$\text{no distortions: } \frac{\partial u_j(c_{j,t}^{fb}, l_{j,t}^{fb})}{\partial c} = -\frac{\partial u_j(c_{j,t}^{fb}, l_{j,t}^{fb})}{\partial l} \text{ for } j = 1, ..., N \text{ and all } t,$$

$$\text{perfect smoothing: } c_{j,t}^{fb} = c_j^{fb} \text{ and } l_{j,t}^{fb} = l_j^{fb} \text{ for } j = 1, ..., N \text{ and all } t.$$

The structure of the first best allocations is standard. Efficiency requires the marginal benefit from additional consumption to be equal to the marginal cost of labor supply for each individual, and also requires perfect consumption and labor supply smoothing.

Note that different parties can be treated differently in the first-best allocation depending on the Pareto weight vector $\boldsymbol{\alpha}$, i.e., receive different consumption and labor allocations.

23.2.3. Political Economy. We now consider a political environment in which political power fluctuates between the N parties $j \in \mathcal{N} \equiv \{1, ..., N\}$. The game form in this political environment is as follows.

(1) In each period t, we start with one party, j', in power.

(2) All parties simultaneously make their labor supply decisions $l_{j,t}$. Output $Y_t = \sum_{j=1}^{N} l_{j,t}$ is produced.

(3) Party j' chooses consumption allocations $c_{j,t}$ for each party subject to the feasibility constraint

$$\sum_{j=1}^{N} c_{j,t} \leq \sum_{j=1}^{N} l_{j,t}. \tag{23.7}$$

(4) A first-order Markov process m determines who will be in power in the next period. The probability of party j being in power following party j' is $m(j \mid j')$, with $\sum_{j=1}^{N} m(j \mid j') = 1$ for all $j' \in \mathcal{N}$.

A number of features is worth noting about this setup. First, this game form captures the notion that political power fluctuates between groups. Second, it builds in the assumption that the allocation of resources is decided by the group in power (without any prior commitment to what the allocation will be). The assumption of no commitment is standard in political economy models (e.g., Persson and Tabellini, 2000, Acemoglu and Robinson, 2006a), while the presence of power switches is crucial for our focus (see also Dixit, Grossman and Gul, 2000, and Amador, 2003a,b). In addition, we have simplified the analysis by assuming that there are no constraints on the allocation decisions of the group in power and by assuming no capital.

In addition, we impose the following assumption on the Markov process for power switches.

ASSUMPTION 23.2. ***(Markov process)*** *The first-order Markov chain $m(j \mid j')$ is irreducible, aperiodic and ergodic.*

We are interested in *subgame perfect equilibria* of this infinitely-repeated game. More specifically, as discussed in the Introduction, we will look at subgame perfect equilibria that correspond to constrained Pareto efficient allocations, which we refer to as *Pareto efficient perfect equilibria*.[1]

To define these equilibria, we now introduce additional notation. Let $h^t = (h_0, ..., h_t)$, with $h_s \in \mathcal{N}$ be the history of power holdings. Let H^∞ denote the set of all such possible histories of power holding. Let $\mathbf{L}^t = \left(\{l_{j,0}\}_{j=1}^{N}, ..., \{l_{j,t}\}_{j=1}^{N}\right)$ be the history of labor supplies, and let $\mathbf{C}^t = \left(\{c_{j,0}\}_{j=1}^{N}, ..., \{c_{j,t}\}_{j=1}^{N}\right)$ be the history of allocation rules. A *(complete) history*

[1]Throughout, by "Pareto efficient," we mean "constrained Pareto efficient," but we drop the adjective "constrained" to simplify the terminology.

of this game ("history" for short) at time t is

$$\omega^t = \left(h^t, \mathbf{C}^{t-1}, \mathbf{L}^{t-1}\right),$$

which describes the history of power holdings, all labor supply decisions, and all allocation rules chosen by groups in power. Let the set of all potential date t histories be denoted by Ω^t. In addition, denote an intermediate-stage (complete) history by

$$\hat{\omega}^t = \left(h^t, \mathbf{C}^{t-1}, \mathbf{L}^{t}\right),$$

and denote the set of intermediate-stage full histories by $\hat{\Omega}^t$. The difference between ω and $\hat{\omega}$ lies in the fact that the former does not contain information on labor supplies at time t, while the latter does. The latter history will be relevant at the intermediate stage where the individual in power chooses the allocation rule.

We can now define strategies as follows. First define the following sequence of mappings $\hat{l} = \left(\hat{l}^0, \hat{l}^1, ..., \hat{l}^t, ...\right)$ and $\hat{C} = \left(\hat{C}^0, \hat{C}^1, ..., \hat{C}^t, ...\right)$, where

$$\hat{l}^t : \Omega^t \to \left[0, \bar{l}\right]$$

determines the level of labor a party will supply for every given history $\omega^t \in \Omega^t$, and

$$\hat{C}^t : \hat{\Omega}^t \to \mathbb{R}_+^N$$

a sequence of allocation rules, which a party would choose, if it were in power, for every given intermediate-stage history $\hat{\omega}^t \in \hat{\Omega}^t$, such that $\hat{C}$ satisfies the feasibility constraint (23.7). A date t strategy for party j is $\sigma_j^t = \left(\hat{l}^t, \hat{C}^t\right)$. Denote the set of date t strategies by Σ^t. A strategy for party j is $\sigma_j = \left(\left\{\sigma_j^t\right\} : t = 0, 1, ...\right)$ and the set of strategies is denoted by Σ. Denote the expected utility of party j at time t as a function of its own and others' strategies given history ω^t and intermediate-stage history $\hat{\omega}^t$ by

$$U_j\left(\sigma_j, \sigma_{-j} \mid \omega^t, \hat{\omega}^t\right).$$

We next define various concepts of equilibria which we use throughout the paper.

Definition 23.1. *A* subgame perfect equilibrium (SPE) *is a collection of strategies* $\sigma^* = \left(\left[\sigma_j^t\right] : j = 1, ..., N,\ t = 0, 1, ...\right)$ *such that* σ_j^* *is best response to* σ_{-j}^* *for all* $\left(\omega^t, \hat{\omega}^t\right) \in \Omega^t \times \hat{\Omega}^t$ *and for all* j*, i.e.,* $U_j\left(\sigma_j^*, \sigma_{-j}^* \mid \omega^t, \hat{\omega}^t\right) \geq U_j\left(\sigma_j, \sigma_{-j}^* \mid \omega^t, \hat{\omega}^t\right)$ *for all* $\sigma_j \in \Sigma$*, for all* $\left(\omega^t, \hat{\omega}^t\right) \in \Omega^t \times \hat{\Omega}^t$*, for all* $t = 0, 1, ...$ *and for all* $j \in \mathcal{N}$*.*

Definition 23.2. *A* Pareto efficient perfect equilibrium *at time* t *(following history* ω^t*),* σ^{**}*, is a collection of strategies that form an SPE such that there does not exist another SPE* σ^{***}*, whereby* $U_j\left(\sigma_j^{***}, \sigma_{-j}^{***} \mid \omega^t, \hat{\omega}^t\right) \geq U_j\left(\sigma_j^{**}, \sigma_{-j}^{**} \mid \omega^t, \hat{\omega}^t\right)$ *for all* $\hat{\omega}^t \in \hat{\Omega}^t$ *and for all* $j \in \mathcal{N}$*, with at least one strict inequality.*

We will also refer to *Pareto efficient allocations* as the equilibrium-path allocations that result from a Pareto efficient perfect equilibrium. To characterize Pareto efficient allocations, we will first determine the worst subgame perfect equilibrium, which will be used as a threat against deviations from equilibrium strategies. These are defined next. We write $j = \mathbf{j}\left(h^t\right)$ or $j = \mathbf{j}\left(\omega^t\right)$ if party j is in power at time t according to history (of power holdings) h^t or according to (complete) history $\omega^t \in \Omega^t$. We also use the notation $h^t \in H_j^t$ whenever $j = \mathbf{j}\left(h^t\right)$. A *worst SPE for party* j at time t following history ω^t where σ^W is a collection of strategies that form a SPE such that there does not exist another SPE σ^{***} such that $U_j\left(\sigma_j^{***}, \sigma_{-j}^{***} \mid \omega^t, \hat{\omega}^t\right) < U_j\left(\sigma_j^W, \sigma_{-j}^W \mid \omega^t, \hat{\omega}^t\right)$ for all $\hat{\omega}^t \in \hat{\Omega}^t$.

