Chapter 1

Mutual and Higher-Order Expectations

1.1 Introduction

The typical *game playing situation* involves a group of *self-interested* agents, or players, engaged in some "*interdependent* decision problem" [55]. Each person in the group privately chooses a course of action under the assumption that the final outcome depends on the decision of *everyone* in the group. Agents are self-interested in the sense that they are free to solve their individual decision problems however they see fit. They are not constrained or committed to any team or group point of view — see [66, 8, 65] for an analysis of group decision making under such commitments — but this does not mean that the agents are completely selfish. For some of them, altruistic considerations may play a crucial role in their final decision.

Two basic assumptions are important for our study. First, each agent only has partial information about the full consequences given each of his options, also called actions in the game-theoretic literature. The consequence of choosing a given option depends on what the others will choose, and the agents typically have only partial information about each others' choices. In the decision theory literature, this is referred to as a decision under uncertainty, as opposed to under *risk* where the probabilities of each outcome given a specific option are available to the decision-maker[1]. Secondly, and most important, the agents recognize that they are, in fact, engaged in an interdependent decision problem, i.e. in a game playing situation. They are aware that they are interacting with other agents who are themselves aware of the interactive character of the situation. That is, they are interacting with agents whose choices will depend, in part, on what they expect the others will do.

Let us consider an example where two players, Ann and Bob, are in a game situation where they are trying to coordinate their choices:

[1]See [49, Chapter 1] for a nice discussion of this important distinction.

		Bob *L*	Bob *R*
Ann	*u*	1, 1	0, 0
Ann	*d*	0, 0	1, 1

They each face an individual decision problem and recognize that their decision problems are interdependent. Much of the traditional work in game theory has centered around the questions: *What should Ann (Bob) do*? or *what is rational for Ann (Bob) to do*? Game theorists have proposed various "solution concepts" as answers to these two questions (the most well known examples being the *Nash equilibrium* and *dominance reasoning*).

However, in recent years a number of game-theorists, philosophers and logicians have moved away from directly asking "*what is rational for Ann (Bob) to do*?" and tried to understand *what does it mean for Ann (Bob) to act rationally in a given interactive situation.* These two questions are quite different. In the first case, one looks at social or interactive situations in abstraction from their specific context, and tries to circumscribe, in a (quasi-)axiomatic fashion, criteria of rational decision making. In the other case, one looks at specific game playing situations, here at specific informational contexts, and tries to understand how this context will or should bear on the agents' decisions.

In the example above, if Ann expects that Bob will behave *rationally* (whatever that may mean) she seems under normative pressure to incorporate this fact in her practical reasoning, and similarly for Bob. What are the consequences of this? A classical take on instrumental rationality recommends that she should choose the option she considers best given what she *expects.* That is, whether or not a particular choice is rational depends, in part, on Ann's current *information.* In our example, this information is of a very specific kind: it is about the potential choices of Bob (i.e., about what Bob might choose). If Ann recognizes that Bob is, just like her, a rational agent trying to make the best out of the situation given what he expects, then her expectations about what Bob will choose depends on what she expects Bob will be expecting of her, which in turn depends on what she expects Bob will be expecting she will be expecting of him, and so on.

These interrelated expectations do not need to affect the underlying view about what constitutes a rational choice for Ann (Bob), namely choosing the best option given her (his) current information. Yet, they crucially change the informational background and this takes us beyond this basic view of instrumental rationality. This is what we intend to show in this book: *a theory of interactive rationality is a theory of mutual and higher-order expectations.* In other words, we claim that the interrelated aspect of Ann and Bob's expectations and practical reasoning is a central part of what it means to be rational in game situations. The next chapter discusses different models that have been recently developed to study mutual and higher-order expectations. The remainder of the book will employ these models to to develop our theory interactive rationality as a theory of mutual and higher-order expectations.

We are primarily concerned with the questions of how expectations drive the players' practical reasoning in game situations, keeping the "rationality" parameter constant almost all the way through, viz. to classical "Bayesian" rationality. However, this should not be seen as an argument against alternative notions of practical rationality[2], e.g., satisficing [60] or "rule-rationality" [2, 5][3]. Our methodology is entirely modular and towards the end of the book we will consider the consequences of "plugging in" different concepts of practical rationality. Until then, keeping the notion of rationality constant will allow us to highlight how different types of information influence the players' reasoning during the (private) decision-making process. This is in line with a increasingly popular, but of course not uncontroversial point of view (cf. [44]), in epistemic logic and game theory that "*the* fundamental insight of game theory [is] that a rational player must take into account that the players reason about each other in deciding how to play." [6, pg. 81]. Exactly how the players incorporate the fact that they are interacting with other (actively reasoning) agents into their own decision making process is the subject of much debate. As we will see, subtle differences in assumptions about the players' informational attitudes will lead to alternative accounts of what constitutes rational play.

1.2 Preliminaries: Information in Interaction

Our analysis begins by being more precise about how to represent the information that the players have in an interactive situation. This will allow us to discuss a number of results from the so-called *epistemic program in game theory* (cf. [19, 20]), which we think currently constitute the most sophisticated body of work on the role of mutual and higher-order expectations in interactive rationality. The central thesis of the epistemic program in game theory is that the basic mathematical models of a game situation should include an explicit parameter describing the players informational attitudes. A simple political example suffices to show that this goes beyond the mundane observation that players are usually not perfectly informed about all parameters in a game situation:

> Formally, a game is defined by its strategy sets and payoff functions. But in real life, many other parameters are relevant; there is a lot

[2]Or against alternative accounts of *rational decision making* taking into account the (observed) *behavior* of human beings when making decisions [45] or the fact that humans are *bounded* reasoners [31, 52].

[3]In these two papers Aumann makes a distinction between "rule-rationality" and "act-rationality" roughly analogous to the the well-knowing distinction between *act-utilitarianism* of Jeremy Bentham and *rule-utilitarianism* of John Stuart Mill, John Harsanyi and others.

> more going on. Situations that substantively are vastly different may nevertheless correspond to precisely the same strategic game. For example, in a parliamentary democracy with three parties, the winning coalitions are the same whether the parties each hold a third of the seats in parliament, or, say, 49 percent, 39 percent, and 12 percent, respectively. But the political situations are quite different. The difference lies in the attitudes of the players, in their expectations about each other, in custom, and in history, though the rules of the game do not distinguish between the two situations. [6, pg. 72]

This remark nicely highlights the delicate issue that mathematical models used by game theorists to represent social situations abstract away from many details that are often deemed important in other areas of the social sciences. Epistemic game theory takes one step away from this absract analysis by including the *information* that players have in the analysis of an interactive situation. So, among the many different aspects of a social interactive situation that are not *explicitly* represented in classical game-theoretical models, the agents' informational attitudes (including higher-order notions such as *common knowledge*) are singled out as important to make explicit.

Harsanyi [38] proposed to capture this idea via the notion of a player's *type*[4], an additional parameter to a basic model of a game situation, and this has become a widely accepted approach in epistemic game theory:

> We think of a particular [type] structure as giving the "context" in which the game is played. In line with Savage's Small-Worlds idea in decision theory [...], who the players are in the given game can be seen as a shorthand for their experiences before the game. The players' possible characteristics — including their possible types — then reflect the prior history or context. [21, pg. 319]

While the focus on explicitly modeling the players' informational attitudes is certainly not the *only* interesting extension of the basic game theoretic models, it is by far the most well-developed[5]. A common side-effect of becoming a "well-developed area" is an abundance of modeling techniques. This is certainly true here: many different formal models have been used to represent the players reasoning and information in game situations. This presents us with a concrete challenge as we want to understand how the specific results found in the literature contribute to a general theory of interactive rationality. Sometimes differences between formal models simply reflect the mathematical conventions found in different research (sub)communities. Other technical differences are motivated by important conceptual issues. The rest of this chapter is devoted to discussing the

[4] See [50] for an interesting discussion of the history of this notion and the reception of Harsanyi's paper.

[5] For a sampling of the extensive literature, see [18, 68] and references therein.

key approaches to formally modeling the players' information in a game situation (i.e., the players' types). We start by discussing some general issues that motivate specific modeling choices.

1.2.1 Ex Ante, Ex Interim and Ex Post: Where to Locate the Analysis?

Following much of the literature in epistemic game theory, we distinguish three stages of the decision making process: *ex ante*, *ex interim* and *ex post*. At one extreme is the *ex ante* stage where no decision has been made yet. The other extreme is the *ex post* stage where the choices of all players are openly disclosed. In between these two extremes is the *ex interim* stage where, to use a well-known metaphor from von Neumann and Morgenstern [73], the players have passed their envelopes to the umpire, i.e., they have made their decision, but they are still uninformed about the decisions and intentions of the other players.

These distinctions are not intended to be very sharp. Rather, they describe various degree of information disclosure during the decision-making process. At the *ex-ante* stage, little is known except the structure of the game, who is taking part, and possibly (but not necessarily) some aspect of the agents' character[6]. In particular, the agents have not yet made a decision; "the future is open" [13]. At the *ex-post* stage the game is basically over: all player have made their decision and these are now irrevocably out in the open. This does not mean that all uncertainty is removed as an agent may remain uncertain about what exactly the others were expecting of him. Note that, in general, there will not be a unique correspondence between having certain expectations and playing a certain strategy: many coherent systems of higher-order expectations might rationalize a given strategy choice[7]. In between these two extreme stages lies a whole gradation of states of information disclosure that we loosely refer to as "the" *ex-interim* stage. Common to these stages is the fact that the agents have made *a* decision, although not necessarily an irrevocable one. In the umpire-envelope metaphor, especially suited for games in strategic form[8], the agents' decisions do seem irrevocable. Things are less clear in extensive games or if we allow pre-play communication. In these cases, players might want to revise their decision on the basis what they observe while the game is being played or simply from what they have told each other. The models of mutual and higher-order expectations that we introduce below will be used to precisely capture these subtleties.

