From Artificial Intelligence to Artificial Culture

M. Afzal Upal

Electrical Engineering & Computer Science Department, University of Toledo
Psychology Department, Bowling Green State University
Mail Stop 308, 2801 W. Bancroft Rd, Toledo, OH, 43606
afzal@eecs.utoledo.edu

Abstract

Historically, cognitive sciences in general and artificial intelligence in particular have had little to say about human culture. However, if the modest success of nascent field of cognitive science of culture in explaining aspects of culture is any indication, further contributions from cognitive scientists may be needed to develop a predictive computational science of culture. This paper outlines a number of promising directions along which artificial intelligence researchers can contribute to this exciting new field.

Introduction

> Why call this "Artificial Culture?" Yes, we know culture is artificial by definition, but that's not the point we're making. We are referring to the trajectory in advanced computation from "Artificial Intelligence" to the field of "Artificial Life. " We are projecting that trajectory from 'Artificial Life,' to "Artificial Culture". (Gessler 2005)

Cognitive sciences have traditionally focused on individual cognition at the expense of sociocultural processes and their relationship to an agent's cognition (Sun 2001). Social scientists, on the other hand, have traditionally ignored the contributions of an individual's cognition on the formation and spread of cultural information. There has been a growing realization by social scientists that social phenomena such as culture and religion must be explained in terms of individual cognitive and ecological factors (Boyer 1994; Sperber 1996) not in a never ending cyclical chain involving other social factors. This has been accompanied by a growing dissatisfaction with verbal models as being too imprecise and mathematical models as being too rigid and unable to be scaled up. As economist Scott Moss recently lamented, "in more than half a century since the publication of von Neumann-Morgenstern (194x), no significant progress has been made in the development of models that capture the process of interaction among more than two or three agents" (Moss 1998). An alternative is to build bottom-up algorithmic models of socio-cognitive processes to untangle the causal relationships between cognitive tendencies of individuals and emergent patterns of shared beliefs that result from interactions among population of individuals with these tendencies. By relating algorithmic descriptions of individual cognition developed by cognitive scientists with the population level social simulations, the cognitive science of culture aims to show that culture is algorithms all the way.

An issue that all simulation researchers have to wrestle with is selecting the model's level of complexity. If a model is too complex, it loses its value as a *model* because it is either intractable and/or its results are just as complex to understand as those of the real world phenomena being modeled. However, if a model is too simple, the effort spent in designing and programming the computer simulation is often not worth it because the consequences of a simple model can be readily foreseen without constructing it and the model has little to tell us about real world phenomena of interest. In practice, researchers often do not fix a level of complexity at the start of the simulation. Instead, they typically start with a simple model, run the simulations, analyze the results and refine the model by adding/modifying agent features or agent interaction rules, rerunning the simulation and comparing the result of the change. This closed loop development between model construction, simulation, and analysis is key to the success of bottom-up agent-based models as social theory development tools. I believe that the technique of building simpler to complex models, analyzing the results of each change to understand the causal relationships between micro-level changes and the emergent changes in the macro-level phenomena can be exploited to build computational models of the creation and spread of cultural representations.

However, agent-based social simulation models developed to date have employed simple models of cognition. For instance, most existing social simulation systems models of social belief change model agent-beliefs as a single bit (Bainbridge 1995; Doran 1998; and Epstein 2001) and belief change involves flipping the bit from 0 to 1 or vice versa often to match the belief of one's neighbors. This means that agents in such societies can never form complex belief systems that characterize real world cultures and hence such simulations can tell us little about human cultures. Further, there is considerable evidence to suggest that the ability to produce and comprehend natural language narratives is crucial to the formation and propagation of cultural knowledge. Being able to generate and comprehend natural language requires rich knowledge representation and deep reasoning capabilities including planning, knowledge acquisition, belief revision, and comprehension. The new cognitive science of culture investigates how cultural contexts affect individual thought and how individual cognitive tendencies result in the emergence of cultural patterns.

I believe that artificial intelligence can play a central role in the new cognitive science of culture by not only providing agent-based computer modeling tools that allow social scientists to explore and test their theories but also by providing a level of abstraction that is removed enough from case studies of particular sociocognitive processes to allow one to see the important similarities between them and discover patterns invisible to the social scientists. This approach can allow us to ask and answer new questions. I illustrate this approach with the help of the analysis of communication between agents- a process whose understanding is key to the understanding of the creation and spread of cultural ideas.