23.2.4. Preliminary Results. The next lemma describes the worst subgame perfect equilibrium. In that equilibrium, all parties that are not in power in any given period supply zero labor and receive zero consumption, while the party in power supplies labor and consumes all output to maximize its per period utility in such a way that marginal utility from consumption is equated with marginal disutility of labor.

Lemma 23.4. *Suppose Assumption 23.1 holds. The worst SPE for any party j'' is given by the collection of strategies σ^W such that for all $j \neq \mathbf{j}\left(h^t\right)$: $l_j^t\left(\omega^t\right) = 0$ for all $\omega^t \in \Omega^t$, and for $j' = \mathbf{j}\left(h^t\right)$ $l_{j'}^t\left(\omega^t\right) = \tilde{l}_{j'}$ for all $\omega^t \in \Omega^t$ where $\tilde{l}_{j'}$ is a solution to*

$$\frac{\partial u_{j'}(\tilde{l}_{j'}, \tilde{l}_{j'})}{\partial c} = -\frac{\partial u_{j'}(\tilde{l}_{j'}, \tilde{l}_{j'})}{\partial l} \tag{23.8}$$

and $c_j^t\left(\hat{\omega}^t\right) = 0$ for $j \neq j', c_{j'}^t\left(\hat{\omega}^t\right) = \sum_{i=1}^N l_i^t\left(\hat{\omega}^t\right)$ for all $\hat{\omega}^t \in \hat{\Omega}^t$.

Proof. We first show that σ^W is a best response for each party in all subgames when other parties are playing σ^W. Consider first party j that is not in power (i.e., suppose that party $j' \neq j$ is in power) at history ω^t. Consider strategy $\sigma_{j,t}$ for party j that deviates from σ_j^W at time t, and then coincides with σ_j^W at all subsequent dates (following all histories). By the one step ahead deviation principle, if σ_j^W is not a best response for party j, then there exists such a strategy $\sigma_{j,t}$ that will give higher utility to this party. Note, first that given σ_{-j}^W, for any $\sigma_{j,t}$, party j will always receive zero consumption (i.e., $c_j\left(\sigma_j, \sigma_{-j}^W \mid \omega^t, \hat{\omega}^t\right) = 0$), and moreover under σ^W, this has no effect on the continuation value of party j. Therefore, at such an history, we have

$$\begin{aligned} U_j\left(\sigma_{j,t}, \sigma_{-j}^W \mid \omega^t, \hat{\omega}^t\right) &= u_j(0, l_{j,t}\left(\sigma_{j,t}\right)) + \beta \mathbb{E}\left[U_j\left(\sigma_{j,t}, \sigma_{-j}^W \mid \omega^{t+1}, \hat{\omega}^{t+1}\right) \mid \omega^t\right] \\ &\leq \beta \mathbb{E}\left[U_j\left(\sigma_j^W, \sigma_{-j}^W \mid \omega^{t+1}, \hat{\omega}^{t+1}\right) \mid \omega^t\right], \end{aligned}$$

for any such $\sigma_{j,t}$, where $l_{j,t}(\sigma_{j,t})$ is the labor supply of party j at this history under the alternative strategy $\sigma_{j,t}$, and $\mathbb{E}\left[U_j\left(\sigma_{j,s}, \sigma_{-j}^W \mid \omega^{t+1}, \hat{\omega}^{t+1}\right) \mid \omega^t\right]$ is the continuation value of this party from date $t+1$ onwards, with the expectation taken over histories determining power switches given current history ω^t. The second line follows in view of the fact that since $u_j(0,0) = 0$, we have $u_j(0, l_{j,t}(\sigma_{j,t})) \leq 0$, and since under σ^Wany change in behavior at t has no effect on future play and $\sigma_{j,t}$ coincides with σ_j^W from time $t+1$ onwards, $\mathbb{E}\left[U_j\left(\sigma_{j,t}, \sigma_{-j}^W \mid \omega^{t+1}, \hat{\omega}^{t+1}\right) \mid \omega^t\right] = \mathbb{E}\left[U_j\left(\sigma_j^W, \sigma_{-j}^W \mid \omega^{t+1}, \hat{\omega}^{t+1}\right) \mid \omega^t\right]$. This establishes that there is no profitable deviations from σ_j^W for any j not in power.

Next consider party j in power at history ω^t. Under σ^W, $l_{j',t} = c_{j',t} = 0$ for all $j' \neq j$, and thus $l_{j,t} = c_{j,t}$. Consider again strategy $\sigma_{j,t}$ for party j that deviates from σ_j^W at time t, and then coincides with σ_j^W at all subsequent dates. Then, using similar notation, we have

$$\begin{aligned} U_j\left(\sigma_{j,t}, \sigma_{-j}^W \mid \omega^t, \hat{\omega}^t\right) &= u_j(c_{j,t}(\sigma_{j,t}), l_{j,t}(\sigma_{j,t})) + \beta\mathbb{E}\left[U_j\left(\sigma_{j,t}, \sigma_{-j}^W \mid \omega^{t+1}, \hat{\omega}^{t+1}\right) \mid \omega^t\right] \\ &\leq u_j(c_{j,t}(\sigma_j^W), l_{j,t}(\sigma_j^W)) + \beta\mathbb{E}\left[U_j\left(\sigma_j^W, \sigma_{-j}^W \mid \omega^{t+1}, \hat{\omega}^{t+1}\right) \mid \omega^t\right], \end{aligned}$$

for any such $\sigma_{j,t}$, where $l_{j,t}(\sigma_{j,t})$ and $c_{j,t}(\sigma_{j,t})$ are the labor supply and consumption of party j at this history under strategy $\sigma_{j,t}$. The second line follows in view of the fact that σ_j^W satisfies (23.8), and thus $u_j(c_{j,t}\left(\sigma_j^W\right), l_{j,t}\left(\sigma_j^W\right)) = u_j(\tilde{l}, \tilde{l}) \geq u_j(c_{j,t}(\sigma_{j,t}), l_{j,t}(\sigma_{j,t}))$ for any $\sigma_{j,t}$, and again because $\mathbb{E}\left[U_j\left(\sigma_{j,t}, \sigma_{-j}^W \mid \omega^{t+1}, \hat{\omega}^{t+1}\right) \mid \omega^t\right] = \mathbb{E}\left[U_j\left(\sigma_j^W, \sigma_{-j}^W \mid \omega^{t+1}, \hat{\omega}^{t+1}\right) \mid \omega^t\right]$ (from the fact that under σ^W the current deviation by party j has no effect on future play and $\sigma_{j,t}$ coincides with σ_j^W from time $t+1$ onwards). This establishes that there is no profitable deviations from σ_j^W for the party in power. Therefore, σ^W is a SPE. The proof is completed by showing that σ^W is also the worst SPE for any party j. To see this, suppose that all $j' \neq j$ choose strategy σ_{-j}^M to minimize the payoff of j. Since power switches are exogenous, party j can guarantee itself $u_j(\tilde{l}, \tilde{l})$ whenever it is in power and $u_j(c_{j,t}\left(\sigma_j^W, \sigma_{-j}^M\right), l_{j,t}\left(\sigma_j^W, \sigma_{-j}^M\right)) = 0$ whenever it is not in power. Therefore,

$$U_j\left(\sigma_j^W, \sigma_{-j}^M \mid \omega^t, \hat{\omega}^t\right) \geq U_j\left(\sigma_j^W, \sigma_{-j}^W \mid \omega^t, \hat{\omega}^t\right)$$

for any σ_{-j}^M, and thus σ^W is the worst SPE. Intuitively, the worst equilibrium involves all groups other than the one in power supplying zero labor, which minimizes the utility of the group in power. The labor supply decisions constitute a best response, since following a deviation all output is expropriated by whichever group is currently in power. Also note that the same equilibrium is the worst equilibrium for all parties. □

