The Culture of Artificial Intelligence

Value Systems for Thinking Machines

Bill Magnuson and Benjamin Gleitzman

October 06, 2008

1 Introduction

Human culture and its associated cultural norms provide a set of ideals and practices that allow for large groups of people to share common goals, behaviors, and ideas. We see these shared thoughts as useful tools to make thinking more efficient. As a participant in a particular culture, it is easier for someone to follow the norms of that culture than to rely on rational thought to justify every aspect of daily life. In this instance, the individual learns the conclusions of others, without necessarily grasping how these conclusions were derived from first principles.

"Our cultures don't encourage us to think much about learning."[4] wrote Marvin Minsky in a 1982 article for AI Magazine. More recently, Minsky has indicated that it would be useful not to provide any such means of short-circuiting rational thought in a theoretical intelligent robot to ensure that the robot will never be forcefully ignorant. Such a design would create a mind that follows no cultural norms and thus has no culture. In exploring this idea we aim to discuss the origins of culture, examine the usefulness and shortcomings of relying on culture, demonstrate that a lack of culture

will be perceived as a culture in itself, and finally discuss our conclusions which will advocate a design goal for applying culture to artificial intelligences.

2 Culture: What and Why?

2.1 Unpacking the Suitcase

Before discussing the origin of culture, we must unpack the meaning behind the word and outline what aspects of culture we intend to explore. For this paper, we will refer to culture as a general base of rules and expectations which govern, either loosely or more stringently, a group of individuals.

Using this definition, we can investigate the benefits and inadequacies of employing culture to regulate decisions in daily life. Those who rely on culture unconsciously slip into generally accepted patterns of behavior, bypassing lower level thought processes as cultural abstractions fill in the "frame" of conventional behavior. As a result of these abstracted decisions, the individual is left with more time to explore higher level thought processes and increase efficiency in real life. However, as more and more decisions are taken for granted, the individual may become trapped in a local maximum that cannot be escaped without questioning the basic rules that define the culture of that individual.

2.2 A Case For Culture

> "Religions are popular because they give people reasons not to do things, not because their advice is good, but because they allow for people to not question it." - Marvin Minsky

This bit of wisdom alludes to one of the likely reasons for the development of human culture: efficiency in decision making. By providing answers to a large number

of questions, human culture provides a way to "short-circuit" decision making and stop the questioning of values, ideals, and goals that help to determine how humans act in their daily life. The true scope of these cultural ideas is difficult to comprehend, as it imparts ideas to us about things as simple as our eating schedules and as complex as political philosophy or human rights.

One can imagine not having these normative ideas and values to guide our lives. Consider for a moment the amount of time that would be wasted each day if a person had to determine a new eating schedule and frequency based on the independent characteristics of that particular day. Also, what if one woke up each morning and re-determined whether or not it was useful to wear trousers to work? Finally, entertain the possibility of constantly reconsidering and taking up new personal ideologies about economics, politics, welfare, or education. All three of these scenarios are situations where our culture provides us with an answer (or at least a finite set of options from which to choose).

We do not intend to argue that changes do not occur as an individual participates in a culture, or even that human cultures are homogeneous. However, it is clear to us that there is a limited set of goals, values, and ideas that people follow, and that they receive this set from those around them. Most of the time, this cultural transmission happens without the receiver being aware of their newly acquired knowledge, and as such they make no attempts to find reasoning for the conclusions imparted to them.

This particular knowledge transmission that comes without questioning is certainly more efficient than using logical reasoning to arrive at conclusions (and in some situations it can even impose conclusions that are illogical for a particular person, but better for the culture itself or the determiners of such a culture). While investigating human interaction one sees that a similar transfer of information occurs when a person is in love. When one is "in love" with another person they disregard

most defects and deficiencies in that person [5]. Perhaps it is just that humans are programmed to "love" the culture in which they participate.

By extending this love analogy we can easily discover the positive and negative aspects of the culture by looking at the successes and shortcomings of people in love. On the one hand, people in love will sometimes suffer undue hardship or abuse sourced from those they love. However, for many people it is precisely the suppression of defect detectors that allow for them to lead happy lives with the companionship of others. When dealing with culture we see that people will often make decisions that are not logically correct or rational, but are rather the byproduct of cultural momentum, manipulation, or the faulty observations of others.