Human Communications

> Communicate. We humans do it all the time, and most of the time we do it as a matter of course, without thinking about it. We talk, we listen, we write, we read - as you are doing now - or we draw, we mimic, we nod, we point, we shrug, and, somehow, we manage to make our thoughts known to one another… compared to other living kinds, we are amazingly good at it. Other species, if they communicate at all, have a narrow repertoire of signals that they use to convey again and again things like: "this is my territory," "danger, run!" or "ready for sex."
>
> (Sperber 1995)

> One of the basic causes for all the trouble in the world today is that people talk too much and think too little.
> (Margaret Chase Smith)

> People say conversation is a lost art; how often I have wished it were! (Edward R. Morrow)

As Sperber and others have noted, people appear to be eager communicators. They appear eager (sometimes too eager) to speak and eager to comprehend (sometimes so eager that they see patterns where there really aren't any). Editors of any mailing list, newsgroup, newspaper, magazine or tv/radio call in show know that regardless of the topic, there are always more people willing to speak for longer than the limited resources allow. Given the fact that it is hard to find any two humans within a close range of each other not communicating with each other in one form or another, it is surprising how little work has been done to investigate the reasons as to why that is.

- Why do people appear to be eager speakers?
- Why do people appear to be eager listeners?

As Richerson and Boyd (1985) argue evolution is the ultimate answer to why questions about human mind and behavior. The problem for evolutionary scientists is to uncover the computational advantages that a behavior confers upon its practitioners as compared to the non-practitioners. In this light, the above questions can be reframed as:

- What evolutionary advantages did the eagerness to speak confer upon our evolutionary ancestors?
- What evolutionary advantages did the eagerness to listen confer upon our evolutionary ancestors?

While such questions are certainly interesting in their own right they are also interesting because answers to such questions are needed in order to build a cognitive science of culture that relates pragmatics of individual communication to the creation and spread of culture allowing us to build a natural history of various cultural ideas. Such questions should be of interest to evolutionary theorists who propose that in order for ideas- or memes- to become cultural i.e., widely shared by a population, they have to have something about them that allows them to replicate themselves more effectively than other ideas.

Because of the notorious complexity of the vast number of interacting variables involved, it is best to abstract out the key processes as much as possible and analyze them as computational processes. The problem solving paradigm of Newel and Simon (1972) provides an effective abstraction. This view characterizes the environment inhabited by our evolutionary ancestors as a problem solving environment which frequently called upon its inhabitants to make accurate predictions in order to secure resources, to safely negotiate their way, and to avoid falling prey to various predators.

In a dynamically changing environment, the adaptive agents that can modify their problem solving strategies and their world model have an advantage over their competitors in terms of being able to better predict their environment and thereby obtain access to more resources and pass on their strategies and learned knowledge to their descendents. Such adaptive agents should not only learn from their own interaction with the world but they should also be open to learning from others when opportunities arise. However, this raises several problems such as how to evolve a common language with a commonly agreed upon syntax and semantics. Much of the work on computational pragmatics (Bunt & Black 2000) addresses these questions. However, even if we assume that the communicating agents share a common language, a listener still needs to engage in an effort to try to comprehend the information provided by a speaker. This effort has significant computational costs such as the cost of understanding the meaning of the message and the cost of updating the agent's world model in light of the information derived from the message. The agent has to devote precious time and cognitive resources to processing the message. Such preoccupation can result in immobility potentially costing an agent its life. Therefore, a rational agent should only engage in such an effort if the estimated benefits of such an effort outweigh its estimated costs. Thus a speaker should compare the estimated benefits of performing the utterance action to communicate a piece of information with the cost of performing such an action to decide whether to speak or not. A listener should compare the estimated benefits of processing a message with estimated costs of processing it before actually comprehending it.

While some have advanced alternative architectures (Brooks 1991), most cognitive scientists believe that a problem solving agent has to build internal representation of its environment (the so called *world model*) to be successful in solving various problem posed by its environment. A message received from another agent is beneficial to such an agent if it leads to a more predictive model of the world.

The problem however, is that before comprehending the meaning of a message an agent has little information available to estimate the costs and benefits of processing a message. This suggests a multi-phased approach. A rational agent should exploit all the information at its disposal when it receives the message (without actually processing the message) to estimate the benefits and costs of processing the message and start the first phase of processing if it believes that benefits outweigh the costs. This information could include the context in which the listeners receives the information and the mental state of the listener. On receiving two messages simultaneously, a rational listener should prefer to process the message received from the more reliable and trustworthy agent. Similarly, an agent should be less willing to process a long message if all its cognitive resources are being consumed in a potentially deadly fight with a predator. If the benefits of engaging in the first phase of comprehension outweigh the costs, the agent should start the first phase.