We denote $V_j^W(h^t)$ to be the expected payoff of party j from period $t+1$ on, conditional on history h^t. Note that generally such utility differs based on the history: $V_j^W(h^t) \neq V_j^W(\hat{h}^t)$

for $h_t \neq \hat{h}_t$. The reason is that the identity of party in power in period t determines the probability of party j being in power in $t+1$. When m satisfies Assumption 23.2, it can be further simplified, since $V_j^W(h^t)$ depends only on the identify of party in power h_t.

PROPOSITION 23.4. *Suppose Assumptions 23.1 and 23.2 hold. Then, an outcome of any Pareto efficient perfect equilibrium is a solution to the following maximization problem for all* $h^t \in \omega^t, \hat{\omega}^t$:

$$\max_{\{c_j(h^t),l_j(h^t)\}_{j=1,..,N;\ h^t}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\sum_{j=1}^{N} \alpha_j u_j(c_j\left(h^t\right), l_j\left(h^t\right)) \right] \tag{23.9}$$

subject to, for all h^t,

$$\sum_{j=1}^{N} c_j\left(h^t\right) \leq \sum_{j=1}^{N} l_j\left(h^t\right), \tag{23.10}$$

$$\mathbb{E}_t \sum_{s=0}^{\infty} \beta^s u_j\left(c_j\left(h^{t+s}\right), l_j\left(h^{t+s}\right)\right) \geq \beta V_j^W(\mathbf{j}(h^t)) \text{ for all } j \neq \mathbf{j}(h^t), \tag{23.11}$$

and

$$\mathbb{E}_t \sum_{s=0}^{\infty} \beta^s u_{j'}\left(c_{j'}\left(h^{t+s}\right), l_{j'}\left(h^{t+s}\right)\right) \geq \max_{\tilde{l}\geq 0} u_{j'}\left(\sum_{j\neq j'} l_j\left(h^t\right) + \tilde{l}\ , \tilde{l}\right) + \beta V_{j'}^W(j') \text{ for } j' = \mathbf{j}(h^t), \tag{23.12}$$

for some Pareto weights vector $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)$.

PROOF. We start by showing that (23.10)-(23.12) are necesary and sufficient conditions for any allocation $\{c_j(h), l_j(h)\}_{j=1}^N$ that is an outcome of some SPE. First, we show that any allocation $\{c_j(h), l_j(h)\}_{j=1}^N$ that satisfies (23.10)-(23.12) is an outcome of some SPE. For any history ω^t with $h^t \in \omega^t$ let $\sigma^*(\omega^t) = \left\{l_j\left(h^t\right)\right\}_{j=1}^N$ if $\omega^t = \left(h^t, \left\{c_j\left(h^{t-1}\right), l_j\left(h^{t-1}\right)\right\}_{j=1}^N\right)$, and $\sigma^*\left(\omega^t\right) = \sigma^W$ otherwise, and, analogously, $\sigma^*(\hat{\omega}^t) = \left\{c_j\left(h^t\right)\right\}_{j=1}^N$ if $\hat{\omega}^t = \left(h^t, \left\{c_j\left(h^{t-1}\right), l_j\left(h^t\right)\right\}_{j=1}^N\right)$, and $\sigma^*\left(\tilde{\omega}^t\right) = \sigma^W$ otherwise. For any $j \neq \mathbf{j}(h^t)$ if $\sigma_j(\omega^t) \neq \sigma_j^*(\omega^t)$,

$$\begin{aligned} U_j(\sigma_j, \sigma_{-j}^*|\omega^t) &\leq u_j(0,0) + \beta V_j^W(\mathbf{j}(h^t)) \\ &\leq U_j(\sigma_j^*, \sigma_{-j}^*|\omega^t), \end{aligned}$$

where the last inequality follows from (23.11) and that fact that $u_j(0,0) = 0$. Moreover, for $j' = \mathbf{j}(h^t)$, any $\sigma_{j'} \neq \sigma^*_{j'}$ implies

$$\begin{aligned} U'_{j'}(\sigma_{j'}, \sigma^*_{-j'}|\omega^t, \tilde{\omega}^t) &\leq \max_{\tilde{l}\geq 0} u_{j'}\left(\sum_{j\neq j'} l_j\left(h^t\right) + \tilde{l}\ , \tilde{l}\right) + \beta V^W_{j'}(j') \\ &\leq U_{j'}(\sigma^*_{j'}, \sigma^*_{-j'}|\omega^t, \tilde{\omega}^t). \end{aligned}$$

Therefore, σ^* is an equilibrium.

The necessity of (23.10)-(23.12) is straightforward. Condition (23.10) is feasibility constraint. In any equilibrium σ^* for $j \neq \mathbf{j}(h^t)$ for all $\sigma_j \neq \sigma^*_j$

$$U_j(\sigma^*_j, \sigma^*_{-j}|\omega^t) \geq U_j(\sigma_j, \sigma^*_{-j}|\omega^t)$$

This implies in particular that

$$\begin{aligned} U_j(\sigma^*_j, \sigma^*_{-j}|\omega^t) &\geq u_j\left(c_j\left(\sigma_{j,t}\right), l_j\left(\sigma_{j,t}\right)\right) + \beta\mathbb{E}_t U_j(\sigma_j, \sigma^*_{-j}|\omega^{t+1}) \\ &\geq u_j\left(0, l_j\left(\sigma_{j,t}\right)\right) + \beta V^W_j(\mathbf{j}(h^t)) \\ &= u_j\left(0,0\right) + \beta V^W_j(\mathbf{j}(h^t)). \end{aligned}$$

Since $u_j\left(0,0\right) = 0$, it implies (23.11). Similarly, for $j' = \mathbf{j}(h^t)$ best response implies

$$\begin{aligned} U_{j'}(\sigma^*_{j'}, \sigma^*_{-j'}|\omega^t, \tilde{\omega}^t) &\geq \max_{\tilde{l}\geq 0} u_{j'}\left(\sum_{j\neq j'} l_j\left(h^t\right) + \tilde{l}\ , \tilde{l}\right) + \beta\mathbb{E}_t U_j(\sigma_{j'}, \sigma^*_{-j'}|\omega^{t+1}) \\ &\geq \max_{\tilde{l}} u_{j'}\left(\sum_{j\neq j'} l_j\left(h^t\right) + \tilde{l}\ , \tilde{l}\right) + \beta V^W_{j'}(j') \end{aligned}$$

which is condition (23.12).

