Computational Models for Belief Revision, Group Decision-Making and Cultural Shifts

MURI 05 THIRD YEAR REPORT

AFOSR Grant No. A9550-05-1-0321 (Start Date: 1 May 05)
PRINCIPAL INVESTIGATOR: Professor Whitman Richards, 617-253-5776, wrichards@mit.edu
LEAD INSTITUTION: Massachusetts Institute of Technology
AFOSR PROGRAM MANAGER: Dr. Terence Lyons, 703-696-9542, terence.lyons@afosr.af.mil

FORTHCOMING GOVERNMENT REVIEW AND GOVERNMENT PARTICIPANTS:
17 Dec 2007 -- at the Massachusetts Institute of Technology. A one-day technical review to be attended by the review team of Dr. Kenneth Boff (formerly Chief Scientist AFRL/HE, currently consultant to AOARD), Dr. Rebecca Goolsby (ONR Program Mgr. & Cultural Anthropologist), Dr. Fariba Fahroo (AFOSR Program Mgr. Computational Mathematics), Dr. Jun Zhang (AFOSR Program Mgr. Cognition and Decision-Making & Human-Machine Interaction), Dr. Terence Lyons (AFOSR Program Mgr. Socio-Cultural Modeling.)

PROGRAM OBJECTIVE: The primary objective of this MURI is (1) to explore how beliefs support and lead to certain actions in one culture but not another, and (2) to develop computational models that further our understanding of the relation between beliefs, decisions, and actions. A key requirement for such models is to distinguish the different roles played by sacred values versus instrumental or secular values, which differ widely across cultures. The goal of these models will be to provide formal explanations for how the beliefs of individuals affect group and individual actions, and how groups evolve. Such models are an important step toward understanding and predicting the dynamics and actions of groups. They are fundamental to an understanding of how actions of a group may be altered by belief revision, by either internal or external pressures (including force). They also will be needed for strategic reasoning in negotiations, where beliefs in different cultures may lead to what appears to be irrational proposals, yet are seen as rational in that competitive culture. This MURI initiative supports DoD's urgent need to better understand how terrorist cells are created and how they evolve, thus providing models for the underlying seeds, which can be more easily identified early in the development process. In turn, the models may suggest countermeasures that will help revise or manipulate belief structures to reduce the likelihood of militant cells forming within a culture.

MURI CONSORTIUM RESEARCH TEAM MEMBERS (original members):

- **Massachusetts Institute of Technology** (PI) Whitman Richards, Prof. of Cognitive Science,
  Computer Science and Artificial Intelligence Laboratory (CSAIL),
  Joshua Tenenbaum, Assoc. Prof. of Computational Cognitive Science,
  Patrick Winston, Prof. of Computer Science. CSAIL
- **Harvard University,** Avi Pfeffer, Assoc. Prof. of Computer Science
- **University of Michigan,** Scott Page, Prof. of Economics & Complex Systems,
  Jenna Bednar, Assoc. Prof. of Political Science
- **Northwestern University,** Kenneth Forbus, Prof. of Computer Science,
  Douglas Medin, Prof. of Psychology
- **CUNY, John Jay Center for Terrorism & Univ. Mich.,** Scott Atran, Prof. of Anthropology,
  Psychology and Public Policy
- **University of Texas**, Brian Stankiewicz, Asst. Prof. of Psychology (now at 3M, Minneapolis,)
- **Consultants:** Robert Axelrod, Prof. of Political Science and Public Policy, University of Michigan
  Marc Sageman, MD, PhD; Sageman Consulting LLC; Univ. Penn. Adjunct Prof., Dept. Psychiatry

FINANCIAL EXECUTION: (as of 1 Nov 07)
3-YEAR BASE PERIOD – FY05: $669,604 (5 months) **FY06:** $1,139,405 **FY07:** $1,183,188 & FY07: 71,181 (additional funding) **FY08:** $551,821 (7 months) & FY08: $73,878 (anticipated supplement for Northwestern)
2-YEAR OPTION PERIOD - FY08: $718,743 (5 months) **FY09** $ 1,26,255 **FY10:** $604,483