2.3 Universals in Culture

Human culture, variously defined, has a variety of originating sources. Some of these are consequences of local environmental effects, others a result of human universals, while some are a result of combinations of both of these. Human universals of culture include ideas such as "myths, legends, daily routines, rules, concepts of luck and precedent, body adornment, and the use and production of tools" [2]. This means that all instances of human culture have their own values for these human universals. There is often a large amount of local variance due to environmental effects, but a large portion of universals have "distinctive, even dedicated, neural underpinnings, and thus are universals of mind too" [2].

Human universals are especially interesting when applied to ideas of culture for an artificial intelligence because it suggests that in some situations our culture has a causal relationship with the physical structure of our mind or the prevalence of certain genetic traits. Brown notes that for aspects of culture that are constant through societies, there must be something that does not change depending on the

environment.

Thus, any attempts at a coherent robot culture would require an unchanging nature across all instances of robots. Practically, this seems an impossible goal as local variations in human culture will undoubtedly bias particular higher level components of "robot nature." Whether or not a constant robot nature is something with which the developers of an artificial intelligence should concern themselves is a separate discussion, but if the pervasive theme of human unwillingness to trust robots in science fiction writing is any indication of future events, it is certainly a topic that requires further attention.

Specific examples of this causal relationship exist in human societies. One of these is a nearly universal usage of the right hand for ceremonial purposes, probably due to the genetic dominance of right handedness. A related example comes from the names that most cultures have given to the pupil of the eye; in a large number of unrelated languages the name given is the same as one given to a little person. This is likely due to the propensity to recognize a small mirrored image of oneself in the eyes of others and highlights the role that universal experience has in creating universals across societies [2].

2.4 The Role of Technology

Necessity is said to be the mother of innovation, and innovation inevitably leads to changes in the way that we carry out our lives. Indeed, the incredibly fast advances in human society that have led to the divergence from our primate ancestors cannot be adequately explained by evolution alone. We must instead look to cultural transmission as the catalyst and here we see the ratchet effect of technology [9].

As a particular society advances and a new technology is developed, it can easily spread throughout that society by making more efficient the accomplishment of a

particular task. In this situation we find an example of changing culture. In order for a new technology to be adopted, an established methodology must be considered inefficient or otherwise less desirable, and that original assumption must change to include the new technology.

Furthermore, when technology spreads from one culture to another its effects are varied but largely homogenizing. In many cases, the specific use of a technology can cause the displacement of a large variety of previous solutions that existed across cultures. A good example of this is found in agriculture: despite the large variety of farming methods that existed centuries ago, most modern farming is done in a similar manner due to the development of technologies such as plows, tractors, pesticides, or harvesting cycles.

We do not mean to assert that the adoption of a particular tool or technology is completely homogenizing. In many cases a newly introduced technology is "adopted to the social processes of the adopting society, and not vice-versa" [7]. Examples of this are seen in the primitive Maori society's adoption of simple farm tools such as hoes and spades. After initially ignoring the new technologies and following the momentum of their previous ways, they eventually incorporated the new artifacts into their society. However, instead of using the hoes and spades as the Europeans did, the Maori incorporated the idea behind the hoe into their agricultural practices. They did this by attaching the metal end of the hoe to a shorter stick so that it could be used in the familiar squatting position while work was being done [7]. However, despite this localization, the Maori would still become dependent on European sources for the parts of the tools that they decided were worthwhile [6]. This foreign dependence required further interaction between the two cultures, and thus suggests a continuing mode of culture transfer that will ultimately result in cultural convergence.

This is interesting for our purposes because not only will robots be a highly disrup-

tive technology when they become an integral part of our cultures, but a robot with an artificial intelligence will be continuously exposed to existing and new technologies that could change their own culture. One can imagine a newly formed culture that does not have strong ties to any existing culture quickly adopting a diverse sampling of other cultures. However, since many contradictions exist between cultures, the robot would be challenged to decide which values or goals to adopt. Indeed, Brown agrees with this idea in his discussion of human nature by pointing out that when adaptations conflict with each other in some circumstances, the resulting adopted behaviors are compromises [6]. This idea relates well to the critic-selector model which Minsky has proposed in that there are higher level mental processes which are deciding on a best course of action for the brain to take [5]. This would suggest that a robot would likely benefit from a model similar to the one proposed by Minsky in order to deal with conflicting observations and value sets.