To see that σ^* is a Pareto efficient equilibrium, suppose there is any other equilibrium σ^{**} that Pareto dominates σ^*. Since σ^{**} is a SPE,

$$U_j(\sigma^{**}_j, \sigma^{**}_{-j}|\omega^t) \geq \beta V^W_j(h^t)$$

for all $j \neq \mathbf{j}(h^t)$ and

$$U_j(\sigma^{**}_j, \sigma^{**}_{-j}|\omega^t) \geq \max_{\tilde{l}\geq 0} u_{j'}\left(\sum_{j\neq j'} l_j\left(\omega^t\right) + \tilde{l}\ , \tilde{l}\right) + \beta V^W_{j'}(j').$$

Therefore, the outcome of σ^{**} must satisfy (23.10)-(23.12). But then the value of (23.9) would be higher under the outcome of σ^{**} than under σ^*, a contradiction. □

To simplify the notation, we define

$$V_j(h^{t-1}) \equiv \mathbb{E}\left\{\sum_{s=0}^{\infty} \beta^s u_{j'}\left(c_{j'}\left(h^{t+s}\right), l_{j'}\left(h^{t+s}\right)\right)|h^{t-1}\right\}$$

$$V_j[h^{t-1}, i] \equiv \mathbb{E}\left\{\sum_{s=0}^{\infty} \beta^s u_{j'}\left(c_{j'}\left(h^{t+s}\right), l_{j'}\left(h^{t+s}\right)\right) | h^{t-1}, \mathbf{j}(h^t) = i\right\}$$

The difference between $V_j(h^{t-1})$ and $V_j[h^{t-1}, i]$ is that the former denotes expected lifetime utility of party j in period t before the uncertainty which party is in power in that period is realized, while the latter denotes the expected lifetime utility after realization of this uncertainty. From the above definition and Assumption 23.2,

$$V_j(h^{t-1}) = \sum_{j'=1}^{N} m(j'|h_{t-1}) V_j[h^{t-1}, j'].$$

Proposition 23.4 implies that in order to characterize the entire set of Pareto efficient perfect equilibria, we can restrict attention to strategies that follow a particular prescribed equilibrium play, with the punishment phase given by σ^W. Notice, however, that this proposition applies to Pareto efficient outcomes, *not* to the strategies that individuals use in order to support these outcomes. These strategies *must be* conditioned on information that is not contained in the history of power holdings, h^t, since individuals need to switch to the worst subgame perfect equilibrium in case there is any deviation from the implicitly-agreed action profile. This information is naturally contained in ω^t. Therefore, to describe the subgame perfect equilibrium strategies we need to condition on the full histories ω^t.

The maximization (23.9) subject to (23.10), (23.11), and (23.12) is a potentially non-convex optimization problem, because (23.12) defines a non-convex constraint set. This implies that randomizations may improve the value of the program (see, for example, Prescott and Townsend, 1984 a,b). Randomizations can be allowed by either considering correlated equilibria rather than subgame perfect equilibria, or alternatively, by assuming that there is a commonly observed randomization device on which all individuals can coordinate their actions. In the Appendix, we will formulate an extended problem by introducing a commonly-observed, independently and identically distributed random variable, which all individual strategies can be conditioned upon. We will show that this does not change the basic structure of the problem and in fact there will be randomizations over at most two points at any date, and the history of past randomizations will not play any role in the characterization of Pareto efficient allocations. Since introducing randomizations complicates the notation considerably, in the text we do not consider randomizations (thus implicitly assuming that the problem is convex for the relevant parameters). The equivalents of the main results are stated in the Appendix for the case with randomizations.

We next present our main characterization result, which shows that the solution to the maximization problem in Proposition 23.4 can be represented recursively.

23.2.5. Recursive Characterization. Let us first define $M\left(h^{t+s} \mid h^t\right)$ to be the (conditional) probability of history h^{t+s} at time $t+s$ given history h^t at time t according to the Markov process $m(j \mid j')$. Moreover, define $P\left(h^t\right)$ be the set of all possible date $t+s$ histories for $s \geq 1$ that can follow history h^t. We write $M\left(h^{t+s} \mid h^{\varnothing}\right)$ for the unconditional probability of history h^{t+s}.

First note that the maximization problem in Proposition 23.4 can be written in Lagrangian form as follows:

$$\max_{\{c_j(h^t),l_j(h^t)\}_{j=1,..,N}} \mathcal{L}' = \sum_{t=0}^{\infty}\sum_{h^t}\beta^t M\left(h^t \mid h^{\varnothing}\right)\left[\sum_{j=1}^{N}\alpha_j u_j\left(c_j\left(h^t\right), l_j\left(h^t\right)\right)\right]$$
$$+\sum_{t=0}^{\infty}\sum_{h^t}\beta^t M\left(h^t \mid h^{\varnothing}\right)\lambda_j\left(h^t\right)\times$$
$$\left[\sum_{s=t}^{\infty}\beta^{s-t}\sum_{h^s}M\left(h^s \mid h^t\right)u_j\left(c_j\left(h^s\right), l_j\left(h^s\right)\right) - v_j\left(\sum_{i\neq j}l_i\left(h^t\right)\right) - \beta V_j^W(j)\right] \tag{23.13}$$

subject to (23.10) and (23.11), where $\beta^t M\left(h^t \mid h^0\right)\lambda_j\left(h^t\right)$ is the Lagrange multiplier on the sustainability constraint, (23.12), for party j for history h^t, and $v_j(Y) \equiv \max_{\tilde{l}} u_j\left(Y+\tilde{l}\,,\tilde{l}\right)$. The restriction that $h^t \in P\left(h^{t-1}\right)$ is implicit in this expression.

The proof of Theorem 23.5 establishes that after history h^{t-1}, this Lagrangian is equivalent to

$$\max_{\{c_j(h^t),l_j(h^t)\}_{j=1,..,N;\ h^t}} \mathcal{L} = \sum_{s=t}^{\infty}\sum_{h^s}\beta^s M\left(h^s \mid h^t\right)\left[\sum_{j=1}^{N}\left(\alpha_j+\mu_j\left(h^{t-1}\right)\right)u_j\left(c_j\left(h^s\right), l_j\left(h^s\right)\right)\right]$$
$$+\sum_{s=t}^{\infty}\sum_{h^s}\beta^s M\left(h^s \mid h^t\right)\times$$
$$\sum_{j=1}^{N}\lambda_j\left(h^s\right)\left(\sum_{s'=s}^{\infty}\sum_{h^{s'}}\beta^{s'-s}M\left(h^{s'} \mid h^s\right)u_j\left(c_j\left(h^{s'}\right), l_j\left(h^{s'}\right)\right) - v_j\left(\sum_{i\neq j}l_i\left(h^s\right)\right) - \beta V_j^W(j)\right), \tag{23.14}$$

subject to (23.10) and (23.11), with μ_j's defined recursively as:

$$\mu_j\left(h^t\right) = \mu_j\left(h^{t-1}\right) + \lambda_j\left(h^t\right)$$

with the normalization $\mu_j\left(h^{\varnothing}\right) = 0$ for all $j \in \mathcal{N}$.

The most important implication of the formulation in (23.14) is that for any $h^t \in P\left(h^{t-1}\right)$, the numbers

$$\alpha_j\left(h^{t-1}\right) \equiv \frac{\alpha_j+\mu_j\left(h^{t-1}\right)}{\sum_{j'=1}^{N}\left(\alpha_{j'}+\mu_{j'}\left(h^{t-1}\right)\right)} \tag{23.15}$$

can be interpreted as *updated Pareto weights*. Therefore, after history h^{t-1}, the problem is equivalent to maximizing the sum of utilities with these weights (subject to the relevant constraints). The problem of maximizing (23.14) is equivalent to choosing current consumption and labor supply levels for each group and also updated Pareto weights $\{\alpha_j\}_{j=1}^N$.