For the main part of this book we focus on this *ex interim* stage, except in Section **??**, because it allows for a straightforward assessment of the agents' rationality given their expectations. Indeed, one needs to look at the decisions and

[6]We will come back on this while discussing the common prior assumption in Chapter 4.
[7]We come back on this in Chapter 3
[8]Strategic and extensive form games are introduced in Chapter 1.

expectations in order to assess the rationality of the former, and the *ex-interim* stage gives us just that: a point at which the agents have made a decision but where much uncertainty remains. How agents should react, that is eventually revise their decision, upon learning that they did not choose "rationally" is an issue that we will leave on the side for most of our study. Understanding the role of higher-order expectations in rational interactive decision making is enough of a task for now. We leave the issue of decision or intention revision for further investigation (cf. [58, 43]).

1.2.2 Varieties of Informational Attitudes

Contemporary epistemology provides us with a rich typology of informational attitudes. There are numerous notions of knowledge around: the pre-Gettier "justified true belief" view, reliability accounts [32], counterfactual accounts [51], and *active* vs. *passive* knowledge [63, pg. 299], to name just a few. Similarly, beliefs come in many forms: graded or flat-out [37], conditional and lexicographic [19], safe and strong [10]. On top of all this, beliefs seem to be just one example in a large variety of "acceptance-like" attitudes [57].

In this book we draw from a general distinction in dynamic epistemic logic, between attitudes of *hard* and *soft* information [67, 9], and explore the consequences of various assumptions on such informational attitudes for our theory of interactive rationality. We do not take a stance on the issue of which of these attitudes, if any, should be seen as primary, either for epistemology in general or for the specific purpose of our analysis of interactive rationality.

By *hard information*, and its companion attitude, we mean information that is *veridical*, *fully introspective* and *not revisable.* This notion is intended to capture what the agents are fully and correctly certain of in a given interactive situation. At the *ex interim* stage, for instance, the players have hard information about their *own* choice. They "know" which strategy they have chosen (i.e., what they "have written on the envelope"), they know that they know this, and no new incoming information could make them change their opinion on this. As this phrasing suggests, we will often use the term *knowledge*, in absence of better terminology and following common usage in the literature, to describe this very strong type of informational attitude. We make no claim as to whether this notion captures one of the many notions of knowledge just mentioned (in fact, it probably does not). We simply note that "hard information" shares *some* of the characteristics that have been attributed to knowledge in the epistemological literature such as veracity. Furthermore, its modeling comes quite close what has been called "implicit knowledge" or, arguably, to Lewis' picture of "elusive knowledge" (as long as the remains unscrutinized, that is).[9]

[9] Indeed, unlike Lewis' notion of knowledge hard information does not vanish upon careful scrutiny.

Soft information is, roughly speaking, anything that is not "hard": it is not necessarily veridical, not necessarily fully introspective and/or highly revisable in the presence of new information. As such, it comes much closer to *beliefs*. Once again, philosophical carefulness is in order here. The whole range of informational attitudes that we label as "beliefs" indeed falls into the category of attitudes that can be described as "regarding something as true" [56]. However, this truth-oriented character of beliefs only plays a marginal role in our analysis. Instead, we focus on the function of these soft informational attitudes as input into practical reasoning, a function that they surely share with many other non-belief-like attitudes [22, 57]. We remain agnostic as whether the whole range of attitudes that we call "beliefs" should not rather be called, say, "acceptances".

Two main types of models have been used in the literature to represent hard and soft information: *type spaces* [38, 61] and the so-called *Aumann-* or *Kripke-structures* [7, 28]. These two structures differ in the way they represent the two kinds of information. A second distinction found in the literature is between "quantitative" structures, representing "graded beliefs" (typically via probability distributions), and "qualitative" structures using "plausibility orderings".

In the game-theoretic literature, these different modeling paradigms have given rise to different styles of epistemic analysis [19]: so-called *belief*-based characterizations on the one hand and *knowledge*-based characterizations on the other. We see this plurality as a source of interesting research questions rather than of ideological debates. What kind of theory of interactive rationality does one get by focusing only on attitudes of hard information? How much generality is lost by considering only flat-out instead of graded beliefs? How do the two kinds of modeling strategies relate? In what follows and in the coming chapters, we will be addressing these questions and so work in parallel with both type spaces *and* qualitative "plausibility models".

1.3 Hard and Soft Information: Mathematical Models

The models introduced in this section describe the players' hard and soft information of interactive situations. The models differ in their representation of a state of the world, but are otherwise closely related.

The formal models introduced in this can be broadly described as "possible worlds models" familiar in much of the philosophical logic literature. We start with a non-empty (finite or infinite) set S of *states of nature* describing the *exogenous* parameters (i.e., facts about the physical world) that do not depend on the agents' uncertainties. In the scenarios that we focus on in this book, the set S is simply the set of possible outcomes in a game situation. Thus, in a strategic game, S will typically be the set of all *strategy profiles*[10]. Each agent is assumed to entertain a number of *possibilities*, called *possible worlds* or simply *(epistemic) states*. These "possibilities" are intended to represent a possible way a game situation may evolve. So each possibility will be associated with a *unique* state of nature (i.e., there is a function from possible worlds to states of nature, but this function need not be 1-1 or even onto). Crucial for our analysis is the assumption that there may be *different* possible worlds associated with the same state of nature. Such possible worlds are important because they open the door to representing different state of information. One final common feature is that the agents' informational attitudes will be directed towards *propositions*, also called *events* in the game-theory literature, represented as sets of possible worlds. These basic modeling choices are not uncontroversial, but such issues are not our concern in this book and so, in this situation, we opt for mathematical precision in favor of philosophical carefulness.

1.3.1 Relational Models

We first look at the kind of model of informational attitudes that are familiar to philosophical logicians [42, 69], computer scientists [28] and game theorists [7]. We first introduce the model themselves, and then say a few words about the *level of description*, i.e. the *language* at which one may want to carry out the analysis in such models.

Representing Hard Information: States, Partitions, and Relations

First, some terminology. Given a set W of states, or possible worlds, let us call any subset $E \subseteq W$ an *event* or *proposition*. Given events $E \subseteq W$ and $F \subseteq W$, we use standard set-theoretic notation for intersection ($E \cap F$, read "E and F"), union ($E \cup F$, read "E or F") and (relative) complement ($-E$, read "not E"). We say that an event $E \subseteq W$ occurs at state w if $w \in E$. This terminology will be crucial for studying the following models:

1.3.1. Definition. Let S be a non-empty set of states of nature and $\mathcal{A}$ a set of agents, or players. An **epistemic model based on S and $\mathcal{A}$** is a triple $\langle W, \{\Pi_i\}_{i \in \mathcal{A}}, \sigma \rangle$, where W is a nonempty set, for each $i \in \mathcal{A}$, Π_i is a partition over W and $\sigma : W \to S$.

Epistemic models are representations of the informational context of a given situation, in terms of possible configurations of states of nature and information that the agents have about them. The latter is represented by a partition over

[10] Recall from Chapter 1 that a strategy profile in a strategic game is a sequence $(a_1, a_2, \ldots, a_n)$ where each a_i is an available action for player i.

the set of states[11]. The function σ assigns to each possible world a unique state of nature in which every ground, i.e. non-informational fact is either true or false. If $\sigma(w) = \sigma(w')$ then the two worlds w, w' will agree on all the ground facts but, crucially, the agents may have different information in them. So, elements of W are *richer*, than the elements of S (more on this below).

Given a state $w \in W$, the cell $\Pi_i(w)$ is called agent i's *information set*. Following standard terminology, if $\Pi_i(w) \subseteq E$, we say the agent i *knows* the event E at state w. Given an event E, the event that agent i knows E is denoted $K_i(E)$. Formally, we define for each agent i a knowledge function assigning to every event E the event that the agent i knows E:

1.3.2. Definition. Let $\mathcal{M} = \langle W, \{\Pi_i\}_{i\in\mathcal{A}}, \sigma\rangle$ be an epistemic model. The **knowledge function** for agent i based on $\mathcal{M}$ is $K_i : \wp(W) \to \wp(W)$ with:

$$K_i(E) = \{w \mid \Pi_i(w) \subseteq E\}$$

where for any set X, $\wp(X)$ is the *powerset of* X.

1.3.3. Remark. It is often convenient to work with *equivalence relations* rather than partitions. In this case, an epistemic model based on S and $\mathcal{A}$ can also be defined as a triple $\langle W, \{\sim_i\}_{i\in\mathcal{A}}, \sigma\rangle$ where W and σ are as above and for each $i \in \mathcal{A}$, $\sim_i \subseteq W \times W$ is reflexive, transitive and symmetric. Given such a model $\langle W, \{\sim_i\}_{i\in\mathcal{A}}, \sigma\rangle$, we write $[w]_i = \{v \in W \mid w \sim_i v\}$ for the equivalence class of w. Since there is a 1-1 correspondence between equivalence relations and partitions[12], we will abuse notation and use $\sim_i$ and Π_i interchangeably.

Applying the above remark, an alternative definition of $K_i(E)$ is that E is true in all the states the agent i considers possible (according to i's hard information). That is, $K_i(E) = \{w \mid [w]_i \subseteq E\}$.

Partitions or equivalence relations are intended to represent the agents' hard information at each state. The following properties of K_i show that such information is indeed truthful, fully introspective and satisfy strong closure properties.

1.3.4. Fact. Let $\mathcal{M}$ be an epistemic model and K_i the associated knowledge function for agent i. For all events E and F:

$E \subseteq F \Rightarrow K_i(E) \subseteq K_i(F)$	*Monotonicity*
$K_i(E \cap F) = K_i(E) \cap K_i(F)$	*Closure under intersection*
$K_i(E) \subseteq E$	*Truth*
$K_i(E) \subseteq K_i(K_i(E))$	*Positive introspection*
$-K_i(E) \subseteq K_i(-K_i(E))$	*Negative introspection*
$K_i(\emptyset) = \emptyset$	*Consistency*

[11]A partition of W is a pairwise disjoint collection of subsets of W whose union is all of W. Elements of a partition Π on W are called **cells**, and for $w \in W$, let $\Pi(w)$ denote the cell of Π containing w.

[12]Given an equivalence relation $\sim_i$ on W, the collection $\Pi_i = \{[w]_i \mid w \in W\}$ is a partition. Furthermore, given any partition Π_i on W, $\sim_i = \{(w, v) \mid v \in \Pi_i(w)\}$ is an equivalence relation with $[w]_i = \Pi_i(w)$.