To-Date (27 Nov 2007) Base Period funds of $2,693,378 have been expended and are "off the government's book." This figure takes into account outstanding invoices from the subawardees and consultants through the period 31 Oct 07, but NOT the $625,699 for the final 7 months of the current option period, which expires 31 May 08. The current Base Period amount includes $2,992,197 in original MURI funding as well as a supplement of $71,181 bring the total for the current Base Period to $3,063,378. This does not account for the FY08 partial increment of $625,699. The current commitments for the period 1 Nov 07 through 31 May 08 are estimated to be $620,000, leaving a balance of $375,000. The under-expenditures arise principally from (1) the June 07 departure of Prof. Brian Stankiewicz from the University of Texas (leaving a projected $100,000 unexpended at 31 May 08) and (2) the Stanford sabbaticals of Prof. Page and Prof. Bednar (leaving a projected $250,000 total unexpended as of 31 May 08). Part of the under-expenditures arose because many of the students were exceptional, and obtained independent fellowship support. Our proposal for these remaining funds is to add two additional subcontractors to the MURI: (1) Prof. Partha Niyogi, Computer Science, University of Chicago to collaborate with Prof. Berwick & Richards at MIT, bringing expertise in computational models for culturally-sensitive belief acquisition and revision; (2) Prof. John Mikhail, Georgetown, who would provide expertise in the computational modeling of moral cognition & moral reasoning. If additional funds are needed to support these 2 new PI's during the Option Period, we will reallocate funds from the current subcontractors who had unexpended funds. Therefore, it is requested that no funds should be withheld from the Option Period.

SCIENTIFIC APPROACH: As mentioned above, a major goal of the MURI is to develop computational models that further our understanding of the relation between beliefs, decisions and actions in different cultures. Our aim is to provide predictive, rather than just descriptive models. For example, rather than simply describing the structure of a militant or non-violent network, we focus on processes that underlie their formation and evolution. Understanding the origin and nature of shared and revised beliefs among group members is critical. Unfortunately, data are scarce, and hence an important component of our effort is the reliable documentation of the formation of groups such as those engaged in the acts at Madrid, London, Bali, etc. The research led by Scott Atran under our MURI currently provides the most detailed and reliable set of data available, and this work continues with more details and studies of a dozen or so such networks. Additional data are also being collected in non-violent settings, such as the Wisconsin Menomenee, the Guatemalan Ladinos, and Amish and Muslim communities. These studies of non-violent communities allow closer examination of moral issues and sacred vs. secular values – the latter being critical in both belief revision and negotiations across cultures (see below). Because the origins of many beliefs rest in the traditions of a culture, we are formalizing how beliefs, actions and values are expressed in stories within a culture. The framework for representing stories will support rapid, automatic analysis of the semantic interpretation of text, such as excerpts from news sources. Implications for future trends can be explored through analogies with traditional stories from the relevant culture. Finally, individuals or groups striving for social change need to have strategies for actions. We are developing and testing new frameworks for understanding strategic reasoning, subject to beliefs, which dictate in part which moves are considered rational. These complement the more traditional "game-theoretic" approaches, and appear more plausible in many real world, multi-agent scenarios.

MAJOR ACCOMPLISHMENTS TO-DATE:

A. Important Transition events

- Atran, Sageman et al have provided detailed studies of the evolution of several militant networks, including Madrid, Bali, & Leeds. These studies document the psychology of the players, the relations among the players, their motivations as well as the network structure. Many important generalizations emerge. Implications of findings were presented by Scott Atran to U.S. Dept of State and UK House of Lords, Washington DC and London, Westminster, October / November 2007. (See http://www.edge.org/3rd_culture/Atran07/index.html.)
- Tenenbaum's group has developed a new framework for inferring structural forms from relational data. For example, the associated algorithms can discover that a particular set of social relations follows a hierarchy, order, cliques system of some other structural form. Predictions can then be made about how unobserved entities will interact. This framework has stimulated interest for application in several Gov't laboratories.

B. Other Scientific Accomplishments

- Atran, Medin & Ginges have completed an important study showing negotiations that propose trading sacred for secular assets may create impediments to any further productive negotiations. Hence culturally-sensitive negotiations can not be treated in the same manner as one would by an economist concerned only with monetary assets. This work provides one of the first clear findings in a Middle East context. Resolution of

quarrels arising from conflicting scared vales may require concessions that acknowledge the opposition's core concerns (Atran, Axelrod, Davis.)