In addition, (23.14) has the attractive feature that $\mu_j\left(h^t\right) - \mu_j\left(h^{t-1}\right) = 0$ whenever $j \neq \mathbf{j}\left(h^t\right)$, i.e., whenever group j is not in power. This is because there is no sustainability constraint for a group that is not in power. This also implies that in what follows, we can drop the subscript j and refer to $\lambda\left(h^t\right)$ rather than $\lambda_j\left(h^t\right)$, since the information on which group is in power is already incorporated in h^t.

This analysis establishes the following characterization result:

PROPOSITION 23.5. *Suppose Assumptions 23.1 and 23.2 hold. Then the constrained efficient allocation has a quasi-Markovian structure whereby consumption and labor allocations* $\left\{c_j\left(h^t\right), l_j\left(h^t\right)\right\}_{j=1,..,N;\ h^t}$ *depend only on* $\mathbf{s} \equiv \left(\{\alpha_j(h)\}_{j=1}^N, \mathbf{j}(h)\right)$, *i.e., only on updated weights and the identity of the group in power.*

PROOF. The proof of this theorem builds on the representation suggested by Marcet and Marimon (1998). In particular, observe that for any $T \geq 0$, we have

$$\sum_{s=0}^{T}\sum_{h^s}\beta^s M\left(h^s \mid h^{\varnothing}\right)\lambda_j\left(h^s\right)\sum_{s'=s}^{T}\sum_{h^{s'}}\beta^{s'-s}M\left(h^{s'} \mid h^s\right)u_j\left(c_j\left(h^{s'}\right), l_j\left(h^{s'}\right)\right) \tag{23.16}$$
$$= \sum_{s=0}^{T}\beta^s M\left(h^s \mid h^{\varnothing}\right)\mu_j\left(h^s\right)u_j\left(c_j\left(h^s\right), l_j\left(h^s\right)\right)$$

where $\mu_j\left(h^s\right) = \mu_j\left(h^{s-1}\right) + \lambda_j\left(h^s\right)$ for $h^s \in P\left(h^{s-1}\right)$ with the initial $\mu_j\left(h^{\varnothing}\right) = 0$ for all j. Substituting (23.16) in $\mathcal{L}'$ in (23.13), we obtain that $\mathcal{L}'$ for any h^{t-1}, can be expressed as

$$\max_{\{c_j(h),l_j(h)\}_{j=1}^N}\mathcal{L}'' = \sum_{s=t}^{\infty}\sum_{h^s}\beta^s M\left(h^s \mid h^{t-1}\right)\left[\sum_{j=1}^{N}\left(\alpha_j + \mu_j\left(h^{t-1}\right)\right)u_j\left(c_j\left(h^s\right), l_j\left(h^s\right)\right)\right]$$
$$+\sum_{s=t}^{\infty}\sum_{h^s}\beta^s M\left(h^s \mid h^{t-1}\right)\sum_{j=1}^{N}\lambda_j\left(h^s\right)\left(\begin{array}{c}\sum_{s'=s}^{\infty}\sum_{h^{s'}}\beta^{s'-s}M\left(h^{s'} \mid h^s\right)u_j\left(c_j\left(h^s\right), l_j\left(h^s\right)\right)\\ -v_j\left(\sum_{i\neq j}l_i\left(h^s\right)\right) - \beta V_j^W(j)\end{array}\right)$$
$$+\sum_{s=0}^{t-1}\sum_{h^s}\beta^s M\left(h^s \mid h^{\varnothing}\right)\sum_{j=1}^{N}\alpha_j u_j\left(c_j\left(h^s\right), l_j\left(h^s\right)\right)$$
$$+\sum_{s=0}^{t-1}\sum_{h^s}\beta^s M\left(h^s \mid h^{\varnothing}\right)\sum_{j=1}^{N}\lambda_j\left(h^s\right)\left(\begin{array}{c}\sum_{s'=s}^{t-1}\sum_{h^{s'}}\beta^{s'}M\left(h^{s'} \mid h^s\right)u_j\left(c_j\left(h^{s'}\right), l_j\left(h^{s'}\right)\right)\\ -v_j\left(\sum_{i\neq j}l_i\left(h^s\right)\right) - \beta V_j^W(j)\end{array}\right).$$

Since after history h^{t-1} has elapsed, all terms in the last two lines are given, maximizing $\mathcal{L}''$ is equivalent to maximizing (23.14).

Given the structure of problem (23.14), the result that optimal allocations only depend on $\{\alpha_j(h)\}_{j=1}^N$ and $\mathbf{j}(h)$ then follows immediately. $\square$

The result in this theorem is intuitive. When the sustainability constraint for the party in power is binding, the discounted value of this party needs to be increased so as to satisfy this constraint. This is typically done by a combination of increasing current and future utility. The latter takes the form of increasing the Pareto weight of the party in power, corresponding to a move along the constraint Pareto efficient frontier.

It is also worth noting that the existence of a recursive formulation for the problem of characterizing the set of Pareto efficient allocations also has an obvious parallel to the general recursive formulation provided by Abreu, Pearce and Stacchetti (1990) for repeated games with imperfect monitoring. Nevertheless, Theorem 23.5 is not a direct corollary of their results, since it establishes that this recursive formulation depends on updated Pareto weights and the identity of the group in power, and shows how these weights can be calculated from past realizations of the history h^t.

Theorem 23.5 allows us to work with a recursive problem, in which we only have to keep track of the identity of the party that is in power and updated Pareto weights. Moreover, the analysis preceding the theorem shows that the Pareto weights are updated following the simple formula (23.15), which is only a function of the Lagrange multiplier on the sustainability constraint of the party in power at time t. The recursive characterization implies that we can express $V_j(h^{t-1})$ and $V_j[h^{t-1}, i]$ as $V_j(\boldsymbol{\alpha})$ and $V_j[\boldsymbol{\alpha}, i]$.

23.2.6. Characterization of Distortions. We now characterize the structure of distortions arising from political economy. Our first result shows that as long as sustainability/political economy constraints are binding, the labor supply of parties that are not in power is distorted *downwards*. There is a positive wedge between their marginal utility of consumption and marginal disutility of labor. In contrast, there is no wedge for the party in power. Recall also that without political economy constraints, in the first best allocations, the distortions are equal to zero.

Proposition 23.6. *Suppose that Assumptions 23.1 and 23.2 hold. Then as long as* $\lambda(h^t) > 0$, *the labor supply of all groups that are not in power, i.e.,* $j \neq \mathbf{j}(h^t)$, *is distorted* downwards*, in the sense that*

$$\frac{\partial u_j\left(c_j\left(h^t\right),l_j\left(h^t\right)\right)}{\partial c} > -\frac{\partial u_j\left(c_j\left(h^t\right),l_j\left(h^t\right)\right)}{\partial l}.$$

The labor supply of a party in power, $j' = \mathbf{j}\left(h^t\right)$, is undistorted, i.e.,

$$\frac{\partial u_{j'}\left(c_{j'}\left(h^t\right),l_{j'}\left(h^t\right)\right)}{\partial c} = -\frac{\partial u_j\left(c_{j'}\left(h^t\right),l_{j'}\left(h^t\right)\right)}{\partial l}.$$

Positive distortions, which are the equivalent of "taxes," discourage labor supply, reducing the amount of output that the group in power can "expropriate" (i.e., allocate to itself as consumption following a deviation). This relaxes the sustainability constraint (23.12). In fact, starting from an allocation with no distortions, a small distortion in labor supply creates a second-order loss. In contrast, as long as the multiplier on the sustainability constraint is positive, this small distortion creates a first-order gain in the objective function, because it enables a reduction in the rents captured by the group in power. This intuition also highlights that the extent of distortions will be closely linked to the Pareto weights given to the group in power. In particular, when α_j is close to 1 and group j is in power, there will be little gain in relaxing the sustainability constraint (23.12). In contrast, the Pareto efficient allocation will attempt to provide fewer rents to group j when α_j is low, and this is only possible by reducing the labor supply of all other groups, thus distorting their labor supplies.