The simple but instructive proof of this Fact is well-known, so we will not repeat it here (see, for example, [7, 28]).

Viewed as a description, even an idealized one, of *knowledge*, the properties stated in the above fact have raised many criticisms. Monotonicity, for instance, is one of the aspects of the infamous logical omniscience problem, stating that an agent's knowledge is closed under logical consequence (cf., for example, [28], Chapter 9) — we give a precise definition of this notion later. The two introspection principles have also been the object of intense discussion[13]

These discussions are of fundamental importance for the theory of knowledge its formalization, but here we choose to bracket them, and instead to keep a close watch on their ramification on our theory of interactive rationality. We thus take partitional models for what they are, models of hard information, in the sense introduced above, and look at how, and when, the properties of such information come into play in the various results from epistemic game theory that we will study in the next chapters. In other words, we will not go into details about any of the many arguments for or against the principles stated above, but rather investigate when they are really needed, and assess the philosophical significance of various formal results in the light of these findings.

An Example, *Ex Interim* Models and Information Hierarchies

Recall the example of a strategic game from Chapter 1:

		Bob	
		L	R
Ann	u	$1, 1$	$0, 0$
	d	$0, 0$	$1, 1$

Figure 1.1: A strategic coordination game between Ann and Bob

The utilities of the players are not important for us at this stage. To construct an epistemic model for this game, we need first to specify what are the states of

[13]In fact, Hintikka explicitly rejects negative introspection: "The consequences of this principle, however, are obviously wrong. By its means (together with certain intuitively acceptable principles) we could, for example, show that the following sentence is self sustaining $E \subseteq K_i - K_i - (E)$" [42, pg. 54]. Hintikka regards this last statement as counter-intuitive since it means that if it is possible that an agent knows some event E then that fact must be true, but it seems plausible that an agent can justifiably believe that she knows something that is in fact false.

There is also an interesting discussion regarding the "KK Principle": does knowing something imply that one knows that one knows it? Timothy Williamonson Williamson [74, Chapter 5] has a well-known and persuasive argument against this principle, and Egrè and Bonnay [27] have a response relevant for the formal models we discuss in this book.

nature we will consider. For simplicity, take them to be the set of strategy profiles $S = \{(u, L), (d, L), (u, R), (d, L)\}$. The set of agents is of course $\mathcal{A} = \{A, B\}$. What will be the set of state W? We start by assuming $W = S$, so there is exactly one possible world corresponding to each state of nature. This needs not be so, but here this will help to illustrate our point.

There are many different partitions for Ann and Bob that we can use to complete the description of this simple epistemic model. Not all of the partitions are appropriate for analyzing the *ex interim* stage of the decision-making process, though. For example, suppose $\Pi_A = \Pi_B = \{W\}$ and consider the event $U = \{(u, l), (u, r)\}$ representing the situation where Ann chooses u. Notice that $K_A(U) = \emptyset$ since for all $w \in W$, $\Pi_A(w) \not\subseteq U$, so there is no state where Ann *knows* that she chooses u. This means that this model is appropriate for reasoning about the *ex ante* stage rather than the *ex interim* stage. This is easily fixed with an additional technical assumption: Suppose S is a set of strategy profiles for some (strategic or extensive) game with players $\mathcal{A} = \{1, \ldots, n\}$.

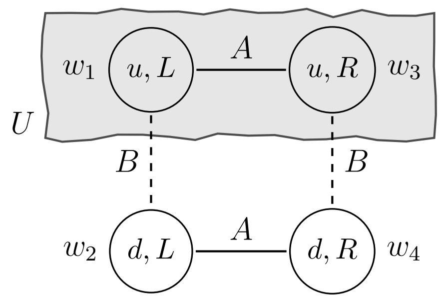
> A model $\mathcal{M} = \langle W, \{\Pi_i\}_{i \in \mathcal{A}}, \sigma \rangle$ is said to be an ***ex interim* epistemic model** if for all $i \in \mathcal{A}$ and $w, v \in W$, if $v \in \Pi_i(w)$ then $\sigma_i(w) = \sigma_i(v)$

where $\sigma_i(w)$ is the ith component of the strategy profile $s \in S$ assigned to w by σ. An example of an *ex interim* epistemic model with states W is:

- $\Pi_A = \{\{(u, L), (u, R)\}, \{(d, L), (d, R)\}\}$ and
- $\Pi_B = \{\{(u, L), (d, L)\}, \{(u, R), (d, R)\}\}$.

Note that this simply reinterprets the game matrix in Figure 1.1 as an epistemic model where the rows are Ann's information sets and the columns are Bob's information sets. Unless otherwise stated, we will always assume that our epistemic models are *ex interim*. The class of *ex interim* epistemic models is very rich with models describing the (hard) information the agents have about their own choices, the (possible) choices of the other players *and* higher-order (hard) information (eg., "Ann knows that Bob knows that...") about these decisions.

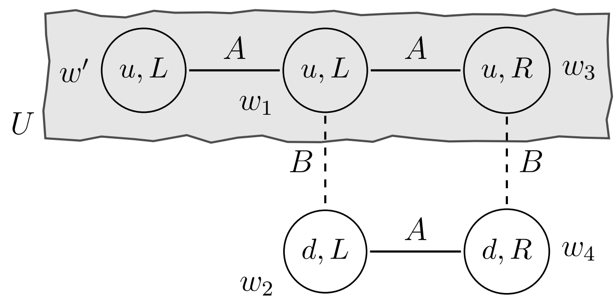
We now look at the epistemic model described above in more detail. We will often use the following diagrammatic representation of the model to ease exposition. States are represented by nodes in a graph where there is a (undirected) edge between states w_i and w_j when w_i and w_j are in the same partition cell. We use a solid line labeled with A for Ann's partition and a dashed line labeled with B for Bob's partition (reflexive edges are not represented for simplicity). The event $U = \{w_1, w_3\}$ representing the proposition "Ann decided to choose option u" is the shaded gray region:



Notice that the following events are true at all states:

1. $-K_B(U)$: "Bob does not know that Ann decided to choose u"
2. $K_B(K_A(U) \vee K_A(-U))$: "Bob knows that Ann knows wether she has decided to choose u"
3. $K_A(-K_B(U))$: "Ann knows that Bob does not know that she has decided to choose u"

In particular, these events are true at state w_1 where Ann has decided to choose u (i.e., $w_1 \in U$). The first event makes sense given the assumptions about the available information at the *ex interim* stage: each player knows their own choice but not the other players' choices. The second event is a concrete example of another assumption about the available information: Bob has the information that Ann has, in fact, made *some* choice. But what warrants Ann to conclude that Bob does not know she has chosen u (the third event)? This is a much more significant statement about what Ann knows about what Bob expects her to do. Indeed, in certain contexts, Ann may have very good reasons to think it is possible that Bob actually *knows* she will choose u. We can find an *ex interim* epistemic model where this event ($-K_A(-K_B(U))$) is true at w_1, but this requires added a new possible world:



Notice that since $\Pi_B(w') = \{\{w'\}\} \subseteq U$ we have $w' \in K_B(U)$. That is, Bob knows that Ann chooses u at state w'. Finally, a simple calculation shows that $w_1 \in -K_A(-K_B(U))$, as desired.

Information Hierarchies and Group Knowledge

In the example above we listed some events that are true at state w_1. We can continue this description by consider the following collections of events : $\mathcal{E}_A = \{E \mid w_1 \in K_A(E)\}$, $\mathcal{E}_{BA} = \{E \mid w_1 \in K_B K_A(E)\}$, $\mathcal{E}_{ABA} = \{E \mid w_1 \in K_A K_B K_A(E)\}$, and so on. In this way we can associate with each possible world (in a fixed epistemic model) a *hierarchy of knowledge*:

1.3.5. DEFINITION. Let $\mathcal{M} = \langle W, \{\Pi_i\}_{i\in\mathcal{A}}, \sigma\rangle$ be an epistemic model with $w \in W$ and define the following collections of events true at w:

First-Order	$\mathcal{E}_1(w) = \bigcup_{i\in\mathcal{A}}\{E \mid w \in K_i(E)\}$
Second-Order	$\mathcal{E}_2(w) = \bigcup_{i,j\in\mathcal{A}}\{E \mid w \in K_i K_j(E)\}$
$\vdots$	$\vdots$
Nth-Order	$\mathcal{E}_n(w) = \bigcup_{i_1,\ldots,i_n\in\mathcal{A}}\{E \mid w \in K_{i_1} K_{i_2} \cdots K_{i_n}(E)\}$
$\vdots$	$\vdots$

Therefore, associated with each possible world in an epistemic model is a precise description of all the information (including the higher-order information) that the players have access to during their practical reasoning. We will discuss these hierarchies of knowledge in more detail later in this chapter, but it is important now to note that these hierarchies are concrete examples of the players' *types* introduced in Section 1.2.

Returning to the game situation in Figure 1.3.1, a natural question is which of the above hierarchies of knowledge is sufficient for Ann and Bob to coordinate their actions (i.e., for both Ann and Bob to both receive a payoff of 1)? Lewis [46] and Clark and Marshall [25] argue that a condition of *common knowledge* is necessary for such coordinated actions. In fact, a seminal result by [33] shows that, without synchronized clocks, such coordinated action is impossible. Chwe [24] has a number of examples that point out the everyday importance of the notion of common knowledge.

Both the game theory community and the epistemic logic community have extensively studied formal models of common knowledge and belief. Barwise [15] highlights three main approaches to formalize common knowledge: (i) the iterated view, (ii) the fixed-point view and (iii) the shared situation view. In this book, we primarily focus on the first two approaches (see [70] for a rigorous comparison between (i) and (ii)). Vanderschraaf and Sillari [71] provide an extensive discussion of the literature.

