Agents, Trust, and Organizational Behavior

From: AAAI Technical Report FS-97-02. Compilation copyright © 1997, AAAI (www.aaai.org). All rights reserved.

Michael Prietula
Department of Commerce & Technology
Johns Hopkins University
201 N. Charles Street, Suite 200
Baltimore MD 21201
prietula@jhu.edu

Kathleen Carley
Department of Social and Decision Sciences
Carnegie Mellon University
Pittsburgh PA 15213
Kathleen.Carley@centro.soar.cs.cmu.edu

Abstract

Trust is viewed as a computational construct capable of being explicitly studied and modeled. Although articulated as a (functional) property of individual agents, the collective effect of events influencing (and being influenced by) trust judgments can impact organizational behavior. In this paper we summarize our work on agent trust and explore a current study on trust, cooperativity, and benevolence of agents and organizational behavior.

Introduction

Organizations are complex systems in which a large number of factors interact in complex and often non-linear fashions. One important source of these non-linearities is the adaptiveness, not just of the organization as a unit; but also of the set of agents within the organization. The organizations overall behavior is thus affected by two logics, which are often at odds, the logic of the task and the logic of interaction. In the logic of task, the set of agents in the organization work collectively to solve some problem or achieve some goal. In the logic of interaction, the set of agents communicate and exchange information in order to create and maintain social norms that may or may not be task related. Theorizing about organizations is complicated by these aspects of organizational reality.

We can understand organizations, and develop our theories of them, by crafting and analyzing computational models of organizations as collections of intelligent adaptive agents. Computational theorizing is an invaluable asset to the organizational theorist as it helps to lay bare the implications of the agent adaptation, task, and organizational culture on organizational performance. Computational theorizing helps us to reason through, systematically, the consequences of the multiple interacting factors present within organizations. Further, both social interaction and the capabilities of a social agent can be defined computationally (Carley and Newell 1994, Carley and Prietula 1994). As such, it is meaningful to conduct computational experiments which address social interaction as well as (associated) cognitive properties in order to explicate properties of organizations of such agents (Carley and Prietula in press, Carley, Kjaer-Hansen, Newell and Prietula 1992, Prietula and Carley 1994).

One issue that computational theorizing helps us to reason about is the impact of aspects of human and human-like interaction, such as trust, on organizational performance under different task and structural constraints. For example, we might ask, if the organization is facing a dynamic environment, where the lessons learned today are not necessarily applicable tomorrow, are there organizational consequence if agents lie? In this paper we summarize some of our research in this area and describe a computational model of organizational performance under conditions of agent uncertainty. One source of agent uncertainty is inaccurate information. Since agents interact both with each other and the environment, this uncertainty may derive from agents not telling the truth or from agent facing a changing environment.

Within organization science, researchers have examined the impact of uncertainty on organizational performance. Some of this work focuses on the relationship between individual uncertainty and overall organizational performance. Importantly for our purposes, some of this work led to the development of organizational models that begin to address issues of uncertainty in a formal fashion. For example, Carley, Prietula and Lin (in press) examine how the performance of the organization is affected by the agents being given incorrect information. Not surprisingly, they find that performance decreases under conditions of uncertainty. More to the point, they also demonstrate that there is an interaction between uncertainty, the organizational structure (team versus hierarchy), and the division of labor. Thus, some organizations, just because they have different structures may be less affected by uncertainty (Carley 1991, Carley and Lin 1995). Differences at the individual level can have profound effects on organizational performance; however, the strength of those effects depends on the structure of the organization.

A Turing Effect

One issue is the effect that crafting information and advice through a particular medium -- the computer. However, it is not simply the medium, it is the context in which the message is presented. One point of impact of a system is the effect it can have in delivering advice to a human, who then subsequently makes decisions based, in part, on that advice. It is a form of exploring trust. One model of exploration in this arena is a "good-bad" model in which the advice is either right or wrong. A more elaborate model is the crafting of a richer context in which the "right or wrong" advice appears.