Two immediate but useful corollaries of Proposition 23.6 are as follows:

Corollary 23.1. *The (normalized) Lagrange multiplier on the sustainability constraint (23.12) given history h^t, $\lambda\left(h^t\right)$, is a measure of distortions.*

This corollary follows immediately from Proposition 23.6, which shows that the wedges between the marginal utility of consumption and the marginal disutility of labor are directly related to $\lambda\left(h^t\right)$. This corollary is useful as it will enable us to link the level and behavior of distortions to the behavior of the Lagrange multiplier $\lambda\left(h^t\right)$.

A related implication of Proposition 23.6 is that constrained Pareto efficient allocations will be "first-best" if and only if the Lagrange multipliers associated with all sustainability constraints are equal to zero (so that there are no distortions in a first-best allocation).

Corollary 23.2. *A first-best allocation starting at history h^t involves $\lambda\left(h^{t+s}\right) = 0$ for all $h^{t+s} \in P\left(h^t\right)$.*

23.2.7. Dynamics of Distortions. Proposition 23.6 states that when the Lagrange multipliers are positive, the allocations are distorted. We now study the evolution of the Largrange multipliers and distortions resulting from the sustainability constraints.

Our first result in this subsection is an immediate implication of the recursive formulation in Proposition 23.5, but it will play an important role in our results. The lemma that follows shows that if a group is in power today, then in the next period its updated Pareto weight must be weakly higher than today.

Lemma 23.5. *If* $\mathbf{j}\left(h^{t+1}\right)=j$, *then* $\alpha_j\left(h^{t+1}\right)\geq\alpha_j\left(h^t\right)$.

Proof. This follows immediately from equation (23.15) observing that if $\mathbf{j}\left(h^{t+1}\right)=j'$, then $\mu_j\left(h^{t+1}\right)-\mu_j\left(h^t\right)=0$ for all $j\neq j'$. □

This lemma implies that as long as party j remains in power, its Pareto weight is increasing. We assume the following condition holds.

Condition 23.1. *For all* j, *if* $\alpha_j=1$, *then*

$$V_j[\boldsymbol{\alpha},j]>v_j\left(\sum_{i\neq j}\tilde{l}_i\right)+\beta V_j^W(j)$$

where $\tilde{l}_i$ *is a labor supply of party* i *in equilibrium in the state* $(\boldsymbol{\alpha},j)$.

Although Assumption 23.1 is stated in terms of the endogenous objects, it is easy to see that it will be satisfied unless the discount factor is too low. An implication of Lemma 23.5 is that when a particular party remains in power for sufficiently long, the Lagrange multipliers on the sustainability constraints and distortions begin declining.

Proposition 23.7. *Suppose Assumptions 23.1, 23.2 and Condition 23.1 hold. Then for any* $\varepsilon>0$ *there exists* $K\in\mathbb{N}$ *such that if* $\mathbf{j}\left(h^t\right)=\mathbf{j}\left(h^{t+k}\right)=j$ *for* $k=1,...,K$, *then* $\lambda\left(h^{t+m}\right)<\varepsilon$ *for all* $m\geq K$.

Proof. Fix j and $h^\infty\in H^\infty$, and suppose that $\mathbf{j}\left(h^t\right)=\mathbf{j}\left(h^{t+k}\right)=j$ for all $k\in\mathbb{N}$. Then from Lemma 23.5, $\left\{\alpha_j\left(h^{t+k}\right)\right\}_{k=0}^\infty$ is a non-decreasing sequence, and moreover, by construction $\alpha_j\left(h^{t+k}\right)\in[0,1]$ for each j and all h^{t+k}. Consequently, $\alpha_j\left(h^{t+k}\right)\to\bar{\alpha}_j$. Next note that $\bar{\alpha}_j<1$. To see this note that the inspection of the maximization problem (23.14) shows that when $\bar{\alpha}_j=1$, the constraint (23.12) is slack. Since the objective function is continuous in the vector of Pareto weights $\boldsymbol{\alpha}$, this implies that for $\alpha_j=1-\varepsilon$ with ε sufficiently small, the constraint is also slack and $\lambda_j\left(h^{t+k}\right)=0$. This implies that there exists $\varepsilon_j>0$ such that starting with $\alpha_j\left(h^t\right)<1-\varepsilon_j$, we cannot have $\alpha_j\left(h^{t+1}\right)=1$ for any h^t, since from equation (23.15) this would imply that $\lambda_j\left(h^{t+k}\right)=\infty$, which is not possible. Next equation (23.15) also implies that if $\lambda\left(h^{t+k}\right)=0$, then $\alpha_j\left(h^{t+k}\right)$ will remain constant (since $\mathbf{j}\left(h^t\right)=\mathbf{j}\left(h^{t+k}\right)=j$ for all $k\in\mathbb{N}$). Therefore, $\alpha_j\left(h^{t+k}\right)\to\bar{\alpha}_j<1$. Next, inspection of equation (23.15) shows that as $\alpha_j\left(h^{t+k}\right) \to \bar{\alpha}_j$, we have $\mu_j\left(h^{t+k}\right) - \mu_j\left(h^{t+k-1}\right) \to 0$ and thus $\lambda\left(h^{t+k}\right) \to 0$ (by virtue of the fact that $\bar{\alpha}_j < 1$). $\square$

Intuitively, along the path in which a particular group remains in power for a long time, distortions ultimately decline. This is because as a particular group remains in power for a long time, its Pareto weight increases sufficiently and the allocations do not need to be distorted to satisfy the sustainability constraint.

The major result in Proposition 23.7 is that as a particular group remains in power longer, distortions eventually decline. Intuitively, this follows from the fact that when the group in power has a higher updated Pareto weight, then there is no need to distort allocations as much. In the limit, if the group in power had a weight equal to 1, then the Pareto efficient allocation would give all consumption to individuals from this group, and therefore, there would be no reason to distort the labor supply of other groups in order to relax the sustainability constraint and reduce rents to this group. Put differently, recall that distortions (and inefficiencies) arise because the group in power does not have a sufficiently high Pareto weight and the Pareto efficient allocation reduces its consumption by reducing total output and thus relaxing its sustainability constraint. As a group remains in power for longer, its updated Pareto weight increases and as a result, there is less need for this type of distortions. This reflects itself in a reduction in the Lagrange multiplier associated with the sustainability constraint.

Proposition 23.7 also suggests a result reminiscent to the conjecture discussed in the Introduction; greater political stability translates into lower inefficiencies and better public policies. This conclusion does not follow from the theorem, however. The theorem is for a given sample path (holding the Markov process regulating power switches fixed). The conjecture linking political stability to efficient public policy refers to a comparison of the extent of distortions for different underlying Markov processes governing power switches.

Proposition 23.7 does not answer the question of whether distortions will ultimately disappear—i.e., whether we will have $\lambda\left(h^t\right) = 0$ for all h^t after some date. More formally, we call any allocation $\{c_j^*, l_j^*\}_j$ a *sustainable first-best allocation* if $\{c_j^*, l_j^*\}_j$ is a first best allocation that satisfies

$$\frac{1}{1-\beta} u_j\left(c_j^*, l_j^*\right) \geq v_j\left(\sum_{i \neq j} l_i^*\right) + \beta V_j^W(j) \text{ for all } j \tag{23.17}$$

The next theorem addresses this question.