- Atran & Axelrod have begun the development of an approach to framing (and reframing) of contentious values to promote negotiations in seemingly intractable conflicts and to understand competitive framing in pursuit of political goals.
- Studies of Hindu-Muslim conflict over sacred land in Northern India by Medin's group partially replicate their previous Middle East findings on how sacred and secular values do not intermix but also suggest that, in some contexts, sacred values are somewhat malleable.
- Studies by LeGuinn, Iliev, Medin and Atran reveal a cross-generation shift from the alignment of personal values with forest spirits to an alignment closer to values from God's perspective, with a corresponding shift away from ecological considerations to a focus on cash value. Similarity of rankings are correlated with social network distance for rankings from God's perspective and correlated with expert network distance for forest spirit rankings. These data provide a case study of the unraveling of a value system and blending of cultures (in essence the Itza' forest spirit notion of the Arux has been replaced by its Ladino counterpart, the duende).
- Argo (MIT) has surveyed motivations for Palestinian resistance in the Balata Refugee Camp, Nablus, given several different scenarios. One significant finding was that following noncombatant casualties, the data suggest about 840 individuals in Balata's population would be willing to participate in violent activities against the Israeli Defense Forces (IDF)—a significant counterinsurgency given the extremely dense housing conditions.
- In a non-violent context (Menomenee), Medin's group has documented a complex interaction among sacred values and moral decision-making: some sacred values are largely irrelevant to daily activities, whereas others are held in place by the social structures. These need to be distinguished in moral reasoning.
- Context plays a key role in moral judgments. Iliev & Medin have shown that when morally relevant issues are at stake, changing the context can have a strong influence on outcomes; These findings have relevance also for the field of choice behavior, suggesting that factors setting a context might have different sensitivity to information content. In Median's group, Joseph has found that even the simple variable of the ordering of events can affect moral judgments.
- Tenenbaum's group has developed a new Bayesian model called CrossCat: an unsupervised learning method capable of discovering multiple ways to classify a set of entities in order to show the different ways the entities can be related to one another. For instance, applied to voting data from the US Senate, CrossCat finds ten different ways to organize senators into clusters who vote in similar ways on subsets of issues. Moral, military, environmental issues exemplify some of these differences among the clusters.
- Page & Bednar have modeled the dynamics of belief revision that would be typified by an individual wishing to join a select group. The model explains two puzzles in the theoretical study of culture: how meaningful cultural signatures can be derived, and how diversity can persist within a culture.
- A preliminary model for small group evolution has been developed by Richards, with results that are consistent with some of the Atran data, as well as the formation (and breakup) of some non-violent groups. The dynamics of the model arises from a competitive interaction among three forces: leadership dominance, team bonding, and diversity. The group's vulnerability to alternative leadership emerges from the study.
- Beliefs underlie actions. Page and Golman have compared basins of attraction for different learning rules for collective actions. These models differ from most research that considers only the stability of equilibria and not their likelihood of attainment. With this newer approach to modeling, we can show that cultural learning and best response learning can have vanishing overlap in their basins of attraction.
- Bednar's group has completed a first set of experiments that show substantial contagion effects on behavior between games. The results question the predominant assumption of game theory that games can be studied independently.
- Tenenbaum's group also has been exploring the relation between beliefs and actions using a Bayesian Framework. Here the emphasis is on modeling an individual's ability to understand another's plans for actions.
- Dehghani and Forbus have developed a concept map system that enables scientists to construct formal representations of human mental models. For example, using interview data gathered by Medin's group, the system used analogical generalization to classify individuals in different cultures based on their belief systems. Inspection of the nature of the generalizations led to new insights about properties of the groups, subsequently confirmed by manual coding and analysis of transcript data.
- Dehghani and Forbus have created a first-cut computational model of moral decision-making. The different impact of sacred versus secular values is expressed via qualitative order of magnitude relationships. The model can derive connections between the choices of a scenario and the system's values either via first-principles reasoning or via analogy with stories involving moral reasoning.

