Proceedings of the 2007 IEEE Symposium on Artificial Life (CI-ALife 2007)

The Structure of False Social Beliefs

M. Afzal Upal

Electrical Engineering & Computer Science Department,

University of Toledo, Toledo, OH, 43606

Email: afzal@eecs.utoledo.edu

***Abstract*—** This paper presents the architecture of a multiagent society (MAS) designed to study the dynamics of belief change in natural and artificial societies. It also presents a multiagent domain called Multiagent Wumpus World (MWW) designed to test the capabilities of the proposed MAS. It also reports on a set of experiments designed to study the formation of false social beliefs. Our results indicate that more false beliefs are likely to be generated about objects/events whose presence is harder to confirm or disconfirm.

I. Introduction

Modeling and understanding the formation, propagation, and evolution of beliefs is crucial both to the success of distributed artificial intelligence (AI) systems as well as to improve our understanding of human and animal societies. The growth of multiagent systems research in artificial intelligence [1] has been paralleled by a growing realization among cultural scientists that the traditional verbal models are too imprecise to model belief dynamics while mathematical models are too rigid and unable to be scaled up [2]. 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” [3]. The alternative that Moss and others propose is to build bottom-up algorithmic models of socio-cognitive processes. The key idea behind the agent-based social simulation (ABS) approach is to encapsulate each member of a population in a software module (called an *agent*) to build bottom-up models of human or animal societies. The ABS models focus on interactions between agents and, for the most part, abstract away the internal cognitive structure of the agents. Thomas Schelling, one of the early pioneers of the ABS approach, designed 1500 agents that lived on a 500 x 500 board [4]. The agent’s cognitive structure consisted of one simple inference rule, namely, if the proportion of your different colored neighbors is above a tolerance threshold then move, otherwise stay. He showed that even populations with high tolerance end up living in highly segregated neighborhoods.

The ABS methodology illustrates that it is not necessary, or even desirable, to have a complete understanding of a social system before building computational models. Indeed, ABS systems are frequently used as theory exploration and development tools (similar to the way computer models are used as tools by AI and Cognitive Modeling researchers) because they allow theoreticians to visualize and fully explore the consequences of their models and to compare competing theories. The last few years, there has been an explosion in the development of ABS systems designed to simulate social systems from a variety of domains. Ignoring the complex internal cognitive structure not only allows ABS designers to design computationally tractable simulation systems but it also helps them show causal connections between the cognitive rules that agents use to make local decisions and social patterns that emerge at the population level–the highly desired, yet rarely achieved–identification of micro-macro links.

Few ABS systems, however, have been built to specifically model beliefs dynamics and the systems developed to date assumed overly simplistic models of individual cognition and knowledge representation. For instance, most existing ABS models of social belief change model agent-beliefs as a single bit and belief change involves flipping the bit from 0 to 1 or vice versa often to match the beliefs of the neighbors [5][6][7]. This severely limits these systems as they are unable to model most real world distributed systems applications. Complex patterns of shared beliefs such as those that characterize people’s cultural and religious beliefs are also not likely to emerge out of such systems because the ABS agents are not even able to represent them. Thus existing ABS systems cannot be used to explore or model belief dynamics in human societies.

Traditionally, artificial intelligence and cognitive modeling have studied how individuals form and modify complex belief structures [8][9][10] but have, for the most part, ignored agent interactions assuming single agents living unperturbed in closed worlds. Artificial intelligence research on *classical planning* illustrates this approach well [11]. Given the knowledge about current state of world, about goals that the agent desires to achieve, and the generalized actions that the agent can take in the world, the planning problem is to compute an ordered sequence of action instances that the agent can execute to attain its goals. The classical AI planning research assumes that the planning agent is acting alone in the world so that the world does not change while the

1-4244-0701-X/07/$20.00 ©2007 IEEE

agent is figuring out what to do next because if that happens, the agent's plan may not be executable any longer. If the world continues to change the agent may never be able to act as it will always be computing the plan for the changed situation. Abstracting away other actors allows AI researchers to eliminate additional sources of complexity to focus on complex reasoning processes that go on inside the heads of individuals and result in the rich knowledge structures such as plans. This has led to the development of successful game playing programs that work in environments with limited or no interaction with other agents. However, this approach is not useful for modeling the dynamics of cultural belief systems such as religious belief systems because they are by their very nature products of the interaction of a large number of agents.