In Lerch, Prietula and Kulik (1997), a series of human-computer interaction experiments were conducted, comparing the how trust levels varied with how the source of advice was characterized over the computer. That is, all advice was received as a form of email, but the source of the advice was differentially described (human experts, expert system, and peers). Results revealed a "Turing effect" in which the characterization of advice as coming from an expert system had significant effects on trust (e.g., they trusted the expert systems less than the human experts). In addition, the results provided insight into how these subjects "viewed" as an internally defined model (e.g., expert systems could not exert "effort," but humans could). Expert systems are seen, as characterized in these studies, as significantly different types of creatures. Furthermore, evidence reveals possible sources for the Turing effect and demonstrates that manipulation of how advice is characterized can influence trust judgments (Lerch, Prietula, Kim and Buzas 1997).

Computational Study 1

Demonstrating that differential trust judgments exist and can be manipulated on an individual level, we move to computational agent research, we use a multi-agent model of teamwork in which a collection of Soar (Laird, Newell and Rosenbloom 1987) agents are working collectively to accomplish a generic search task, which is viewed isomorphically as search on the internet (e.g., Carley and Prietula in press), or searching for items in a warehouse (Prietula and Carley 1994). Viewed from the latter perspective, each agent has task knowledge about how to get orders and how to tell when an order is filled. Each agent moves about in the warehouse locating items to fill the order. Agents can and do interact. One form of interaction is that agents can ask each other where the items they need are located. They are seeking simple advice. In response to this query, the other agents can either tell the truth (report where they last saw the needed item) or lie (purposely report that the needed item is somewhere other than where they last saw it). The objects in the warehouse, however, are not stationary. Specifically, as agents fill orders they may move the items they do not need to other locations so that they can access the items they do need. The fact that agents can move items makes the environment volatile.

The size of the organization and the trustworthiness of the agents interact. Thus, as the size of the organization increases organizations of trustworthy agents ask and answer more questions; whereas, organizations of non-trustworthy agents ask and answer fewer questions. This is simply the effect of the value of information. More accurate information increases the willingness to ask questions. More inaccurate information decreases the willingness to ask questions. However, the effects are not linear. There is also an impact of environmental volatility. Specifically, as the size of the organization increases the effects just mentioned continue to occur but at decreasing rates. As organizations increase in size the environment becomes more volatile. Thus, because items are moving, honest agents appear dishonest and dishonest agents appear honest. We would speculate that eventually there may be a point where volatility is sufficiently high that both curves "turn" and no difference is discernible due to trust.

Computational Study 2

We are beginning a simulation study in which trust models from empirical human studies are woven into simple computational models of groups, in the fashion described above. Specifically, we reimplemented the Plural-Soar model into a less-cognitive, agent-based form. As the prior research demonstrated the general nature of trust judgment deliberation, the particular deliberation model could be extracted from the data and crafted directly in algorithmic form.

The task is the same as previously described: a set of agents, that repeatedly go to a particular location to acquire an item to find, then proceed to search at different locations for that item. There are three components to their social behavior: honesty (will they lie or not), benevolence (how forgiving are they), and cooperation (under what circumstances to they give advice). In this paper we hold benevolence and cooperation constant -- they are quick forgivers and they cooperate when they have the first opportunity. We vary the first social component: whether they lie or not about the location of an item.

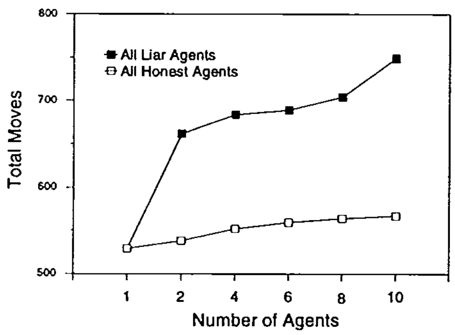
In addition, each agent has an item-memory (it recalls where it has encountered all items), a social memory (it recalls the value of the advice from other agents), and a communication capability (it can ask and receive advice). All three work to define the social interaction among agents. In this study, the nature of the interaction is as follows. Agents are either Liars or they are Honest. Essentially, this defines how an agent responds to a general request for advice. An Honest agent will respond directly to the questioning agent if it knows the location of the item in question. A Liar agent will respond to any request for advice, supplying incorrect location information. An agent, however, recalls the advice provided by an agent, and engages a simple social judgment model of advice acceptance, based on three judgment states: trustworthy, risky, and untrustworthy. Good/bad advice moves the judgments up/down. Advice from risky/untrustworthy agents is not accepted and questions from them are not answered (unless the answering agent is a Liar). If an agent is unforgiving, the untrustworthy state is absorbing – an agent deemed untrustworthy remains so judged. As noted, all agents in this study are benevolent. It is a forgiving group.