- Tomai and Forbus have significantly extended the capabilities of Northwestern's natural language system, to handle the kinds of texts found in experimental moral scenarios, cultural fables, and life-stories from participants in experiments.
- Finlayson & Tomai have made significant progress in the development of the "Story Workbench". (The workbench is a joint project with Northwestern & MIT.) The system which supports the rapid collection of semantically interpreted stories. Beta testing is underway.
- Institutional constraints often differ in cultures that are physically adjacent. A simple example is driving on the left vs. right side of the road, or wearing a head-scarf. MIT & Harvard (Richards, Gal & Ficici) have begun simulations to uncover the dynamics of such conflicting constraints, given different initial population statistics and payoffs. (Note that payoffs can reflect differences in sacred and secular values, and are critical in finding equilibria.)
- Pfeffer, Gal and Ficici at Harvard have developed a framework using multi-agent influence diagrams that can model bilateral and multilateral negotiations; an experimental platform that incorporates this framework is Colored Trails, which is being used to explore negotiation strategies used by human subjects. Within this framework, Pfeffer's group has studied reciprocal behavior, reasoning under uncertainty, and the relationship between individuals' beliefs and their preferences.
- An important advance in understanding strategic play is Pfeffer's identification of four fundamental reasoning patterns in games to characterize the way information is used and manipulated. He introduces the concept of well-distinguishing strategies, which capture the kinds of strategies that are highly justifiable to a human. These strategies always include a Nash equilibrium, but are often distinct. Using this concept, the four reasoning patterns characterize all situations in which an agent cares about its decision. Pfeffer has begun studying how the reasoning patterns inform our understanding of negotiation games with conflicting interests.

Publications and Presentations: approximately 21 papers in refereed journals or refereed proceedings, about 16 working papers in progress, and 11 presentations at national and international meetings.

Student and Postodoctoral support over portions of the base period include: MIT: 4 Phd , 2 MS Graduate Research Assistants and 1 PostDoc; UMich: 4 PhD Graduate Research Assistants; Harvard: 2 PhD and 2 PostDoc; John Jay: 2 PhD and 1 PostDoc (includes Prof. Atran's UMich students); Northwestern: 6 PhD and 1 PostDoc; Univ Tx: 1 Phd student.

PLANS FOR THE OPTION YEARS: During the option years we will expand the MURI team to increase our expertise in computational modeling, especially of network formation and belief acquisition and revision. Studies will also continue to clarify the role of culture and context on patterns of reasoning and decision-making when moral issues come into play. In addition, our two years of profitable and unique interdisciplinary interactions has matured our overlook and has revealed several lacunae in our efforts. We wish to fill some of these gaps. For example, the early work in the 80's by Cavalli-Sforza & Feldman on Cultural Transmission and Evolution can be applied to belief acquisitions and revision. (Hence the suggestion to include Prof. Niyogi at Chicago, who is the current leader in computational approaches in this area.) We have also recognized that the exploration of large networks (> 100 or so nodes) is unlikely to give clear insights into the evolution of militant groups, and wish to move our focus to small group evolution. Moral vs. secular issues loom large across several of our projects, but only in the last year have we begun to unravel some of the complexities of moral vs. secular issues, as well as noting distinctions among moral values that are context sensitive. These become critical in negotiations between parties belonging to different cultures (as documented in our Middle-East study.) Like many others, we have begun our studies of negotiations with a game-theoretic viewpoint, with two player scenarios. Now, we see the importance of moving to multi-player scenarios, with repeated games, and with payoffs that better reflect aspects of moral vs. secular choices. Our work on patterns of reasoning provides a framework for such inclusions (whereas classical game-theoretic approaches do not.) Yet another area lacking computational models is the role of the horizon in strategic play. Preferences for alternatives at the beginning of a negotiation can change dramatically over time as new priorities emerge, or simply from the actions made during the negotiations. Our Colored Trails platform can be used to gain insight here.
In summary, work will continue and expand to include the following projects:

- Analysis and detailed description of the structure and formation of militant groups; modeling of their evolution.
- Modeling of culturally-sensitive belief revision and acquisition; continued explorations into how youths in a

culture (such as Amish or Menomenee) may reject values and traditions stressed during upbringing.