Consider the statement "everyone in group G knows φ". This can be easily described by the following event:

$$K_G(E) \quad := \quad \bigcap_{i\in G} K_i(E)$$

where $G \subseteq \mathcal{A}$. Following Lewis [46][14], the interpretation of "it is common knowledge in G that E" (denoted $C_G(E)$) is the *infinite* conjunction:

$$E \cap K_G(E) \cap K_G K_G(E) \cap K_G K_G K_G(E) \cap \cdots$$

It is not hard to see that the following is true for any event E:

$$C_G(E) \subseteq K_G(C_G(E))$$

This demonstrates the "self-evident" nature of common knowledge: if some fact is common knowledge in some group G then everyone in G not only knows the fact but also that it is common knowledge. In fact, Aumann [3] uses this as an alternative characterization of common knowledge:

> Suppose you are told "Ann and Bob are going together,"' and respond "sure, that's common knowledge." What you mean is not only that everyone knows this, but also that the announcement is pointless, occasions no surprise, reveals nothing new; in effect, that the situation after the announcement does not differ from that before. ...the event "Ann and Bob are going together" — call it E — is common knowledge if and only if some event — call it F — happened that entails E and also entails all players' knowing F (like all players met Ann and Bob at an intimate party). [3, pg. 271]

We conclude this section with an alternative characterization of common knowledge. Recall that we can use equivalence relations $\sim_i$ in place of partitions Π_i in epistemic models (see Remark 1.3.3). Now consider the event $K_G K_G K_G(E)$ which says "everyone from group G knows that everyone from group G knows that everyone from group G knows that E". When will this be true at a state w? First some notation: a **path on length n for G** in an epistemic model is a sequence of states $(w_0, w_2, \ldots, w_n)$ where for each $l = 0, \ldots, n-1$, we have $w_l \sim_i w_{l+1}$ where $i \in G$. Thus, $w \in K_G K_G K_G(E)$ iff every path of length 3 for G starting at w leads to a state in E. This suggests the following alternative characterization of common knowledge:

> $w \in C_G(E)$ iff every finite path for G from w ends with a state in E.

Modal Languages and the Level of Description

Despite their apparent simplicity, epistemic models are extremely fine-grained descriptions of informational situations, and they can be extremely rich. The reader accustomed with set theory might indeed have already observed that the hierarchy of knowledge defined in the previous section needs not to stop at ω.

[14] Although see [26] for an alternative reconstruction of Lewis' notion of common knowledge.

There is, in fact, no upper bound to the richness of such models — in a sense that we make precise at the end of the chapter.

In view of this, it is natural to ask how much details should one take into account while analyzing the informational attitudes of the agents in an interactive situation. Is any event, however *complex*[15], in principle 'knowable' by the agents? What is the use of distinguishing two epistemic models where the agents have exactly the same hard information, except concerning some event (set of states) which can only be described by some function which is not in principle *computable*? Concerning the various closure principle mentioned above, does it make a difference to assume that the agents' hard information is closed only under finitary, instead of arbitrary conjunction? This is connected to some issues discussed by Rubinstein [53] and Lipman [47] concerning the modeling of *language* in Economic models.

Consciously or not, epistemic logicians since Hintikka (1962) have answered some of these questions by looking at (modal) *syntactic* descriptions of epistemic models. These syntax come in various guises, giving more or less *expressive power* [17] to the language. Thus far, we have assumed that states of nature S are simply the set of possible outcomes in a game situation (eg., the strategy profiles). The first step in defining a *formal language* is to choose a language for describing these states of nature. To keep things simple we start with a *propositional language*[16]. Let PROP be a countable set of atomic propositions. These atomic propositions are intended to describe the (ground) facts about the game situation. Thus, in our models we replace the function σ assigning a state of nature to each state with a *valuation function* assigning a description of a state of nature to each state. Typical examples are "Ann choose action a", but one may also describe other features of the interactive situation not typically explicitly included in a game structure. We largely focus on one of the most well-known formal languages for talking about epistemic models: "standard" propositional epistemic logic with common knowledge,

1.3.6. Definition. Let PROP be a countable set of atomic propositions, and $\mathcal{A}$ a finite set of agent. The **finitary epistemic language** $\mathcal{L}_{EL}$ is smallest set generated by the following grammar:

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box_i\varphi \mid \Box^*_G\varphi$$

where i ranges over $\mathcal{A}$, p over PROP and $\emptyset \neq G \subseteq \mathcal{A}$.

The basis of such a language is a simple propositional logic containing atomic propositions together with the Boolean operations of negation and conjunction,

[15] Logicians have a number of ways to make precise this notion of complexity, but these issues are not crucial for our discussion here.

[16] Although first-order or even higher-order languages may also be interesting here, we do not discuss them in this book.

on the basis of which one can define disjunction and implication, that we write $\varphi \vee \psi$ and $\varphi \rightarrow \psi$, respectively. The epistemic operators $\Box_i$ are used to describe the agents' hard information, their "knowledge", and the operator $\Box^*_G$ is used to describe what is "common knowledge" within a group of agent G. These operators have their duals, $\Diamond_i$ and $\Diamond^*_i$, defined as $\neg\Box_i\neg$ and $\neg\Box^*_i\neg$.

This language is *finitary* in two important respects. First, it only contains finite conjunctions and disjunction. Second, only finite, though unbounded "stacking" of epistemic operators ($\Box_i\Box_j...\Box_k\varphi$) are allowed. This finitary character, has we will see shortly, as important consequences on what such a language can say about epistemic models.

Of course, there is a strong connection between, on the one hand, an epistemic language with its Boolean connectives and epistemic operators and, on the other hand, the set-theoretic descriptions of epistemic models of the previous Section, where we talked about knowledge functions and operations such as intersection, union and complements. This relationship can be made precise by giving a *semantics* for our epistemic language requiring a slight modification of Definition 1.3.1 as noted above: (also it is convenient to use equivalence relations):

1.3.7. Definition. Let $\mathcal{L}_{EL}$ be an epistemic language. A **model for $\mathcal{L}_{EL}$** is a triple $\langle W, \{\sim_i\}_{i\in\mathcal{A}}, V\rangle$, where W is a nonempty set of states, for each $i \in \mathcal{A}$, $\sim_i$ is an equivalence relation on W and $\sigma : W \rightarrow \mathcal{P}(\text{PROP})$ is a *valuation function.*

Epistemic models are special cases of this general definition, taking $\text{PROP} = S$ and V such that $V(w)$ is a singleton for all w. With this in and, one can give the standard modal interpretation of $\mathcal{L}_{EL}$, with $\mathcal{M}, w \models \varphi$ meaning that the formula φ is *true at state w in the model $\mathcal{M}$*.

1.3.8. Definition. Let $\mathcal{M} = \langle W, \{\Pi_i\}_{i\in\mathcal{A}}, V\rangle$ be an epistemic model and $\varphi \in \mathcal{L}_{EL}$. The *truth* of φ at state $w \in W$, denoted $\mathcal{M}, w \models \varphi$, is defined by recursion on the structure of φ:

$$\begin{array}{lll}
\mathcal{M}, w \models p & \text{iff} & p \in V(w) \\
\mathcal{M}, w \models \neg\varphi & \text{iff} & \mathcal{M}, w \not\models \varphi \\
\mathcal{M}, w \models \varphi \wedge \psi & \text{iff} & \mathcal{M}, w \models \varphi \text{ and } \mathcal{M}, w \models \psi \\
\mathcal{M}, w \models \Box_i\varphi & \text{iff} & \forall v \in W \text{ (if } w \sim_i v \text{ then } \mathcal{M}, v \models \varphi) \\
\mathcal{M}, w \models \Box^*_G\varphi & \text{iff} & \forall v \in W \text{ (if } w \sim^*_G v \text{ then } \mathcal{M}, v \models \varphi)
\end{array}$$

where $\sim^*_G$ is the reflexive transitive closure[17] of $\cup_{i\in G} \sim_i$.

Thus, in any epistemic model $\mathcal{M}$, the above Definition assigns to every formula $\varphi \in \mathcal{L}_{EL}$ an event called the *truth set of* φ in $\mathcal{M}$. More formally, for each epistemic model $\mathcal{M}$, define $[\![\cdot]\!]_\mathcal{M} : \mathcal{L}_{EL} \rightarrow \wp(W)$ as follows: $[\![\varphi]\!]_\mathcal{M} = \{w \in W \mid \mathcal{M}, w \models \varphi\}$

[17]The reflexive transitive closure of a relation R is the smallest relation containing R that is reflexive and transitive.

(we leave out the subscript $\mathcal{M}$ when it is clear from context). It is not hard to see that in any epistemic model, $[\![\varphi \wedge \psi]\!] = [\![\varphi]\!] \cap [\![\psi]\!]$, $[\![\neg\varphi]\!] = W - [\![\varphi]\!]$ and

$$[\![\Box_i\varphi]\!] = K_i([\![\varphi]\!]) \text{ and } [\![\Box^*_G\varphi]\!] = C_G([\![\varphi]\!])$$

Epistemic languages set a precise limit on what can be expressed, or described, in an (class of) epistemic model(s). For example, the set of *definable* sets in an epistemic model $\mathcal{M} = \langle W, \{\Pi_i\}_{i\in\mathcal{A}}, V\rangle$, denoted $\mathcal{D}_\mathcal{M} = \{E \subseteq W \mid E = [\![\varphi]\!]_\mathcal{M}$ for some $\varphi \in \mathcal{L}_{EL}\}$, in general need not be equal to $\wp(W)$. This gives us a precise notion when two states represent same interactive situation (eg., they agree on the truth of all *definable* sets). Unfortunately, this notion is not very practical, since it requires checking the truth of *every* formula in $\mathcal{L}_{EL}$. Fortunately, a key notion from the mathematical analysis of modal logic [17] can help here providing us with a *model-theoretic* characterization of when two states satisfy the same formulas of $\mathcal{L}_{EL}$:

1.3.9. Definition. Given two epistemic models for $\mathcal{L}_{EL}$, $\mathcal{M} = \langle W, \{\Pi_i\}_{i\in\mathcal{A}}, V\rangle$ and $\mathcal{M}' = \langle W', \{\sim'_i\}_{i\in\mathcal{A}}, V'\rangle$ a *bisimulation* is a relation $\underline{\leftrightarrow} \subseteq W \times W'$ such that, for all $w \in W$ and $v \in W'$, if $w \underline{\leftrightarrow} v$ then, for each $i \in \mathcal{A}$:

prop: for all $p \in$ PROP, $p \in V(w)$ iff $p \in V'(v)$.

forth: if $w \sim_i w'$ then there is a $v' \in W'$ such that $v \sim'_i v'$ and $w' \underline{\leftrightarrow} v'$.

back: if $v \sim'_i v'$ then there is a $w' \in W$ such that $w \sim_i w'$ and $w' \underline{\leftrightarrow} v'$.