The general model consists of 10 item locations (where items are stored), an order queue (where requests are taken), and a delivery queue (where items are taken). A total of twenty unique items were in the order queue (non-redundant) and each agent takes one order at a time. The purpose of this preliminary study is simple. We wish to explore the extent that misinformation can impact organizational performance and behavior. Our model should generate the fundamental collective behaviors it purports to explain.

Results

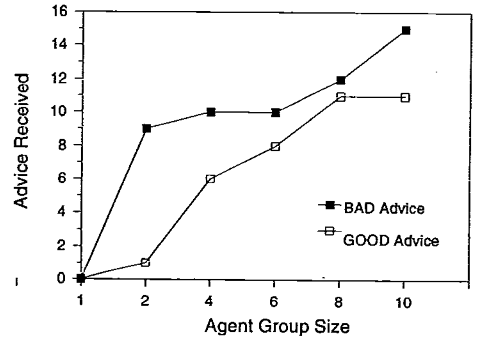
The first questions address the general impact lying would have on group behaviors, thus serving as a baseline for subsequent studies. Table 1 summarizes the Total Moves taken to complete the 20 item task by two types of groups: all Liar agents and all Honest agents. This is varied by group size for each (1, 2, 4, 6, 8, 10 agents).

Table 1



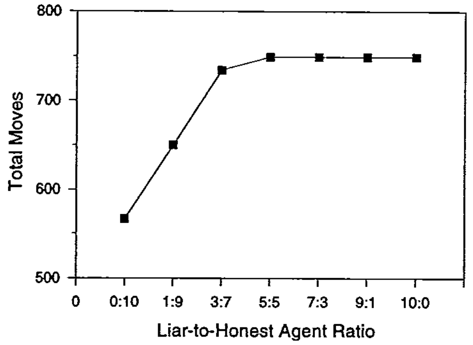
As can be seen in Table 1, a group of Liar agents quickly incur penalties for their incorrect advice. The primary reason for this resides in the nature of the advice being generated by each group. In Table 2, we note the increasing BAD advice (occurring in the Liar group) and the increasing GOOD advice (occurring in the Honest group). The bad get worse and good get attenuated.

Table 2



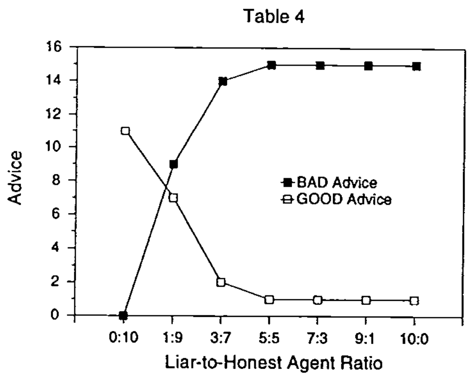
Though it is obvious that such relative relationships would exist (i.e., Liar agent groups do worse than Honest agent groups in stable environments), what is perhaps less obvious is the impact Liar agents would have in a Benevolent group. This was explored as follows. The group size was held constant at 10, while the number of Liar agents in the group was varied according to the schedule: 0, 1, 3, 5, 7, 9, 10. Thus, the extremes considered are the prior reported conditions: all Liar agents, all Honest agents. It is the mix that is of interest. Table 3 summarized the Total Moves for the seven groups simulated.

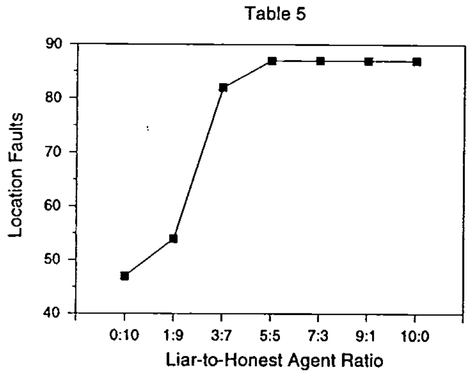
Table 3



The underlying events can be traced to the good and bad advice generated under the different conditions. Table 4 summarizes these results. Note that in Table 4, the amount of Bad advice generated quickly dominates the organizational communication at low ratio levels – a small number of Liar agents can be extremely disruptive to group performance. How this occurs is two fold. First, in order to provide Good advice, an agent must have good information. That is, the agent must know where a given item is located. On the other hand, giving Bad advice requires no information, unless an agent wishes to ensure that an item is not in a known location. These agents, however, simply generate the same incorrect response. They do not search their memories for the correct one, nor do they require it. It is the preference for these agents to trust communication.