Proposition 23.8. *Suppose that Assumptions 23.1 and 23.2 hold. Then there exists $\bar{\beta}$, with $0 < \bar{\beta} < 1$ such that*

Theorem 23.1. (a) *For all $\beta \geq \bar{\beta}$, there is some first-best allocation that is sustainable;*

(b) *For all $\beta < \bar{\beta}$, no first-best allocation is sustainable, and $\{c_j(h^t), l_j(h^t)\}_j$ converges to an invariant non-degenerate distribution F.*

This proposition therefore shows that for high discount factors, i.e., $\beta \geq \bar{\beta}$, first-best allocations will be sustainable. However, when $\beta < \bar{\beta}$, then there will be permanent fluctuations in consumption and labor supply levels as political power fluctuates between different parties. The invariant distribution can be quite complex in general, with the history of power holdings shaping consumption and labor supply levels of each group.

This proposition does not answer the question of whether first-best allocations, when they are sustainable, will be ultimately reached. We address this question for the case of two parties in the next subsection.

23.2.8. The Case of Two Parties. Let us now focus on an economy with two parties (rather than N parties as we have done so far). We also specialize utility function to be quasi-linear. Under these conditions, we show that when there exists a sustainable first-best allocation (i.e., an undistorted allocations for some Pareto weights), the equilibrium will necessarily converge to a point in the set of first-best allocations. More specifically, starting with any Pareto weights, the allocations ultimately converge to undistorted allocations. We also impose the following assumption on the preferences.

Assumption 23.3. ***(quasi-linearity)*** *The instantaneous utility of each party j satisfies*

$$u_j\left(c_j - h_j\left(l_j\right)\right)$$

with the normalization

$$h_j'\left(1\right) = 1. \tag{23.18}$$

Assumption 23.3 implies that there are no income effects in labor supply. Consequently, when there are no distortions, the level of labor supply by each group will be constant, and given the normalization in (23.18), this labor supply level will be equal to 1.

Under Assumption 23.3, Theorem 23.7 implies that as a particular group remains in power for a sufficiently long time, overall output in the economy will increase (since there are no income effects, lower distortions translate into higher labor supply levels). The absence of income effects also simplifies the analysis and dynamics, which is our main focus in this subsection.

We now state and prove three lemmas, which together will enable us to establish the main result in this subsection, Proposition 23.9.

We first show that the party with a higher Pareto weight will receive higher value.

Lemma 23.6. *For any two vectors of Pareto weights α, α', if $\alpha_i > \alpha_i'$ then $V_i[\alpha, j] \geq V_i[\alpha', j]$ for $j \in \{1, 2\}$.*

Proof. Without loss of generality, let $i = 1$. Optimality implies

$$\alpha_1 V_1[\alpha, j] + \alpha_2 V_2[\alpha, j] \geq \alpha_1 V_1\left[\alpha', j\right] + \alpha_2 V_2\left[\alpha', j\right]$$

and

$$\alpha_1' V_1\left[\alpha', j\right] + \alpha_2' V_2\left[\alpha', j\right] \geq \alpha_1' V_1[\alpha, j] + \alpha_2' V_2[\alpha, j].$$

These conditions imply that

$$\left(\alpha_1 - \alpha_1'\right)\left(V_1[\alpha, j] - V_2[\alpha, j]\right) \geq \left(\alpha_1 - \alpha_1'\right)\left(V_1\left[\alpha', j\right] - V_2\left[\alpha', j\right]\right)$$

or

$$V_1[\alpha, j] - V_1\left[\alpha', j\right] \geq V_2[\alpha, j] - V_2\left[\alpha', j\right]. \tag{23.19}$$

Suppose that $V_1[\alpha, j] < V_1[\alpha', j]$. Then (23.19) implies that $V_2[\alpha, j] < V_2[\alpha', j]$. But this is impossible, because then $(V_1[\alpha', j], V_2[\alpha', j])$ would Pareto dominate $(V_1[\alpha, j], V_2[\alpha, j])$. □

Let $\alpha^* = (\alpha_1^*, \alpha_2^*)$ be a vector of the Pareto weights for which first best allocation is sustainable. Consider any other initial vector $\alpha_0 \neq \alpha^*$and suppose that the first best allocation that corresponds to that vector is not sustainable. This implies that at least for one party the sustainability constraint (23.12) binds. The next lemma shows that sustainability constraint (23.12) does not bind if any party has a Pareto weight higher that α_j^*. While the proof of the lemma is somewhat involved, the intuition for the result is straightforward. If a party has a Pareto weight higher than the sustainable first best weight, the planner is treating such agent better than in the sustainable allocation. Therefore, when in power the sustainability constraint does not bind.

Lemma 23.7. *Suppose Assumptions 23.1 and 23.3 hold. Suppose that $\alpha_j\left(h^{t-1}\right) \geq \alpha_j^*$ for some j, h^{t-1} and $\mathbf{j}\left(h^t\right) = j$. Then $\lambda_j\left(h^t\right) = 0$ and $\alpha_j\left(h^t\right) = \alpha_j\left(h^{t-1}\right)$.*

Proof. Let us consider the relaxed problem of maximizing (23.9) without the constraint (23.12) following history h^t such that $\alpha_j\left(h^{t-1}\right) \geq \alpha_j^*$. We will characterize the solution to this relaxed problem and then show that the solution in fact satisfies (23.12) establishing that $\lambda_j\left(h^t\right)=0$ and $\alpha_j\left(h^t\right)=\alpha_j\left(h^{t-1}\right)$.

Without loss of generality assume that $j=1$. The expected utility of party 1 in history h^t in the relaxed problem is

$$u_1\left(c_1\left(\alpha\left(h^{t-1}\right)\right),l_1\left(\alpha\left(h^{t-1}\right)\right)\right)+\beta\left(m(1|1)V_1\left[\alpha\left(h^{t-1}\right),1\right]+m(2|1)V_1\left[\alpha\left(h^{t-1}\right),2\right]\right)$$

where $\left(c_i\left(\alpha\left(h^{t-1}\right)\right),l_i\left(\alpha\left(h^{t-1}\right)\right)\right)$ is a solution to the maximization problem

$$\max_{\{c_i,l_i\}}\alpha_1\left(h^{t-1}\right)u_1\left(c_1,l_1\right)+\alpha_2\left(h^{t-1}\right)u_2\left(c_2,l_2\right)$$

subject to

$$c_1+c_2\leq l_1+l_2.$$

Since there is no sustainability constraint, Assumption 23.3 immediately implies that $l_j\left(\alpha\left(h^{t-1}\right)\right)=1$ for all j, and moreover, $u_1\left(c_1\left(\alpha\left(h^{t-1}\right)\right),l_1\left(\alpha\left(h^{t-1}\right)\right)\right)$ is increasing in $\alpha_1\left(h^{t-1}\right)$.

Since Pareto weights α^* correspond to the sustainable allocation,

$$\begin{aligned}&u_1\left(c_1\left(\alpha^*\right),l_1\left(\alpha^*\right)\right)+\beta\left(m(1|1)V_1\left[\alpha^*,1\right]+m(2|1)V_1\left[\alpha^*,2\right]\right)\\&\geq v_1\left(l_2\left(\alpha^*\right)\right)+\beta V_1^W(1)\end{aligned}\tag{23.20}$$

Once again, Assumption 23.3 implies that $l_j\left(\alpha^*\right)=1$ for all j. From Lemma 23.6, $V_1\left[\alpha\left(h^{t-1}\right),j\right]\geq V_1\left[\alpha^*,j\right]$ for all j. Therefore the solution to the relaxed problem satisfies (23.12) if

$$u_1\left(c_1\left(\alpha\left(h^{t-1}\right)\right),1\right)-v_1(1)\geq u_1\left(c_1\left(\alpha^*\right),1\right)-v_1(1)$$

Since $u_1\left(c_1\left(\alpha\left(h^{t-1}\right)\right),1\right)$ is increasing in $\alpha_1\left(h^{t-1}\right)$ and $\alpha_1\left(h^{t-1}\right)\geq\alpha_1^*$, this inequality is satisfied. □

The previous lemma established that if party j is in power and has an updated Pareto weight above α_j^*, its next period updated Pareto weight remains the same. The next key step in our argument is to show that if a party has Pareto weight is below α_j^*, its next period updated Pareto weight is also below α_j^* (even if its current sustainability constraint is binding).