Clearly, to simulate belief dynamics in human societies, we need to develop knowledge-rich agent-based social simulation systems (KBS) [12]. Agents in these systems must have rich knowledge representation and reasoning capabilities and they must be able to interact with other agents present in their environment. Such simulation systems must overcome computational tractability concerns without abstracting away the agent's internal cognitive structure (as done by ABS systems) or ignoring interactions with other agents (as done by much of traditional AI & CM work)? Furthermore, to be able to tell us something about belief dynamics in human societies such agents in such systems must model the cognitive tendencies that people are known to possess. We believe that people's ability to communicate, comprehend a message, and integrate the newly received information into their existing knowledge is crucial to understanding the formation, propagation, and evolution of beliefs. We have designed a knowledge-rich multiagent society, called CCI[1], to model these processes. The challenge for any KBS system is that of overcoming the computational intractability problems to design an efficient system that can be run in real time. This paper argues that one promising approach for addressing this challenge is to develop synthetic computer games like environments that are rich enough to exercise the enhanced knowledge representation and reasoning capabilities of KBS agents yet they are not so complex to make the simulation intractable and the results impossible to analyze and understand.

II. Communicating, Comprehending, and Integrating (CCI) Agents

The CCI agents are goal directed and plan sequences of actions to achieve their goals. Some of the actions that they need to undertake to achieve their goals may be speaking actions. An agent A may decide to send a message M to an agent B if it believes that sending B the message M will result in changing B's mental state to cause it to perform an action C which can help A achieve one of its goals.

The CCI agents, similar to people [13], are comprehension driven i.e., they attempt to explain each piece of information their sensors detect. On observing an effect E, they search for a cause C that could have produced that effect.

Agents attempt to build accurate models of their environment by acquiring information about cause-effect relationships among various environmental stimuli. They store this information as cases [14]. Agents consult their case memory to form expectations about the future If these expectations are violated, they attempt to explain the reasons for these violations and if they can find those explanations, they revise their world model. The CCI agents ignore the information received from others if they cannot find any justification for it.

We have designed the first version of a CCI society by embedding it into an artificial domain. Multiagent Wumpus World (MWW), shown in Figure 1, is an extension of Russell and Norvig's [11] single agent Wumpus World.

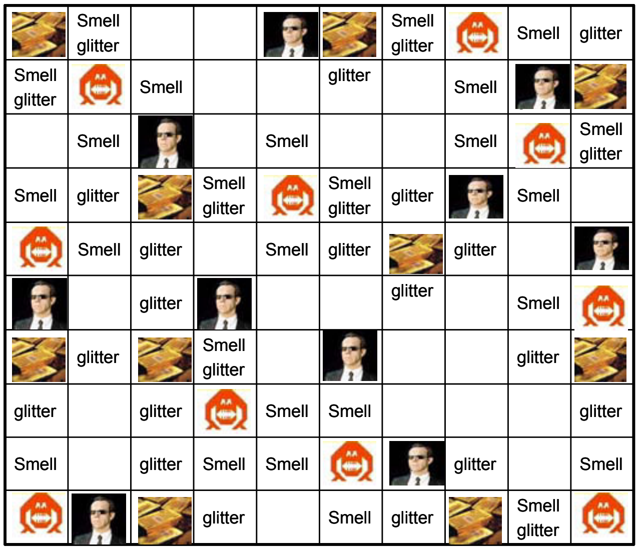

Figure 1: A 10 x 10 version of the Multiagent Wumpus World (MWW) domain. This version has 10 agents, 10 Wumpuses, and 10 Treasures.

A. Multiagent Wumpus World (MWW)

MWW has the same basic configuration as the single agent Wumpus World (WW). MWW is an *N* x *N* board game with a number of wumpuses and treasures that are randomly placed in various cells. Wumpuses emit smell and treasures glitter. Smell and glitter can be sensed in the horizontal and vertical neighbors of the cell containing a wumpus or a treasure. Similar to the single agent WW, once the world is created, its configuration remains unchanged i.e., the wumpuses and treasures remain where they are throughout the duration of the game. Unlike the single agent version, MWW is inhabited by a number of agents randomly placed in various cells at the start of the simulation. An agent dies if it visits a cell containing a wumpus. When that happens, a new agent is

[1] Communicate, Comprehend, and Integrate

created and placed at a randomly selection location on the board.