It is well known that if two states are bisimilar then they must satisfy the same formulas of epistemic logic (i.e., $\mathcal{L}_{EL}$ is *invariant* under bisimulation). The converse is true provided the set of states is finite. Of course, employing more expressive languages, for instance by allowing infinite conjunctions and disjunctions, requires strengthening the above notion of bisimulation. For now, it suffices to note that specifying a syntax is a natural way to stipulate precisely which events are taken into account while analyzing an interactive situation.

Of course, adopting this "syntactic approach" brings with it a number of technicalities and it is natural to wonder whether the extra precision is worth the extra work. In fact, the syntactic approach sketched here and the "semantic approach" we followed in the previous sections are often seen as *competing* styles of analysis [3, 4, 39, 54]. However, we rather see the two styles as *complementary* (this is in line standard results in modal logic relating *algebraic* models to relational models [72]) each raising interesting technical and conceptual issues. With this in mind, in what follows, we do not strictly adhere to any one style of analysis. Details of a specific formal language will be given when it aides exposition and otherwise we will work with events (with the understanding that there is an underlying formal language that can describe the relevant events).

Soft Information and Plausibility Orderings

So far we have looked at relational models of hard information. A small modification of these models allows us to model a softer informational attitude. Indeed, by simply replacing the assumption of reflexivity of the relation $\sim_i$ with seriality (for each state w there is a state v such that $w \sim_i v$), but keeping the other aspects of the model the same, we can capture what epistemic logicians have called "*beliefs*". Formally, a **doxastic model** is a tuple $\langle W, \{R_i\}_{i\in\mathcal{A}}, V\rangle$ where W is a nonempty set of states, R_i is a transitive, Euclidean and serial relation on W and V is a valuation function (cf. Definition 1.3.1). This notion of belief is very close to the above hard informational attitude and, in fact, shares all the properties of K_i listed above except *Veracity* (this is replaced with a weaker assumption that agents are "consistent" and so cannot believe contradictions). This points to a logical analysis of both informational attitudes with various "bridge principles" relating knowledge and belief (such as knowing something implies believing it or if an agent believes φ then the agent knows that he believes it). However, we do not discuss this line of research (see, for example, [34, 64]) here since these models are not our preferred ways of representing the agents' soft information.

A key aspect of beliefs which is not yet represented in the above models is that they are *revisable* in the presence of new information. While there is an extensive literature on the theory of belief revision in the "AGM" style [1], we focus on how to extend an epsitemic models with a representation of softer, revisable informational attitudes. The standard approach is to include a *plausibility ordering* for each agent: a preorder (reflexive and transitive) denoted $\preceq_i \subseteq W \times W$. If $w \preceq_i v$ we say "player i considers v at least as plausible as w." For an event $X \subseteq W$, let

$$Min_{\preceq_i}(X) = \{v \in W \mid v \preceq_i w \text{ for all } w \in X\ \}$$

denote the set of minimal elements of X according to $\preceq_i$. Thus while the $\sim_i$ partitions the set of possible worlds according to the agents' hard information, the plausibility ordering $\preceq_i$ represents which of the possible worlds the agent considers more likely (i.e., it represents the players soft information).

1.3.10. DEFINITION. Suppose $\mathcal{A}$ is a set of agents and At a set of atomic propositions, an **epistemic doxastic model** is a tuple $\langle W, \{\Pi_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$ where $\langle W, \{\Pi_i\}_{i\in\mathcal{A}}, V\rangle$ is an epistemic model and for each $i \in \mathcal{A}$, $\preceq_i$ is a well-founded[18], reflexive and transitive relation on W satisfying the following properties, for all $w, v \in W$

1. *plausibility implies possibility*: if $w \preceq_i v$ then $w \sim_i v$.

2. *locally-connected*: if $w \sim_i v$ then either $w \preceq_i v$ or $v \preceq_i w$.

[18]Well-foundedness is only needed to ensure that for any set X, $Min_{\preceq_i}(X)$ is nonempty. This is important only when W is infinite.

1.3.11. Remark. Note that if $w \not\sim_i v$ then, since $\sim_i$ is symmetric, we also have $v \not\sim_i w$, and so by property 1, $w \not\preceq_i v$ and $v \not\preceq_i w$. Thus, we have the following equivalence: $w \sim_i v$ iff $w \preceq_i v$ or $v \preceq_i w$.

Let $[w]_i$ be the equivalence class of w under $\sim_i$. Then local connectedness implies that $\preceq_i$ totally orders $[w]_i$ and well-foundedness implies that $Min_{\preceq_i}([w]_i)$ is nonempty. This richer model allows us to formally define a variety of (soft) informational attitudes. We first need some additional notation: the plausibility relation $\preceq_i$ can be lifted to subsets of W as follows[19]

$$X \preceq_i Y \text{ iff } x \preceq_i y \text{ for all } x \in X \text{ and } y \in Y$$

Suppose $\mathcal{M} = \langle W, \{\Pi_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$ is an epistemic-doxastic model with $w \in W$, consider the following extensions to the language $\mathcal{L}_{EL}$

- *Belief*: $\mathcal{M}, w \models B_i\varphi$ iff for all $v \in Min_{\preceq_i}([w]_i)$, $\mathcal{M}, v \models \varphi$.
  This is the usual notion of belief which satisfies the standard properties discussed above (eg., consistency, positive and negative introspection).
- *Safe Belief*: $\mathcal{M}, w \models \Box_i\varphi$ iff for all v, if $v \preceq_i w$ then $\mathcal{M}, v \models \varphi$.
  This stronger notion of belief has also been called *certainty* by some authors [cf. 59, Section 13.7].
- *Strong Belief*: $\mathcal{M}, w \models B_i^s\varphi$ iff there is a v such that $w \sim_i v$ and $\mathcal{M}, v \models \varphi$ and $\{x \mid \mathcal{M}, x \models \varphi\} \cap [w]_i \preceq_i \{x \mid \mathcal{M}, x \models \neg\varphi\} \cap [w]_i$.
  So φ is strongly believed provided it is epistemically possible and agent i considers *any* state satisfying φ more plausible than *any* state satisfying $\neg\varphi$. This notion has also been studied by [62] and [16].

The logic of these notions has been extensively studied by Alexandru Baltag and Sonja Smets in a series of articles [2006, 2008, 2006, 2009]. We conclude this section with a few remarks about the relationship between these different notions. For example, it is not hard to see that if agent i knows that φ then i (safely, strongly) believes that φ. However, much more can be said about the logical relationship between these different notions [cf. 12].

As noted above, a crucial feature of these informational attitudes is that they may be defeated by appropriate evidence. In fact, we can characterize these attitudes in terms of the type of evidence which can prompt the agent to adjust her beliefs. To make this precise, we introduce the notion of a *conditional belief*: suppose $\mathcal{M} = \langle W, \{\Pi_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$ is an epistemic-doxastic and φ and ψ are formulas, then we say i *believes* φ *given* ψ, denoted $B_i^\psi\varphi$, provided

$$\mathcal{M}, w \models B_i^\psi\varphi \text{ iff for all } v \in Min_{\preceq_i}([\![\psi]\!]_\mathcal{M} \cap [w]_i),\ \mathcal{M}, v \models \varphi$$

[19] This is only one of many possible choices here, but it is the most natural in this setting [cf., 48, Chapter 4].

So, 'B_i^ψ' encodes what agent i will believe upon receiving (possibly misleading) evidence that ψ is *true*. Two observations are immediate. First of all, we can now define belief $B_i\varphi$ as $B_i^\top\varphi$ (belief in φ given a tautology). Second, unlike beliefs, conditional beliefs may be inconsistent (i.e., $B^\psi\bot$ may be true at some state). In such a case, agent i cannot (on pain of inconsistency) revise by ψ, but this will only happen if the agent has hard information that ψ is false. Indeed, $K\neg\varphi$ is logically equivalent to $B_i^\varphi\bot$ (over the class of epistemic-doxastic models). This suggests the following (dynamic) characterization of an agents' hard information as unrevisable beleifs:

$$\mathcal{M}, w \models K_i\varphi \text{ iff } \mathcal{M}, w \models B_i^\psi\varphi \text{ for all } \psi$$

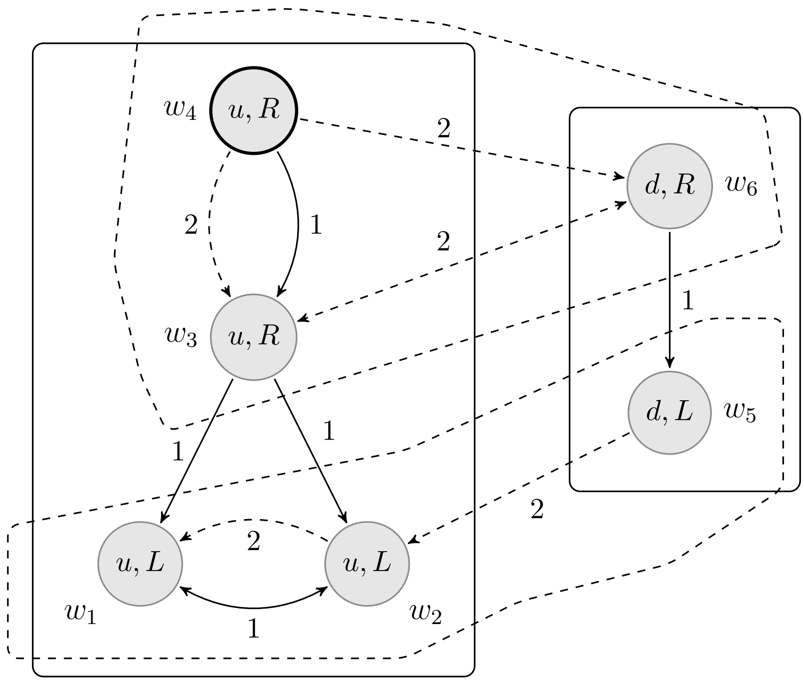
Safe belief and strong belief can be similarly characterized by restricting the admissible evidence:

- $\mathcal{M}, w \models \Box_i\varphi$ iff $\mathcal{M}, w \models B_i^\psi\varphi$ for all ψ with $\mathcal{M}, w \models \psi$.
  That is, i safely beliefs φ iff i continues to believe φ given any true formula.
- $\mathcal{M}, w \models B_i^s\varphi$ iff $\mathcal{M}, w \models B_i\varphi$ and $\mathcal{M}, w \models B_i^\psi\varphi$ for all ψ with $\mathcal{M}, w \models \neg K_i(\psi \to \neg\varphi)$.
  That is, agent i strongly believes φ iff i believes φ and continues to believe φ given any evidence (truthful or not) that is not known to contradict φ.