3 Inspecting Human Culture

In order to support our recommendations for the culture of a thinking machine, we draw upon the previous work of social psychologists. By studying the outcomes of experiments in human cultural studies, we can gain insight into possible gains and pitfalls in developing a robot culture.

3.1 The Ten Values

On a pan-national scale, media and news sources often report the striking differences between nations, races, and cultures. At the end of the twentieth century, work by psychologists on value priorities suggested that within and across cultures, individual choices for value hierarchies differed greatly [8]. These differences were attributed to

diverse genetic heritage, personal experiences, and differing social locations, among others.

However, the 2001 work of Schwartz and Bardi on cross-culture value similarities suggests that there indeed exists an underlying structure upon which humans base their cultural decisions [8]. If this is the case, these values can be enumerated and captured for use in a thinking machine.

In order to provide a level surface on which to define cross-cultural values, researchers Schwartz and Sagiv surveyed members of 63 nations in order to explore the existence of core cultural values. From a pan-national set of responses in various languages, the researchers postulated the existence of 10 motivationally distinct value types, drawn from universal requirements of the human condition. The 10 value types are listed below, with a short description of the sub-values encompassed by that value.

- Power (social power, authority, wealth)
- Achievement (successful, capable, ambitious)
- Hedonism (pleasure, enjoying life)
- Stimulation (daring, a varied life, an exciting life)
- Self-direction (creativity, freedom, independent)
- Universalism (broad-minded, wisdom, honest)
- Benevolence (helpful, honest, forgiving)
- Tradition (humble, devout, respect for tradition)
- Conformity (politeness, obedient, self-discipline)
- Security (family security, social order, reciprocation of favors)

With these values defined, further research was conducted to determine if similar value hierarchies exist across nations.

3.2 Similarity of Pan-Cultural Values

In a large scale experiment spanning 54 countries in which teachers and students were asked to rank and give a rating to the 10 core values, evidence suggests that there does indeed exist a baseline ranking of cultural values [8]. While value ratings differed between groups, rankings for both teachers and students listed benevolence, self-direction, and universalism as the three most important cultural values (in that order). The two least important cultural values by rank, tradition and power, were also the same between teachers and students. While the middle set of five values tended to differ between students and teachers (this may be due to general status of life differences between teachers and students, with students favoring hedonism and teachers favoring security), between nations there exists an extremely similar ranking of the 10 core values.

Aside from the revelation that humans might not be as culturally separate as we have been led to believe, these findings are important for a number of reasons. As mentioned earlier, the existence of a pan-cultural value ordering is strong evidence for an underlying structure that influences human values. Arguably, the existence of this structure is self-fulfilling, since humans with values that clash with that of survival, or the values of other groups of humans, may have a lower fitness and have been subsequently weeded out by evolution. Nonetheless, this structure can be enumerated, captured, exploited, and leveraged by those wishing to build a thinking machine.

Furthermore, while this value ordering is not the same across all cultures, it does provide a baseline from which to compare nations across the globe. If we wish to have the culture of our thinking machine adopt the values of the region in which it resides and operates, it is worthwhile to have a means of evaluating that culture in regard to the pan-national average.

All of these findings should be considered in the creation of a thinking machine. However, now that a generic ranking of pan-cultural values has been established, we turn to the challenge of imparting culture between robots.

4 Imparting Culture

In order to explore the process of cultural transmission and change for a thinking machine, we again turn to the work of social psychologists. Although the generic studies of parent-child value transmission may not be completely applicable for a robot culture, it again provides a useful baseline for comparing alternatives within cultural transmission.

In order to combat what Klaus Boehnke viewed as the failings of studies in value transmission and change, his 2001 paper aimed to correct and explore alternatives to traditional value studies [1]. In this context, value transmission generally focuses on the ways values are imparted from one individual or group to another, while value change explores the in values and ideals over time.

4.1 Value Transmission

According to Boehnke, classical research in value transmission focused too heavily on the interaction between parents and children, neglecting the influence of societal change. In a nutshell, value change and intrafamilial value transmission lost focus, leaving researchers to explore only the interaction between parents and offspring. For the creation of a thinking machine, this distinction becomes paramount as maternal and paternal roles becomes less defined, and societal change processes take a more influential role. The social change process, especially the influence of the *zeitgeist* or "spirit of the times," is worth noting in relation to influence exhibited by a parental

figure.