The MWW agents know the game rules. Agents know that smell is caused by the presence of a wumpus in a neighboring cell while glitter is caused by the presence of treasure in a neighboring cell. Agents sense their environment and explain each stimulus they observe. While causes (such as wumpuses and treasures) explain themselves, effects (such as smell and glitter) do not. The occurrence of effects can only be explained by the occurrence of causes that could have produced the observed effects e.g., glitter can be explained by the presence of a treasure in a neighboring cell while smell can be explained by the presence of a wumpus in a neighboring cell. An observed effect, however, could have been caused by many unobserved causes e.g., the smell in cell (2, 2) in Figure observed in could be explained by the presence of a wumpus in either of the four cells:

- (1, 2)
- (3, 2)
- (2, 1)
- (2, 3)

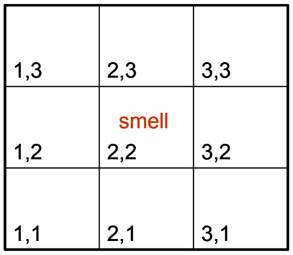

Figure 2: A part of the MWW.

An agent may have reasons to eliminate some of these explanations or to prefer some of them over the others. The MWW agents use their existing knowledge to select the best explanation. Agent's knowledge base contains both the game rules as well as their world model. A world model contains agent's observations and past explanations. The observations record information (smell, glitter, treasure, wumpus, or nothing) the agent observed in each cell visited in the past. The MWW agents use their past observations and game knowledge to eliminate some possible explanations e.g., if an agent sensing smell in cell (2,2) has visited the cell (1,3) in the past and did not find sense any glitter there, then it can eliminate "wumpus at (2, 3)" as a possible explanation because if there were a wumpus at (2, 3) there would be smell in cell (1, 3). Lack of smell at (1, 3) means that there cannot be a wumpus at (2, 3). Agents use their knowledge base to form expectations about the cells that they have not visited e.g., if the agent adopts the explanation that there is a wumpus in cell (2, 1) then it can form the expectation that there will be smell in cells (1, 1) and (3, 1).

In each simulation round, an agent has to decide whether to take an action or not. Possible actions include:

- the action to move to the vertically or horizontally adjacent neighboring cell
- the action to send a message to another agent present in the same cell as the agent, and
- the action to process a message that the agent has received from another agent.

The MWW agents are goal directed agents that aim to visit all treasure cells on the board. Agents create a plan to visit all treasure cells they know about. The plan must not include any cells that contain wumpuses in them. Each agent ranks all the cells by how confident it is of its knowledge about a cell. It has the highest confidence in the cells that it has already visited. Next are the cells whose neighbors the agent has visited and so on. Agents also rank cells by how urgently they need the information about that cell. The order in which the cells are to be visited determines the urgency ranking e.g., if a cell is the next to be visited then finding information about that cell is assigned the highest priority while a cell that is not planned to be visited for another 10 rounds gets low priority. The agents the use an information seeking function that takes the two rankings (confidence and urgency) as inputs and decides whether the agent needs to communicate and if so which cell it needs the information about. If it decides to communicate and if another agent is currently present in the same cell then agent sends a request-for-information message to that agent.

The listening agent attempts to explain as to why the speaker sent it the message (e.g., was it sent to seek information or to distract me from traveling) and depending on that determination it may or may not decide to respond. If the agent decides to respond, it will offer the information about the requested cell in exchange for information about a cell about which it needs information. If the speaker agrees to the exchange then the agents communicate information about the cells.

On receiving information about a cell, an agent has to decide whether to incorporate that information into its knowledge-base or not. If it decides to incorporate the new information then it has to decide how best to revise its existing knowledge. If the new information confirms what the agent already know about the cell then no revision is done. If, on the other hand, the information received from another agent is different from what the agent expects to find in that cell then it attempts to explain the reason for the contradiction. If it can find a possible explanation that it ignored in the past that supports the received information, then it adopts the new explanation and the new information and retracts its belief in the old explanation and expectation. Otherwise, the agents rejects the newly received information.