Baltag and Smets [12] provide an elegant logical characterization of the above notions by adding the safe belief modality ($\Box_i$) to the epistemic language $\mathcal{L}_{EL}$ (denote the new language $\mathcal{L}_{EDL}$). The key observation is that conditional belief (and hence belief) and strong belief are *definable* in this language (proof of this is left to the reader):

- $B_i^\varphi\psi \ := \ L_i\varphi \to L_i(\varphi \wedge \Box_i(\varphi \to \psi))$
- $B_i^s\varphi \ := \ B_i\varphi \wedge K_i(\varphi \to \Box_i\varphi)$

We conclude this section with an example to illustrate the above concepts. Recall again the coordination game of Figure 1.3.1: there are two actions for player 1 (Ann), u and d, and two actions for player 2 (Bob), R and L. The preferences (or utilities) of the players are not important at this stage since we are only interested in describing the players' information. The following epistemic-doxastic model is a possible description of the players' informational attitudes that can be associated with this game. The solid lines represent player 1's informational attitudes and the dashed line represents player 2's. The arrows correspond to the players plausibility orderings with an i-arrow from w to v meaning $v \preceq_i w$ (we do not draw all the arrows: each plausibility ordering can be completed by filling in arrows that result from reflexivity and transitivity). The different regions represent the players' hard information.



Suppose that the actual state is w_4. So, player 1 is choosing u and player 2 is choosing R. The following formulas are true at w_4: suppose that p_L describes the event $\{w_1, w_2, w_5\}$ where player 2 chooses L (similarly for p_u, p_d, and p_R)

1. $B_1 p_L$: "player 1 believes that player 2 is choosing L"

2. $B_1 B_2 p_u$: "player 1 believes that player 2 believes that player 1 chooses u"

3. $B_1^{p_R} \neg B_2 p_u$: "player 1 believes that player 2 does not believe she is choosing u given that player 2 chooses R"

This last formula is interesting because it "pre-encodes" what player 1 would believe upon learning that player 2 is choosing R. Note that upon receiving this *true* information, player 1 drops her belief that player 2 believes she is choosing u. The situation can be even more interesting if there are statements in the language that reveal only *partial* information about the player strategy choices. Suppose that p describes the event $\{w_4, w_6\}$. Now p is true at w_4 and player 2 believes that *player 1 chooses* d given that p is true (i.e., $B_2^p p_d$ is true at w_4). So, player 1 can "bluff" by revealing the true (though partial) information p.

1.3.2 Type Spaces and Probabilistic Models

State w in an epistemic-doxastic model $\mathcal{M}$ for a fixed strategic game describes both the strategy choices (either via a function σ or described in some formal language via the valuation function V) and the players' informaitonal attitudes. This includes the players' *higher order* information (eg., common knowledge/belief). This section examines different ways of representing this information.

One alternative approach was initiated by Harsanyi in his seminar paper [38]. Rather than "possible worlds", Harsanyi takes the notion of the players' *type* as primitive. Formally, the players are assigned a nonempty set of types. Typically, players are assumed to *know* their own type but not the types of the other players. As we will see, each type can be associate with a specific hierarchy of knowledge.

1.3.12. DEFINITION. A **Qualitative type space** for a (nonempty) set of states of nature S and agents $\mathcal{A}$ is a tuple $\langle \{T_i\}_{i\in\mathcal{A}}, \{\lambda_i\}_{i\in\mathcal{A}}, S\rangle$ where for each $i \in \mathcal{A}$, T_i is a nonempty set and

$$\lambda_i : T_i \rightarrow \wp(\times_{j\neq i} T_j \times S).$$

So, each each type $t \in T_i$ is associated with a set of tuples consisting of types of the other players and a state of nature. For simplicity, suppose there are only two players, Ann and Bob. Intuitively, $(t', o') \in \lambda_{Ann}(t)$ means that Ann's type t considers it possible that the outcome is o' *and* Bob is of type t'. Since the players uncertainty is directed at the choices and types of the *other* players, the informational attitude captured by these models will certainly not satisfy the Truth axiom. In fact, we will show that qualitative types spaces can be viewed as simply a "re-packaging" of a doxastic model.

Consider again the running example of the coordination game between Ann and Bob (pictured in Figure 1.1). In this case, the set of states of nature is $S = \{(u, L), (d, L), (u, R), (d, L)\}$. In this context, it is natural to modify the definition of the type functions λ_i to account for the fact that the players are only uncertain about the other players' choices: let $S_A = \{u, d\}$ and $S_B = \{L, R\}$ and suppose T_A and T_B are nonempty sets of types. Define λ_A and λ_B as follows:

$$\lambda_A : T_A \rightarrow \wp(T_B \times S_B) \qquad \lambda_B : T_B \rightarrow \wp(T_A \times S_A)$$

Suppose that there are two types for each player: $T_A = \{t_1^A, t_2^A\}$ and $T_B = \{t_1^B, t_2^B\}$. A convenient way to describe the maps λ_A and λ_B is:

$\lambda_A(t_1^A)$

	L	R
t_1^B	1	0
t_2^B	1	0

$\lambda_A(t_2^A)$

	L	R
t_1^B	0	0
t_2^B	1	0

$\lambda_B(t_1^B)$

	u	d
t_1^A	1	0
t_2^A	0	0

$\lambda_B(t_2^B)$

	u	d
t_1^A	0	0
t_2^A	0	1

where a 1 in the (t', s) entry of the above matrices corresponds to assuming $(t', s) \in \lambda_i(t)$ $(i = A, B)$. What does it means for Ann (Bob) to *believe* an event E in a type structure? We start with some intuitive observations about the above type structure:

- Regardless of what type we assign to Ann, she believes that Bob will choose L since in both matrices, $\lambda_A(t_1^A)$ and $\lambda_A(t_2^A)$, the only places where a 1 appears is under the L column. So, fixing a type for Ann, in all of the situations Ann considers possible it is true that Bob chooses L.
- If Ann is assigned the type t_1^A, then she considers it possible that Bob believes she will choose u. Notice that type t_1^A has a 1 in the row labeled t_1^B, so she considers it possible that Bob is of type t_1^B, and type t_1^B believes that Ann chooses u (the only places where 1 appears is under the u column).
- If Ann is assigned the type t_2^A, then Ann believes that Bob believes that Ann believes that Bob will choose L. Note that type t_2^A "believes" that Bob will choose L and furthermore t_2^A believes that Bob is of type t_2^B who in turn believes that Ann is of type t_2^A.

We can formalize the above informal observations using the following notions: Fix a qualitative type space $\langle \{T_i\}_{i\in\mathcal{A}}, \{\lambda_i\}_{i\in\mathcal{A}}, S\rangle$ for a (nonempty) set of states of nature S and agents $\mathcal{A}$.

- A **(global) state**, or **possible world** is a tuple $(t_1, t_2, \dots, t_n, s)$ where $t_i \in T_i$ for each $i = 1, \dots, n$ and $s \in S$. If $S = \mathsf{X} S_i$ is the set of strategy profiles for some game, then we write a possible world as: $(t_1, s_1, t_2, s_2, \dots, t_n, s_n)$ where $s_i \in S_i$ for each $i = 1, \dots, n$.
- Type spaces describe the players beliefs about the other players' choices, so the notion of an *event* needs to be relativized to an agent. An **event for agent** i is a subset of $\mathsf{X}_{j\neq i} T_j \times S$. Again if S is a set of strategy profiles (so $S = \mathsf{X} S_i$), then an event for agent i is a subset of $\mathsf{X}_{j\neq i}(T_j \times S_j)$.
- Suppose that E is an event for agent i, then we say that agent i **believes** E **at** $(t_1, t_2, \dots, t_n, s)$ provided $\lambda(t_1, s) \subseteq E$.

In the specific example above, an event for Ann is a set $E \subseteq T_B \times S_B$ and we can define the set of pairs (t^A, s^A) that believe this event:

$$Bel_A(E) = \{(t^A, s^A) \mid \lambda_A(t^A, s^A) \subseteq E\}$$

similarly for Bob. Note that the event $Bel_A(E)$ is an event for Bob and vice versa. Using this formal machinery we can associate with each type a hierarchy of *belief* as we did for states in an epistemic model. In fact, we can show that qualitative type spaces are formally equivalent to doxastic models (see [75, Chapter 3] and [76] for proofs).