Experiments into value transmission, taking into account the *zeitgeist* as well as intrafamilial forces and using the 10 core values as a measure, concluded with a number of interesting results. As expected, similarity between highest ranked parental and child values differed greatly. This is most probably the result of the two entities existing within different stages of development. However, when looking at the lowest ranked values, transmission from parents to offspring was the highest. The four value types least preferred by fathers and mothers (hedonism, stimulation, tradition, and power) have the highest correlation between parents and offspring [1].

From this observation, we may wish our robotic culture to embody the idea of negative feedback. Between generations, the most highly ranked cultural values can differ substantially, but in transmitting values from one robot to another – if that is the course ultimately chosen for value transmission between robots – we may want to encourage lower ranked values to remain unfavored. In this way, the process by which one might avoid the mistakes or value choices that were found to be negative for one's parents can be transferred to a robotic intelligence.

4.2 Value Change

Classical studies into value change also fell victim to similar pitfalls as value transmission studies. According to Boehnke, many perceive value change in Western societies as "a more or less automatic consequence of increased socioeconomic prosperity." In other words, rich families tend to raise fewer materialistic youth that poor families [1]. Again, intrafamilial value transmission was neglected while studying value change over time, and families were often evaluated as units, rather than separate individuals within a group.

In studying value change, we see similar results as that of value transmission.

Power and tradition rank in the last two spots for both genders and across generations, while hedonism and self-direction, ranked low by the mothers, are highly ranked for male offspring.

From this information, we can suggest that our robotic intelligence may wish to draw a distinction between generations and ages. Might we want younger robots to "rebel" against their older counterparts, placing hedonism and self-direction as higher values? Or, should we create a robotic culture that avoids this value change, and instead transmits the "adult" values directly from one individual to another?

Additionally, research by Boehnke has suggested that the proportional influence of parental values on their offspring might not be as strong as expected. The effect of the *zeitgeist* upon the values of both offspring is important to consider when dealing with value change over time. This research suggests that major players contributing to the percentage variation in offspring value preferences are mother, father, socioeconomic status, as well as the values of the times. It is interesting to note, however, that maternal and *zeitgeist* influence are strongest, with ratings of 2.8, followed by the father (rating of 1.5) and socioeconomic status (with a rating of 0.9) [1].

Ignoring the intricacies of the rating scale, we can see that if we are to follow human value change as a model in the development of robot culture, we may wish to impart the notion of the current state of cultural values in that region.

4.3 Ways to Impart

Humans have the ability to share and impart values in many diverse ways, both within a group and between individuals. While robots may not be currently imbued with such a ability, it is worth postulating upon the ways in which a thinking machine might impart culture between "generations" of robots.

Building upon archaeological evidence to explain human culture change, Eerkens

and Lipo have suggested three ways in which culture is imparted between individuals. Building upon these notions, the researchers then suggest general methods by which variation in cultural values may arise during the process of value transmission. First, we will focus on the process of transmission.

The three modes of transmission suggested by Eerkens et. al include the categories vertical, oblique, and horizontal [3]. Vertical transmission is characterized by direct transmission from a parent to an offspring. In the case of traits being copied from an individual, these rules could be filled by the parent and child. For traits copies from a subset of individuals, this could be a child acquiring the traits of both parents. Vertical transmission does not occur from a population to an individual.

Oblique transmission is again classified by direct transmission, but also begins to include conformist transmission when dealing with groups. In conformist transmission, individuals conform to the average value for the entire previous generation [3]. Between individuals, this can be seen as a student acquires values from a teacher. Within a subset of individuals, this includes conformist transmission from a group to an individual. With oblique transmission, we can begin to see the effect of a population's values imparted to an individual, such as the case where values are imparted from a parental generation to an individual.

The last method, horizontal transmission, includes hints of the conformist transmission for large populations, but is more prestige-based between groups and individuals. This type of transmission can occur when an individual values the traits of another individual who has been deemed "notable," or between specific peer groups and an individual.

Currently, there is no structure by which a robotic intelligence might base such judgments as prestige, but such a system might be helpful in creating an organism that can support this type of value transmission.