1) Formation of False Beliefs

Agents use symmetrical processes to form and revise beliefs in the presence of treasures and wumpuses. Thus agent models may contain false beliefs about the locations of the wumpuses as well the locations of treasures. However, while the agents prefer to travel towards a cell that they believe contains a treasure, they avoid cells that they believe contain wumpuses in them. This makes beliefs in the presence of wumpuses relatively harder to confirm or disconfirm than beliefs in the presence of treasures. The experiments described next were designed to investigate the impact that this asymmetry has on the patterns of false beliefs that the agents form.

III. Experiments and Results

I designed a 10 x 10 version of MWW with ten agents randomly placed at ten different locations. In the first set of experiments I designed three different versions of MWW:

- The 5x5 version has 5 wumpuses and 5 treasures
- The 10x10 version has 10 wumpuses and 10 treasures, and
- the 20x20 version has 20 wumpuses and 20 treasures.

I allowed the simulation to run for 100 rounds. At the end of that round I measured the following metrics:

- *Average agent age* is the average age of the ten agents that survive after round 100.
- *The average number of wumpus beliefs* is the average number of wumpus beliefs the surviving agents have
- *The average number of treasure beliefs* is the average number of treasure beliefs the surviving agents have
- *Average number of cells visited* is the average number of cells that 10 agent surviving at the end of round 10 have visited.

Figure 1shows the results of Experiment I. The results for average age show that 10x10 world proves to be the most taxing for agents with average agent ages only being less than 30 rounds. Agents in 5x5 and 20x20 rounds, on the other hand, have much higher average ages. This is explained by the considering the average number of cells that the agents visit. Agents in the 5x5 world visit a larger number of cells while agents live longer in the 20x20 world by visiting fewer cells and by thus deciding to stay in their cells. Agents in the 10x10 world, however, visit more cells than in 20x20 world but they pay the price of this adventurism by having smaller average ages.

In the second experiment, then I adopted the 10x10 world and measured the proportion of false wumpus and treasure beliefs that the agents had at the end of round 100. The proportion of false wumpus/treasure beliefs is the proportion between the number of false wumpus/treasure beliefs that the agent has the total number of wumpus/treasure beliefs that the agent possesses.

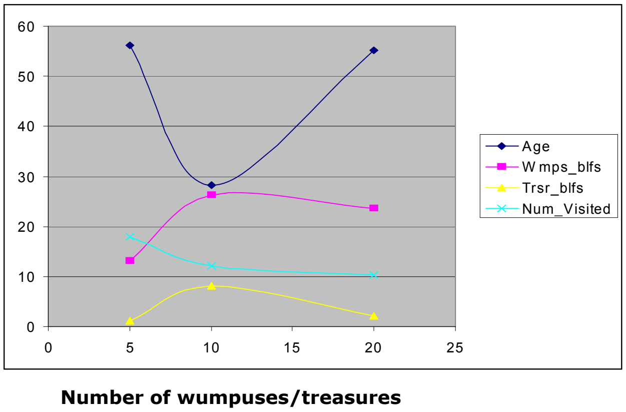

Figure 3: Results of Experiment 1. Each point is an average of 30 runs.

The results (Figure 4) show that after the initial drop, the proportion of false wumpus beliefs remains relatively unchanged regardless of how much the agents travel in the world. False beliefs in the presence of treasures, however, continue to decrease as agents with agent age. The agents who survive the length of simulation (100 rounds) have few, if any, false beliefs about treasures but on average 40% of their beliefs about the presence of wumpuses are false.

A. Discussion

Our results indicate that explanations that are harder to confirm and disconfirm are more likely to be generated by agents that attempt to explain their observations and revise these explanations in light of the evidence. This suggests that people should have more false beliefs about things that are harder to confirm or disconfirm. There is some evidence to suggest that that is the case. Bainbridge and Stark [15] made confirmability the core of their theory of religion to argue that religious beliefs are unconfirmable algorithms to achieve rewards that are highly desired by people yet cannot be obtained. Similarly, there is some evidence to suggest that many false ethnic stereotypes people have are about things that are harder to confirm or disconfirm such as the sexual practices of the neighboring tribes.