At the completion of the first phase, however, a rational listener should use the newly derived information to compute the estimates and costs of further processing and only continue if benefits of further processing outweigh the its costs. An agent can use the information derived during the process of natural language understanding (this includes both the meaning of the message (the objective of natural language understanding) and also any ancillary information the agent may obtain as by-product of comprehension such as the background knowledge that is activated during the meaning computation process to compute a better estimates of the benefits and costs of further processing the message to revise the agent's world model in light of the message.

A message that does not lead to any improvements in an agent's world model would be of little benefit to an agent. For instance, a message that conforms to all the agent's expectations and hence will not likely lead to any changes in the listener's world model should be discounted for further processing. Similarly, a message that violates so many expectations that the agent cannot make any sense out of it should also be discounted. A message in the intermediate range, however (i.e., a message that does not conform to all expectations but can be assigned a meaning with some effort) should be estimated as likely to provide significant new information i.e., result in *information gain.*

If we define *postdictability* of a message as the ease with which a listener can compute the meaning of that message and *predictability* of a message as the ease with which the listener can create that message using its existing world model then the difference between the two quantities give us an estimate of the benefit of further processing the message to revise one's world model.

Our model predicts that rational listeners should bias their comprehension process to process those messages maximally that have the largest difference in postdictability and predictability values. Over the last few years we have performed a number of experiments with human subjects to see whether people's comprehension of various concepts corresponds to the predictions of the rational listener model outlined above (Upal 2004; Upal 2005; Owsianiecki, Upal, Slone, & Tweney 2006; Tweney, Upal, Owsianiecki, Slone submitted; Slone, Upal, Owsianiecki, Tweney, submitted). Our findings are similar to those of a number of earlier studies (Boyer & Ramble 2001; Barrett & Nyhoff 2001; Atran 2003) that show that people better remember minimally counterintuitive concepts as compared to the intuitive and maximally counterintuitive concepts. Boyer (1994) has argued that this helps explain as to why minimally counterintuitive ideas are so widespread among the world religions in particular and in popular culture in general. Minimally counterintuitive ideas are so prevalent in emergent cultural phenomena because they are preferentially processed by individuals.

This appears to be obviously true if everything else is equal. As students of culture are painfully aware everything else is never equal in the real world. The complexity of culture, in part due to a large number of interacting factors, has so far thwarted any meaningful progress towards a detailed predictive model of the spread of cultural ideas. Besides counterintuitiveness there are a number of other factors that are also known to impact memorability of an idea such as imagery, and concreteness (Sadoski and Paivio 2001), and the emotional state of the comprehendor, to see how they may be impacting memorability of intuitive and counterintuitive concepts. Imagery and concreteness are known to be positively correlated with each other as well as with memorability. However, it is not known how these variables relate to counterintuitivess and how imagery, concreteness, and counterintuitivess jointly relate to memorability. Some work in progress (Slone, Owsianiecki, Upal, Tweney, & Edwards, in preparation) suggests that counterintuitiveness may insulate a concept from the usual effects of imagery. Moreover, a number of other factors besides memorability are though to have an impact on how well an idea spreads in a population (Atran 2003). For instance, Bainbridge and Stark (1987) argue that unconfirmable ideas have an advantage in terms of being believed by individuals and spreading in a population.

I believe that cognitively-rich agent-based social simulation models can be helpful in understanding the relationship between the myriad of factors thought to affect formation and spread of culture. To this end, we have designed a testbed multiagent society that inhabits a version of Russell & Norvig's (1995) Wumpus World (WW). Our version of WW has 100 (10 x 10) cells and is inhabited by ten agents randomly placed at ten cells at the start of the game. Agents navigate the world by moving one cell at a time horizontally or vertically. An agent dies if it falls into a pit and another agent is created to take its place. Agents like visiting cells with treasure in them and try to avoid falling into pits which emit smell that can be perceived in the neighboring cells. Agents perceive their environment and reason with it to build and refine their world model. The world model, in the current version, only contains the cell locations they believe contain pits or treasures in them. On perceiving the smell, they assume that all the neighboring cells (except the cells they've already visited) contain a pit in them. On hearing an agent's 'yipee', they assume that all the neighboring cells (except the ones they've already visited and do no contain treasures) have treasures.