The above models use a "crisp" notion of uncertainty, i.e., for each agent and state w, any other state $v \in W$ is either is or is not possible/more plausible than w. However, there is an extensive body of literature focused on *graded*, or *quantitative*, models of uncertainty [35]. For instance, in the Game Theory literature it is standard to represent the players' *beliefs* by probabilities [4, 38]. The idea is to replace the plausibility orderings with probability distributions:

1.3.13. Definition. An **epistemic-proabalistic model** is a tuple $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{P_i\}_{i\in\mathcal{A}}, V\rangle$ where $\langle W, \{\Pi_i\}_{i\in\mathcal{A}}, V\rangle$ is an epistemic model and $P_i : W \to \Delta(W)$ ($\Delta(W) = \{p : W \to [0,1] \mid p \text{ is a probability measure }\}$) assigns to each state a probability measure over W. Write p_i^w for the i's probability measure at state w. We make two natural assumptions (cf. Definition 1.3.10):

1. For all $v \in W$, if $p_i^w(v) > 0$ then $p_i^w = p_i^v$
2. For all $v \notin \Pi_i(w)$, $p_i^w(v) = 0$

Property 1 says that if i assigns a non-zero probability to state v at state w then the agent uses the same probability measure at both states. This means that the players "know" there own probability measures. The second property implies that players must assign a probability of zero to all states outside the current (hard) information cell. These model provide a very precise description of the players' hard and soft informational attitudes. However, note that writing down a model requires us to specify a different probability measure for each partition cell which can be quite cubersome. Fortunately,the properties in the above definition implies that, for each agent, we can view the agent's probability measures as arising from one probability measure through conditionalization. Formally, for each $i \in \mathcal{A}$, agent i's **(subjective) prior probability** is any element of $p_i \in \Delta(W)$. Then, in order to define an epistemic-probability model we need only give for each agent $i \in \mathcal{A}$ (1) a prior probability $p_i \in \Delta(W)$ and (2) a partition Π_i on W such that for each $w \in W$, $p_i(\Pi_i(w)) > 0$. The probability measures for each $i \in \mathcal{A}$ are then defined by:

$$P_i(w) = p_i(\cdot \mid \Pi_i(w)) = \frac{p_i(\cdot)}{p_i([w]_i)}$$

Of course, the side condition that for each $w \in W$, $p_i([w]_i) > 0$ is important since we cannot divide by zero — this will be discussed in more detail in later chapters. Indeed, (assuming W is finite[20]) given any epistemic-plausibility model we can find, for each agent, a prior (possibly different ones for different agents) that

[20] Some care needs to be taken when W is infinite, but these technical issues are not important for us at this point, so we restrict attention to finite sets of states.

generates the model as described above. This is not only a technical observation: it means that we are assuming that the players' subject beliefs about the outcome of the situation are fixed *ex ante* with the *ex interim* beliefs being derived through conditionalization on the agent's *hard information.* We will return to these key assumptions throughout the text.

Many different formal languages have been used to describe these rich structures. Examples range from '$\Box_i\varphi$' with the intended meaning "φ is more probable than $\neg\varphi$ for agent i" [41] to more expressive languages containing operators of the form $B_i^q\varphi$ (with q a rational number) and interpreted as follows:

$$\mathcal{M}, w \models B_i^q(\varphi) \text{ iff } p_i^w(\{v \mid \mathcal{M}, v \models \varphi\}) \geq q.$$

These models have also been subjected to sophisticated logical analyses [30, 29, 40] complementing the logical frameworks introduced in this Chapter (cf. [14]).

A small change to the above definition of a type space (Definition 1.3.12) allows us to represent *probabilistic* beliefs (we give the full definition here for future reference):

1.3.14. Definition. A **type space** for a (nonempty) set of states of nature S and agents $\mathcal{A}$ is a tuple $\langle\{T_i\}_{i\in\mathcal{A}}, \{\lambda_i\}_{i\in\mathcal{A}}, S\rangle$ where for each $i \in \mathcal{A}$, T_i is a nonempty set and

$$\lambda_i : T_i \to \Delta(\mathsf{X}_{j\neq i}T_j \times S).$$

where $\Delta(\mathsf{X}_{j\neq i}T_j \times S)$ is the set of probability measures on $\mathsf{X}_{j\neq i}T_j \times S$.

Types and their associated image under λ_i encode the players' (probabilistic) information about the others' information. Indeed, each type is associated with a hierarchy of belief. More formally, recall that an event E for a type t_i is a set of pairs (σ_{-j}, t_{-j}), i.e., a set of strategy choices and types for all the other players. Given an event E for player i, let $\lambda_i(t_i)(E)$ denote the sum of the probabilities that $\lambda_i(t_i)$ assigns to the elements of E. The type t_i of player i is said to *(all-out) believe* the event E whenever $\lambda_i(t_i)(E) = 1$. Conditional beliefs are computed in the standard way: type t_i believes that E given F whenever:

$$\frac{\lambda_i(t_i)(E \cap F)}{\lambda_i(t_i)(F)} = 1$$

A *state* in a type structure is a tuple (σ, t) where σ is a strategy profile and t is "type profile", a tuple of types, one for each player. Let $B_i(E) = \{(\sigma_{-j}, t_{-j}) :$ t_i believes that $E\}$ be the event (for j) that i believes that E. Then agent j believes that i believes that E when $\lambda_j(t_j)(B_i(E)) = 1$. We can continue in this manner computing any (finite) level of such higher-order information. We invite the reader to check that this belief hierarchy can also be recovered from a state in an epistemic doxastic model.

An Example

Returning again to our running example game where player 1 (Ann) as two available actions $\{u, d\}$ and player 2 (Bob) has two available actions $\{L, R\}$. The following type space describes the players' information: there is one type for Ann (t_1) and two for Bob (t_2, t_2') with the corresponding probability measures given below: In this example, since there is only one type for Ann, both of Bob's

$\lambda_1(t_1)$:

	L	R
t_2	0.5	0
t_2'	0.4	0.1

Figure 1.2: Ann's beliefs about Bob

$\lambda_2(t_2)$:

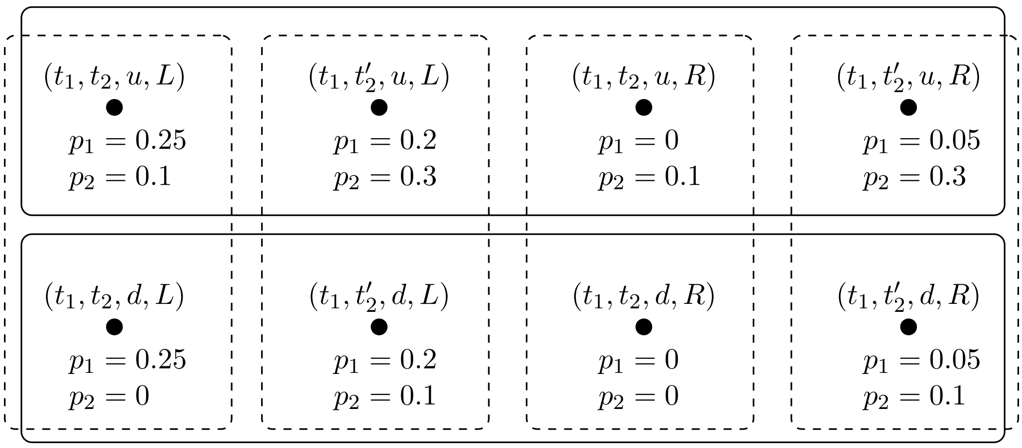
	u	d
t_1	1	0

$\lambda_2(t_2')$:

	u	d
t_1	0.75	0.25

Figure 1.3: Bob's beliefs about Ann

types are *certain* about Ann's beliefs. If Bob is of type t_2 then he is certain Ann is choosing u while if he is of type t_2' he thinks there is a 75% chance she plays u. Ann assigns equal probability (0.5) to Bob's types; and so, she believes it is equally likely that Bob is certain she plays u as Bob thinking there is a 75% chance she plays u. The above type space is a very compact description of the players' informational attitudes. An epistemic-probabilistic model can describe the same situation (here p_i for $i = 1, 2$ is player i's prior probability):



Some simple (but instructive!) calculations can convince us that these two models represent the same situation. The more interesting question is how to these probabilistic models relate to the epistemic-doxastic models of Definition 1.3.10. Here the situation is more complex. On the one hand, probabilistic models with a graded notion of belief which is much more fine-grained than the "all-out" notion of belief discussed in the context of epistemic-doxastic models. On the other hand, in an epistemic-doxastic model, conditional believes are defined for *all* events. In the above, models, they are only defined for events that are assigned nonzero probabilities. In other words, epistemic-probabilistic models do not describe what a player may believe upon learning something "surprising" (i.e., something currently assigned probability zero).

A number of extensions to basic probability theory have been discussed in the literature that address precisely this problem. We do not go into details here about these approaches (a nice summary and detailed comparison between different approaches can be found in [36]) and instead sketch the main ideas. The first approach is to use so-called *Popper functions* which takes *conditional probability measures* as primitive. That is, for each non-empty event A, there is a probability measure $p_A(\cdot)$ satisfying the usual Kolmogrov axioms (relativized to A, so for example $p_A(A) = 1$). A second approach assigns to each agent a finite sequence of probability measures $(p_1, p_2, \ldots, p_n)$ called a *lexicographic probability system*. The idea is that to condition on B, first find the first probability measure not assigning zero to B and use that measure to condition on B. Roughly, one can see each of the probability measures in a lexicographic probability system as corresponding to a level of a plausibility ordering. We will return to these notions later in the book.

1.4 Conclusion

This Chapter has introduced a number of different mathematical models that represent a players *ex interim* information in a game situation.

▶ Add summary

Bibliography

[1] Carlos E. Alchourron, Peter Gardenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50(2):510–530, Jun. 1985. ISSN 00224812. URL http://links.jstor.org/sici?sici=0022-4812%28198506%2950%3A2%3C510%3AOTLOTC%3E2.0.CO%3B2-3.

[2] R. Aumann. Rationality and bounded rationlity. *Games and Economic Behavior*, 21:2 – 14, 1997.

[3] R. Aumann. Interactive epistemology I: Knowledge. *International Journal of Game Theory*, 28:263–300, 1999.

[4] R. Aumann. Interactive epistemology II: Probability. *International Journal of Game Theory*, 28:301 – 314, 1999.

[5] R. Aumann. Rule-rationality versus act-rationality. Unpublished Manuscript, 2008.

[6] R. Aumann and J. Dreze. Rational expectations in games. *American Economic Review*, 98:72 – 86, 2008.

[7] R.J. Aumann. Interactive epistemology I: Knowledge. *International Journal of Game Theory*, 28:263–300, 1999.

[8] M. Bacharach. *Beyond Individual Choices: Teams and Frames in Game Theory*. Princeton University Press, Princeton, 2006. Edited by N. Gold and R. Sugden.