4.4 Variation in Culture

Variation in culture can happen through a variety of mechanisms. Some of these include faulty observations, such as the questioning of preconceived notions or errors in learning. Any of these conditions can cause local variations in culture, some of which can become permanent and eventually create its own sub-culture (or in extreme cases become a culture of its own by consuming its parent).

Perhaps the most likely way to create temporary variations of culture is through faulty observation. This mechanism can manifest itself in two different ways. First, the a person could perceive a cultural norm incorrectly and decide to follow the misconception. This particular error could be easily fixed with future observations, but in some cases people will fail to see the differences or cling to their original interpretation.

Another is through the creation of a new norm that could possibly be transferred to others once it has been established. Even when a person is developing their goals and ideas using rigorous logic and rationality, they might start with false assumptions and thus arrive at incorrect conclusions.

In addition to the variance caused by subconscious acceptance of faulty observations, culture can also be changed through self-reflective thought. If a person discovers that a cultural norm they follow is likely irrational, they have the facility to question it and reform the idea. If the resulting change in culture is particularly persuasive to others this can result in cascading changes throughout a culture. This effect has been seen throughout history by a variety of great leaders and thinkers and has resulted in some of the most pervasive cultural, political, and social shifts in human history.

5 Investigating Conclusions

5.1 No Culture

As we can we from our previous examples, a total lack of culture is infeasible if we expect any reasonable level of development for our robots. Although we may be able to escape the problem of local maxima, the negatives attached to a "culture of no culture" are too great to employ this methodology. In building an intelligence that is always questioning, we cripple our ability to learn from others. Such a system is analogous to performing a breadth first search with a nearly infinite branching factor.

5.2 Asimov's Three Laws

At first glance, a simple set of laws such as those originally proposed by Asimov would seem sufficient to direct the behavior of robots. However, such a simplistic view neglects value transmission, change, development, and other components that we have demonstrated to be required for the functioning of human culture. We apply the analogue of human culture on the reasoning that any successful artificial intelligence will have to co-exist within our cultural landscape and thus will have to develop a selection of the same universals. In order for this to occur, a robot artificial intelligence would require many of the same mechanisms that our own brain uses to develop culture. It must also have a starting point consisting of goals and driven by inherent needs.

5.3 Robot Learning

Considering the concepts discussed in this paper, we strongly recommend the creation of robot culture that employs human cultural development and transmission as a backbone, yet leverages the advantages of a robotic intelligence. A successful culture

should allow for transmission of values between "generations" of robots, yet also provide for the copying of such an intelligence from one robotic individual to another.

We stress that robot culture should be dynamic, and able to change given observational evidence that would suggest to a robot that the culture it currently employs is non-optimal. This ability to switch values and ideals is necessary for any level of success to be achieved when integrating artificial intelligences into human populations.

We are not yet to a point in the development of a robotic intelligence where many of these recommendations can be applied, but the suggestions and considerations set forth by this paper can serve as a starting point for further research. While the future of robotic cultural development is bright, we hope that the lessons learned from millennia of cultural experimentation performed unknowingly by humans is not ignored.

References

[1] Klaus Boehnke. Parent-offspring value transmission in a societal context: Suggestions for a utopian research design – with empirical underpinnings. *Journal of Cross-Cutural Psychology*, 32, 2001.

[2] Donald E. Brown. Human universals, human nature & human culture. *Daedalus*, 133(4):47–54, 2004.

[3] Jelmer W. Eerkens and Carl P. Lipo. Cultural transmission, copying errors, and the generation of variation in material culture and the archaeological record. *Journal of Anthropological Archaeology*, 316, 2005.

[4] Marvin Minsky. Why people think and computers can't. *AI Magazine*, 3, 1982.

[5] Marvin Minsky. *The Emotion Machine, Commonsense Thinking, Aritificial Intelligence, and the Future of the Human Mind*. Simon and Schuster, New York, NY, 2006.

[6] B Pfaffenberger. Social anthropology of technology. *Annual Review of Anthropology*, 21(1):491–516, 1992.

[7] W. Schaniel. New technology and cultural change in traditional societies. *Journal of Economics*, 22:496–498, 1988.

[8] Shalom H. Schwartz and Anat Bardi. Value hierarchies across cultures: Taking a similarities perspective. *Journal of Cross-Cutural Psychology*, 32, 2001.

[9] Michael Tomasello. *The Cultural Origins of Human Cognition*. Harvard University Press, 1999.