Without a given advised location, they will engage in a systematic search. However, information provided will cause them to do directly to the suggested location (without investigating interim locations). This has the effect, for Liar agents, of sending agents to false locations, causing Location faults – going to a location without success. Table 5 summarizes this result.





Conclusion

We have presented an exploratory analysis of the relationship between an affective response, trust, and organizational performance. Further research in this area needs to explore the impact of other affective responses. Individual agents develop cognitive coping mechanisms for dealing with uncertainty, and their affective response to this uncertainty. These individualized cognitive coping mechanisms may be, and are often assumed to be, detrimental to the organization as these cognitive mechanisms may lead to individuals acting in more rigid, less flexible, less efficient fashions. Note, however, the impact that wrong information, presented at a high rate, can influence individual and collective action. The combination of misplaced trust and benevolent nature can yield suboptimal decisions. The model is necessarily incomplete and simple. Yet, as we study the impact of wrong information (as a form of Lying) on the Internet, and the research which points differential trust judgments and control (Turing effect), then the impact of disseminated misinformation is understood. Whether it is a "graduation talk" at MIT or a "finding" of findings in a sensational murder case, or a "report" of an organizations investment risk. It might matter. It might matter a lot.

References

Carley, K. and Newell, A. 1994. The Nature of the Social Agent, *Journal of Mathematical Sociology*, 19(4), 221-262.

Carley, K., J. Kjaer-Hansen, J., Newell, A. and Prietula, M. 1992. Plural-Soar: A Prolegomenon to Artificial Agents and Organizational Behavior. In M. Masuch and M. Warglien (Eds.), *Artificial Intelligence in Organization and Management Theory*, Amsterdam: North-Holland.

Carley, K. and Prietula, M. 1994. ACTS Theory: Extending the Model of Bounded Rationality. In K. Carley and M. Prietula (Eds.), *Computational Organization Theory*, Hillsdale, NJ: Lawrence Erlbaum.

Carley, K. and Prietula, M. In press. WebBots, Trust and Organizational Science. In M. Prietula, K. Carley and L. Gasser (Eds.), *Simulating Societies: Computational Models of Institutions and Groups.* Cambridge, MA: AAAI/MIT Press.

Carley, K. and Lin, Z. 1995. Organizational Designs Suited to High Performance Under Stress. *IEEE - Systems Man and Cybernetics*, 25(1), 221-230.

Carley, K. 1991. Designing Organizational Structures to Cope with Communication Breakdowns: A Simulation Model. *Industrial Crisis Quarterly*, 5, 19-57.

Carley, K., Prietula, M. and Lin, J. In press. Design versus Cognition: The Interaction of Agent Cognition and Organizational Design on Organizational Performance. In R. Conte and E. Chattoe (Eds.), *Evolving Societies: The Computer Simulation of Social Systems.*

Laird, J., Newell, A. and Rosenbloom, P. 1987. Soar: An Architecture for General Intelligence. *Artificial Intelligence*, 33, 1-64.

Lerch, J., Prietula, M. and Kulik, C. 1997. The Turing Effect: The Nature of Trust in Machine Advice. In P. Feltovich, K. Ford and R. Hoffman (Eds.), *Expertise in Context: Human and Machine*, Cambridge, MA: AAAI/MIT Press.

Lerch, J., Prietula, M., Kim, J., and Buzas, T. 1997. *Unraveling the Turing Effect: Measuring Trust in Machine Advice.* Working Paper, Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh PA.

Prietula, M. and Carley, K. 1994. Computational Organization Theory: Autonomous Agents and Emergent Behavior. *Journal of Organizational Computing*, 4(1), 41-83.