- Continue the Atran-Axelrod approach for the framing of contentious values to facilitate cross-cultural negotiations.
- Models of the role of (cultural) incentives and learning rules in predicting acts considered rational.
- A study of the way people use different reasoning patterns in a variety of situations.
- Initial development of a system for helping decision-makers understand the relevant factors in a complex game. The system will be based on Pfeffer's reasoning patterns paradigm.
- An investigation into the way people use models of others as the number of players in a game increases, using the Colored Trails paradigm.
- A model for recovering political stances form news reports (preliminary work has been done using IndaSea data.)
- A computational model of belief acquisition and revision using Niyogi's undated version of the Cavalli-Sforza & Feldman 's Cultural Transmission and Evolution approach.
- Dynamics of the resolution of conflicts between institutional constraints held by different interacting populations. Currently we have a simple example with dynamics that has not yet been completely analyzed.
- Page and Golman will continue to explore necessary and sufficient conditions for learning rules to lead to distinct equilibria. In the case of rational learning vs. cultural learning, one necessary condition appears to be that the initial best response must not be an equilibrium. This has a natural implication: in more complex environments in which participants may have to weed out initially successful but non equilibrium strategies, the particular form of learning may play a crucial role.
- Bednar & Page will restructure a computational model of institutional path dependence, in light of preliminary experimental findings. Additional experiments will be conducted to explore the role of institutional context in explaining apparently suboptimal choice, and the role that intermediate institutions play in overcoming inefficiencies in choice.
- Continued development of Bayesian framework for modeling the ability of one person to infer another person's plans for action.
- Initiative in a computational treatment of the consistency of moral values with Mikhail (i.e. a perspective from the theory of Law.)
- Extend Northwestern's computational model of moral decision-making, which currently has been tested only on hand-coded representations, to operate with a broad range of materials representing stories from different cultural groups.
- Test the combination of the QCM concept map system and analogical generalization with data from other cultural groups, to see if we can define protocols that enable systematic collection of data to automatically learn more accurate classifiers.
- Use the "Story Workbench" to replace our limited, hand-coded database with 100 semantically-interpreted, culturally sensitive stories that can be readily used for analogical reasoning.
- Complete a downloadable version of the "Story Workbench" (now in beta testing) that will be suitable for use in laboratories beyond MIT & Northwestern, using their feedback to guide future modifications.
- Demonstrate how analogy can identify causal trends in a culture's stories, thus allowing us to identify potential precedents for decision-making.
- Complete a research monograph summarizing how the moral domain influences decision-making, and the relevance of cultural factors.

Much of the above research will continue to require interdisciplinary collaboration among the MURI participants, as well as others who may be added. Modeling requires data, and, in turn, the models themselves will suggest new directions for inquiry and experiments. The past two years have established productive relationships among the MURI members, and we will further strengthen and expand these collaborations.

TRANSITION PLANNING: Scott Atran's studies are already making an impact in understanding "Terrorism and Radicalization: What not to do, What to do" (http://www.edge.org/3rd_culture/Atran07/index.html.) This flow will continue. Regarding the development and transfer of computational models, Tenenbaum's work has already gained visibility within the DoD community. Other modeling efforts should reach maturity by the end of the first option year (see below). To provide grounding during this development period, we have established preliminary contacts with other MURI initiatives as well as with Lincoln Laboratories and Sandia. Most promising is a planned collaboration with Lincoln Lab, provided their initiative in "Counter-Terror Social Network Analysis and Intent

Recognition" is funded. On the one hand, Lincoln Lab's efforts in inferring militant network structures from public information sources, such as news reports or search engines, will supplement data now provided by Atran's group, while, on the other, our models for group evolution will provide constraints on how Lincoln infers group structure from "noisy" data. As mentioned in an earlier section, we have begun beta testing of the Story Workbench. This should be available for use in other laboratories in a year. Finally, at the beginning of the second option year, we plan to organize three or more workshops specifically directed at technology transfer. A new generation of algorithms for recovering latent structures in data sets (such as Tenenbaum's Crosscat procedures) will be ready for applied use, as will other models using latent feature analysis for categorization. The workshop would be closed to the public, but would be open to various members of DoD community and affiliated laboratories, who would have the opportunity to engage in discussion of how the various algorithms might be most profitably applied. (A proceedings of the workshop would be available to the public.) A second workshop would be on important factors that motivate the formation of militant groups, and factors that favor their formation (i.e. "What not to do and what to do"). A third workshop would offer updated of tools and models for network analysis, especially with regard to the robustness of the information collected. Lastly, in our final year, we plan to hold a workshop specifically devoted to computational approaches to understanding social networks and belief dynamics in cultural contexts. The presentations at this workshop would be intended to provide exemplars that distinguish between descriptive and computational models, which we believe will have a long term impact, especially on establishing a deep understanding of the complex issues that relate to beliefs within different cultures.