[9] A. Baltag and S. Smets. Conditioanl doxastic models: A qualitative approach to dynamic belief revision. In *Electornic notes in theoretical computer science*, volume 165, pages 5 – 21. Springer, 2006.

[10] A. Baltag and S. Smets. Conditional doxastic models: A qualitative approach to dynamic belief revision. In G. Mints and R. de Queiroz, editors, *Proceedings of WOLLIC 2006, Electronic Notes in Theoretical Computer Science*, volume 165, 2006.

[11] A. Baltag and S. Smets. The logic of conditional doxastic actions. In R. van Rooij and K. Apt, editors, *New Perspectives on Games and Interaction.* Texts in Logic and Games, Amsterdam University Press, 2008.

[12] A. Baltag and S. Smets. ESSLLI 2009 course: Dynamic logics for interactive belief revision. Slides available online at `http://alexandru.tiddlyspot.com/#%5B%5BESSLLI09%20COURSE%5D%5D`, 2009.

[13] A. Baltag, S. Smets, and J. Zvesper. Keep 'hoping' for rationality: a solution to the backwards induction paradox. *Synthese*, 169:301–333, 2009.

[14] Alexandrau Baltag and Sonja Smets. From conditional probability to the logic of doxastic actions. In *TARK '07: Proceedings of the 11th conference on Theoretical aspects of rationality and knowledge*, pages 52–61. ACM, 2007.

[15] Jon Barwise. Three views of common knowledge. In *TARK '88: Proceedings of the 2nd conference on Theoretical aspects of reasoning about knowledge*, pages 365–379, San Francisco, CA, USA, 1988. Morgan Kaufmann Publishers Inc. ISBN 0-934613-66-9.

[16] P. Battigalli and M. Siniscalchi. Strong belief and forward induction reasoning. *Journal of Economic Theory*, 105:356 – 391, 2002.

[17] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic.* Cambridge University Press, Cambirdge, 2001.

[18] G. Bonanno and P. Battigalli. Recent results on belief, knowledge and the epistemic foundations of game theory. *Research in Economics*, 53(2):149–225, June 1999.

[19] A. Brandenburger. The power of paradox: some recent developments in interactive epistemology. *International Journal of Game Theory*, 35:465–492, 2007.

[20] A. Brandenburger. Epistemic game theory: an overview. In S. Durlauf and L. Blume, editors, *The New Palgrave Dictionary of Economics.* Palgrave Macmillan, Basingstoke, 2008.

[21] A. Brandenburger, A. Friedenberg, and H. J. Keisler. Admissibility in games. *Econometrica*, 76:307–352, 2008.

[22] M. Bratman. Practical reasoning and acceptance in a context. *Nous*, 1991. Reprinted in [23, p.15-34] The page numbering in the various quotes refers to this second print.

[23] M. Bratman. *Faces of Intention; Selected Essays on Intention and Agency.* Cambridge University Press, 1999.

[24] M. Suk-Young Chwe. *Rational Ritual.* Princeton University Press, 2001.

[25] H.H. Clark and C. R. Marshall. Definite reference and mutual knowledge. In Joshi, Webber, and Sag, editors, *Elements of Discourse Understanding.* Cambridge University Press, 1981.

[26] Robin P. Cubitt and Robert Sugden. Common Knowledge, Salience and Convention: A Reconstruction of David Lewis' Game Theory. *Economics and Philosophy*, 19(2):175–210, 2003.

[27] Paul Egré and Denis Bonnay. Inexact knowledge with introspection. *Journal of Philosophical Logic*, 38(2):179 – 228, 2009.

[28] R. Fagin, J.Y. Halpern, Y. Moses, and M. Vardi. *Reasoning about Knowledge.* MIT Press, 1995.

[29] Ronald Fagin and Joe Halpern. Reasoning about knowledge and probability. *Journal of the ACM*, 41(2):340 – 367, 1994.

[30] Ronald Fagin, Joe Halpern, and Nimrod Megiddo. A logic for reasoning about probabilities. *Information and Computation*, 87(1):78 – 128, 1990.

[31] G. Gigerenzer and R. Selten. *Bounded Rationality: The Adaptive Toolbox.* MIT Press, 2002.

[32] A. Goldman. What is justified belief? In G. Pappas, editor, *Justification and Kowledge*. D. Reidel, 1976.

[33] J. Halpern and Y. Moses. Knowledge and common knowledge in a distributed environment. *ACM-PODC*, pages 50 – 61, 1983.

[34] Joe Halpern. Should knowledge entail belief? *Journal of Philosophical Logic*, 25(5):483 – 494, 1996.

[35] Joe Halpern. *Reasoning about Uncertainty.* The MIT Press, 2003.

[36] Joe Halpern. Lexicographic probability, conditional probability, and nonstandard probability. *Games and Economic Behavior*, 68(1):155–179, January 2010. URL `http://ideas.repec.org/a/eee/gamebe/v68y2010i1p155-179.html`.

[37] G. Harman. *Change in View.* MIT Press, 1986.

[38] J.C. Harsanyi. Games with incomplete informations played by 'bayesian' players. *Management Science*, 14:159–182, 320–334, 486–502, 1967-68.

[39] A. Heifetz. How canonical is the canonical model? A comment on Aumann's interactive epistemology. *International Journal of Game Theory*, 28(3):435–442, 1999.

[40] Aviad Heifetz and Philippe Mongin. Probability logic for type spaces. *Games and Economic Behavior*, 35:31 – 53, 2001.

[41] Andreas Herzig. Modal probability, belief, and actions. *Fundam. Inf.*, 57 (2-4):323–344, 2003.

[42] J. Hintikka. *Knowledge and Belief: An Introduction to the Logic of Two Notions.* ornell University Press, Ithaca, N.Y., 1962.

[43] T. Icard, E. Pacuit, and Y. Shoham. Joint revision of beliefs and intentions. Unpublished Manuscript, 2009.

[44] J. B. Kadane and P. D. Larkey. Subjective probability and the theory of games. *Management Science*, 28(2):113–120, 1982.

[45] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, pages 263–291, 1979.

[46] D. Lewis. *Convention.* Harvard University Press, 1969.

[47] Barton Lipman. Language and economics. In Nicola Dimitri, Marcello Basili, and Itzhak Gilboa, editors, *Cognitive Processes and Economic Behavior*, pages 75 – 93. Routledge, 2003.

[48] F. Liu. *Changing for the Better: Preference Dynamics and Agent Diversity.* PhD thesis, Universiteit van Amsterdam, February 2008.

[49] R. Duncan Luce and H. Raiffa. *Games and Decisions: Introduction and Critical Survey.* John Wiley and Sons, 1957.

[50] R. Myerson. Harsanyi's games with incomplete information. *Management Science*, 50(12):1818–1824, 2004.

[51] R. Nozick. Knowledge and skepticism. In *Philosophical Investigations*, pages 172 – 185. The MIT Press, 1981.

[52] A. Rubinstein. *Modeling Bounded Rationality.* MIT Press, 1998.

[53] Ariel Rubinstein. *Economics and Language*. Cambridge University Press, 2000.

[54] Dov Samet. S5 knoweldge without partitions. *Synthese*, 172:145 – 155, 2010.

[55] T. Schelling. *The Strategy of Conflict*. Harvard University Press, 1980.

[56] E. Schwitzgebel. The stanford encyclopedia of philosophy, August 2006. URL `http://plato.stanford.edu/`. E. Zalta (ed.).

[57] N. Shah and J.D. Velleman. Doxastic deliberation. *The Philosophical Review*, 114(4):497 – 534, 2005.

[58] Y. Shoham. Logical theories of inention and the database perspective. *Journal of Philosophical Logic*, 38(6):633 – 647, 2009.

[59] Yoav Shoham and Kevin Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2009.

[60] H. A. Simon. A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1):99 – 118, 1955.

[61] M. Siniscalchi. Epistemic game theory: beliefs and types. In S. Durlauf and L. Blume, editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, Basingstoke, 2008.

[62] R. Stalnaker. On the evaluation of solution concepts. *Theory and Decision*, 37(42), 1994.

[63] R. Stalnaker. Extensive and strategic forms: Games and models for games. *Research in Economics*, 53:293 – 319, 1999.

[64] Robert Stalnaker. On logics of knowledge and belief. *Philosophical Studies*, 128:169 – 199, 2006.

[65] R. Sugden. The logic of team reasoning. *Philosophical Explanations 6*, pages 165–181, 2003.

[66] R. Tuomela. *The Importance of Us: A Philosophical Study of Basic Social Notions*. Stanford University Press, Stanford, 1995.

[67] J. van Benthem. Rational animals: What is 'KRA'? invited lecture Malaga ESSLLI Summer School 2006, 2005. URL `http://staff.science.uva.nl/~johan/publications.html`.

[68] J. van Benthem. *Logic in Games*. Texts in Logic and Games. Amsterdam University Press, 2010.

[69] J. van Benthem. *Modal Logic for Open Minds.* CSLI Publications, 2010.

[70] Johan van Benthem and Darko Sarenac. The geometry of knowledge. In *Aspects of Universal Logic,*, volume 17, pages 1–31, 2004.

[71] Peter Vanderschraaf and Giacomo Sillari. Common knowledge. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* Spring 2009 edition, 2009.

[72] Yde Venema. Algebra and coalgebra. In J. van Benthem, P.. Blackburn, and F. Wolter, editors, *Handbook of Modal Logic*, pages 331 – 426. Elsevier, 2006.

[73] J. von Neumann and O. Morgenstern. *A Theory of Games and Economic Behaviour.* Princeton University Press: Princeton, NJ, 1944.

[74] Timothy Williamson. *Knowledge and its Limits.* Oxford University Press, 2000.

[75] J. Zvesper. *Playing with Information.* PhD thesis, ILLC University of Amsterdam, 2010.

[76] J. Zvesper and E. Pacuit. Assumption-completeness in modal logic. In G. Bonanno, W. van der Hoek, and B. Löwe, editors, *Proceedings of LOFT 8*, Texts in Logic and Games. Amsterdam University Press, 2010.