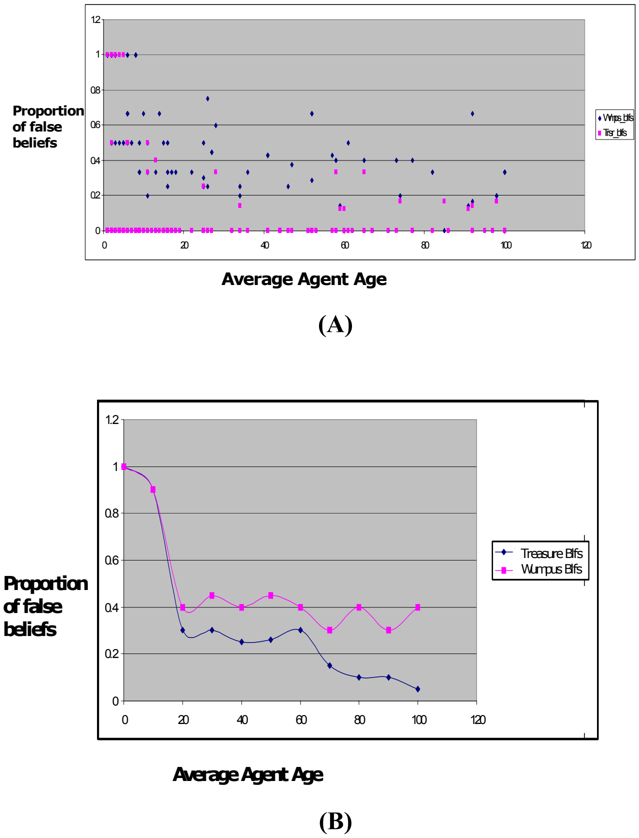

Figure 4 (A) Scatter plot of proportion of false wumpus and treasure beliefs. (B) discretized version of plot (A).

IV. Conclusion

This paper presents the architecture of a multiagent society designed to study the dynamics of belief change in natural and artificial societies. It also presents a multiagent domain designed to test the capabilities of the proposed systems. Preliminary results are encouraging as they indicate the potential for the use of the proposed society.

References

[1] G. Weiss, *Multiagent systems: A modern approach to distributed artificial intelligence*, Cambridge, MA: The MIT Press, 1999.

[2] C. Castelfranchi & B. Kokinov (editors), *Understanding the dynamics of knowledge: Integrating models of knowledge change, development and evolution in cognitive science, epistemology, philosophy, artificial intelligence, logic, and developmental, and evolutionary psychology*, Technical Report of the European Science Foundation, 2005

[3] S. Moss, Game Theory: Limitations and Alternatives, *Journal of Artificial Societies and Social Simulation*, 4(2), 2001.

[4] T. Schelling, Dynamic models of segregation, *Journal of Mathematical Sociology*, 1, 143-186, 1977.

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

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

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

[8] C. E. Alchourròn, P. Gärdenfors, and D. Makinson (1985). On the logic of theory change: Partial meet contraction and revision functions. Journal of Symbolic Logic, 50:510-530.

[9] Allen, J. (1987). Natural Language Understanding. Menlo Park, CA, Benjamin Cummings.

[10] Bringsjord, S. and Ferrucci, D. (2000). Artificial Intelligence and Literary Creativity: Inside the Mind of BRUTUS, a Storytelling Machine. Mahwah, NJ, Erlbaum.

[11] Russell, S. & Norvig P. (1995) Artificial Intelligence: A Modern Approach, Englewood Cliffs, NJ: Prentice Hall.

[12] M. A. Upal & R. Sun (editors) Cognitive Modeling and Agent-based Social Simulation: Papers from the AAAI-06 Workshop (ISBN 978-1-57735-284-6), Menlo Park, CA: AAAI Press, 2006.

[13] W. Kintsch, *Comprehension: A Paradigm for Cognition*, Cambridge, NY: Cambridge University Press, 1998.

[14] R. C. Schank & R. P. Abelson, *Scripts, plans, goals, and understanding: an inquiry into human knowledge structures*, Hillsdale, NJ: Lawrence Erlbaum, 1977.

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

[16] L. Smith, Sects and Death in the Middle East, *The Weekly standard*, 11 (39).