LEMMA 23.8. *Suppose Assumptions 23.1 and 23.3 hold. Suppose that* $\alpha_j\left(h^{t-1}\right)<\alpha_j^*$ *for some* j,h^{t-1} *and* $\mathbf{j}\left(h^t\right)=j$. *Then* $\alpha_j\left(h^t\right)\leq\alpha_j^*$ *for all subsequent* h^t.

PROOF. See the Appendix. □

This lemma is the key to the main result in this subsection. It shows that if the sustainability constraint does not hold for group j that is in power even though its Pareto weight is below α_j^*, then for all subsequent histories its Pareto weight will not exceed α_j^*. The proof utilizes quasi-linearity of preferences to put structure on updated Pareto weights and the corresponding allocations.

Now we are ready to prove the most important result of this subsection about the convergence to the first best allocations.

Proposition 23.9. *Suppose that Assumptions 23.1 and 23.3 hold. If $\beta \geq \bar{\beta}$, then for all j*

$$\left(c_j\left(h^t\right), l_j\left(h^t\right)\right) \rightarrow \left(c_j^*, l_j^*\right)$$

where $\left\{\left(c_j^*, l_j^*\right)\right\}_{j=1}^2$ is a first best sustainable allocation.

Proof. Suppose the first best allocation with Pareto weight α^* is sustainable. Without loss of generality, suppose $\alpha_1\left(h^0\right) \leq \alpha_1^*$. Lemmas 23.7 and 23.8 imply that $\alpha_1\left(h^t\right)$ is a monotonically increasing sequence bounded above by α_1^*. Therefore $\alpha_1\left(h^t\right)$ must converge. Such convergence is possible only if (23.12) does not bind for both parties, which is possible only for the first best sustainable allocation. □

This proposition establishes that if there exist first-best allocations that are sustainable they will be ultimately reached. This implies that the political economy frictions in this situation will disappear in the long run. The resulting long-run allocations will not feature distortions and fluctuations in consumption and labor supply. Note, however, that the theorem does not imply that such first-best allocations will be reached immediately. Sustainability constraints may bind for a while, because the sustainable first-best allocations may involve too high a level of utility for one of the groups. In this case, a first-best allocation will be reached only after a specific path of power switches increases the Pareto weight of this group to a level consistent with a first-best allocation. After this point, sustainability constraints do not bind for either party, and thus Pareto weights are no longer updated and the same allocation is repeated in every period thereafter. Interestingly, however, this first-best allocation may still involve transfer from one group to another.

23.2.9. Political Stability and Efficiency. The framework presented here enables an investigation of the implications of persistence of power on the sustainability of first-best applications. In particular, the "stability" or persistence of power is captured by the underlying Markov process for power switches. If the Markov process $m\left(j \mid j'\right)$ makes it very likely that one of the groups, say group 1, will be in power all the time, we can think of this as a very "stable distribution of political power".

The main result will show that higher persistence of power makes distortions more likely, in the sense that it leads to a smaller set of sustainable first-best allocations (four to general general setup with N parties and non quasi-linear utilities).

Proposition 23.10. *Consider an economy consisting of N groups, with group j having utility functions $u_j(c_j, l_j)$ satisfying Assumption 23.1. Suppose that $m(j \mid j) = \rho$ and $m(j' \mid j) = (1-\rho)/(N-1)$ for any $j' \neq j$. Then $\bar{\beta}$ is increasing in ρ, i.e., the set of sustainable first-best allocations is smaller when ρ is greater.*

Proof. Recall that a first-best allocation satisfies (23.17). The left-hand side of this expression is independent of ρ, so is the first term on the right-hand side. Therefore, the desired result follows if the second term on the right-hand side, $V_j^W(j)$, is increasing in ρ, i.e., if (23.20) holds for any pair (β, ρ), then it holds for any (β, ρ') with $\rho' \leq \rho$, and thus the threshold $\bar{\beta}(\rho)$ about which it holds is increasing in ρ.

We now prove that this is the case. From the specification of the power switching process, we have that group j will remain in power next period with probability ρ, and hence $V_j^W(j)$ satisfies

$$V_j^W(j) = \rho V_j^P + (1-\rho) V_j^{NP}, \tag{23.21}$$

where V_j^P and V_j^{NP} are respectively the utility of being in power and not in power after a deviation. These are given by

$$V_j^P = u_j\left(\tilde{l}_j, \tilde{l}_j\right) + \beta\rho V_j^P + \beta(1-\rho) V_j^{NP},$$

and

$$V_j^{NP} = \beta\left(1 - \frac{1-\rho}{N-1}\right) V_j^{NP} + \beta\left(\frac{1-\rho}{N-1}\right) V_j^P,$$

where $\tilde{l}_j$ solves (23.8). Subtracting the second equation from the first, we obtain

$$V_j^P - V_j^{NP} = \frac{u_j\left(\tilde{l}_j, \tilde{l}_j\right)}{1 - \beta\rho + \beta\left(\frac{1-\rho}{N-1}\right)},$$

and therefore

$$V_j^P = \frac{1 - \beta + \beta\left(\frac{1-\rho}{N-1}\right)}{(1-\beta)\left(1 - \beta\rho + \beta\left(\frac{1-\rho}{N-1}\right)\right)} u_j\left(\tilde{l}_j, \tilde{l}_j\right),$$

and substituting this into (23.21), we obtain

$$V_j^W(j) = \frac{\beta\left(\frac{1-\rho}{N-1}\right) + (1-\beta)\rho}{(1-\beta)\left(1-\beta\rho+\beta\left(\frac{1-\rho}{N-1}\right)\right)} u_j\left(\tilde{l}_j, \tilde{l}_j\right),$$

which is increasing in ρ, establishing the desired result. □

This proposition implies the converse of the Olson conjecture discussed above: the set of sustainable first-best allocations is maximized when there are *frequent power switches* between different groups. The Olson conjecture is based on the idea that "effective discount factors" are lower with frequent power switches, and this should make "cooperation" more difficult. "Effective discount factors" would be the key factor in shaping cooperation (the willingness of the party in power to refrain from deviating) only if those in power can only be rewarded when in power. This is not necessarily the case, however, in reality or in our model. In particular, in our model deviation incentives are countered by increasing current utility and the Pareto weight of the party in power, and, all else equal, groups with higher Pareto weights will receive greater utility in all future dates. This reasoning demonstrates why "effective discount factor" is not necessarily the appropriate notion in this context. Instead, Proposition 23.10 has a simple intuition: the value of deviation for a group in power is determined by the persistence of power; when power is highly persistent, deviation becomes more attractive, since the group in power can still obtain relatively high returns following a deviation as it is likely to remain in power. In contrast, with more frequent power switches, the group in power is likely to be out of power tomorrow, effectively reducing deviation value. Since first-best allocations, and thus first-best utilities, are independent of the persistence of power, this implies that greater persistence makes deviation more attractive relative to candidate first-best allocations, and thus first-best allocations become less likely to be sustainable.

23.3. References

(1) Dixit, Avinash, Gene M. Grossman, and Faruk Gul (2000) "The Dynamics of Political Compromise," *Journal of Political Economy* 108, 531-568.

(2) Acemoglu, Daron, Michael Golosov and Oleg Tsyvinski (2008) "Political Economy and the Structure of Taxation" mimeo.