Thus agent models may contain inaccuracies both about the locations of the pits as well the locations of the treasure. However, there's an asymmetry between the treasures and pits. While the agents prefer to travel towards a cell where they expect to find treasure, they avoid cells they believe contain pits in them. The result is that the agent never have the ability to revise their beliefs about the presence of pits through their own observations. This results agents abandoning their beliefs about phantom treasures as they travel more and more though WW but they are never able to do so for their beliefs in phantom pits. Our results show that the proportion of false pit beliefs to the total number of pit beliefs remains constant regardless of how much the agents travel in the world. This seems to be similar to what is predicted by Bainbridge and Stark (1987)'s theory of religion, namely, that untestable beliefs have cognitive advantages and this leads to more widespread new religious movement beliefs being of this variety.

We plan to carefully add reasoning and knowledge representation capabilities to our testbed evaluating the results of each change to discover causal relationships between cognitive and ecological factors and emergent social phenomena.

Conclusion

This paper illustrates our strategy for building a computational science of culture. It involves analyzing the interaction of knowledge among a society of rational problem solving agents aided by agent-based social simulation of cognitively rich agent societies. This is to be done in conjunction with the study of individual and group human behavior to find the points in the space of artificial cultures where human cultures lie. I believe that this approach has the best hope of leading us to a predictive computational model of the formation and propagation of cultural information.

References

Atran, S. 2002. *In Gods We Trust: The Evolutionary Landscape of Religion*, Oxford, NY: Oxford University Press.

Bainbridge, W. 1995. Neural Network Models of Religious Belief , *Sociological Perspectives*, 38: 483-495.

Bainbridge, W. & Stark, R. 1987. *A Theory of Religion*, New York: Lang.

Barrett, J. L. & Nyhof, M. (2001). Spreading Non-natural concepts: the role of intuitive conceptual structures in memory and transmission of cultural materials. *Journal of Cognition and Culture,* 1, 69-100.

Boyer, P. (1994). *The Naturalness of Religious Ideas: A Cognitive Theory of Religion*, Berkeley, CA: University of California Press.

Boyer, P. & Ramble, C. (2001). Cognitive templates for religious concepts. *Cognitive Science*, 25, 535-564.
Brooks, R. A. (1991), Intelligence without representation, *Artificial Intelligence*, (47), 139–159

Bunt, H. & Black, W. 2000. Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics, Amsterdam: John Benjamins Publishing Co.

Castelfranchi, C. 2001. Towards cognitive memetics: Socio-cognitive mechanisms for meme selection and spreading, *Journal of Memetics - Evolutionary Models of Information Transmission*, vol. **5**.

Doran, J. (1998). Simulating collective misbelief, *Journal of Artificial Societies and Social Simulation*, 1(1).

Epstein, J. & Axtell, R. (1996). *Growing Artificial Societies: Social Sciences from the Bottom Up*, Washington, DC: Brookings Institute Press.

Epstein, J. (2001). Learning to be thoughtless: Social norms and individual computation, *Computational Economics*, 18(1), 9-24

Gessler, N. 2005. Lecture Notes for UCLA Anthropology 137, http://www.sscnet.ucla.edu/geog/gessler/2005-acult/7

Newell, A. & Simon, H. (1972). *Human Problem Solving*, Englewood Cliffs, NJ: Prentice-Hall.

Owsianiecki, L., Upal, M., Slone, J., & Tweney, R. (2006, in press). The role of context in the recall of counterintuitive concepts. *Journal of Cognition and Culture*.

Richerson, P.J. & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago: University of Chicago Press.

Sadoski, M. & Paivio, A. (2001). *Imagery and text: a dual coding theory*. Mahwah, NJ: Erlbaum

Sperer, D. 1995. How do we communicate? in John Brockman & Katinka Matson (eds) *How things are: A science toolkit for the mind*. New York: Morrow, 191-199.

Sperber, D. 1996. *Explaining Culture: A Naturalistic Approach*, Malden, MA: Blackwell Publishers.

Sun, R. 2001. Cognitive science meets multi-agent systems: A prolegomenon. *Philosophical Psychology*, Vol.14, No.1, pp.5-28.

Tweney, R.D., Upal, M.A., Owsianiecki, L., Slone, D.J. & Edwards, K. (in press, 2007). The creative structuring of counterintuitive worlds. *Journal of Cognition and Culture*, vol. 7.

Upal, M. A. (2005b). Simulating the Emergence and Maintenance of NRMs, in *Journal of Artificial Societies and Social Simulation,* 8(1).

Upal, M. A. (2004). Computational Modeling of the Acquisition of Counterintuitive Concepts in *Proceedings of the Fifth International Conference on Cognitive Modeling*, pages 392-393, Mahwah, NJ: Lawrence Erlbaum