

□ EXPLORING THE EFFECTS OF
AGENT TRUST AND
BENEVOLENCE IN A
SIMULATED
ORGANIZATIONAL
TASK

MICHAEL PRIETULA

Department of Commerce and Technology, Johns
Hopkins University, Columbia, Maryland, USA

KATHLEEN M. CARLEY

Department of Social and Decision Sciences, Carnegie
Mellon University, Pittsburgh, Pennsylvania, USA

Executives argue intuitively that trust is critical to effective organizational performance. Although articulated as a cognitive/affective property of individuals, the collective effect of events influencing (and being influenced by) trust judgments must certainly impact organizational behavior. To begin to explore this, we conducted a simulation study of trust and organizational performance. Specifically, we defined a set of computational agents, each with a trust function capable of evaluating the quality of advice from the other agents, and rendering judgments on the trustworthiness of the communicating agent. As agent judgments impact subsequent choices to accept or to generate communications, organizational performance is influenced. We manipulated two agent properties (trustworthiness, benevolence), two organizational variables (group size, group homogeneity/liar-to-honest ratio), and one environmental variable (stable, unstable). Results indicate that in homogeneous groups, honest groups did better than groups of liars, but under environmental instability, benevolent groups did worse. Under all conditions for heterogeneous groups, it only took one to three liars to degrade organizational performance.

To what extent does trust influence organizational performance? Executives, as well as the literature on management, point to the importance of trust; in fact, it is often argued that trust is a fundamental prerequisite for good

An earlier version of this paper was presented at the AAAI Fall Symposium at MIT as "Agents, trust, and organizational behavior," by M. Prietula and K. Carley, pp. 146-149, in K. Dautenhahn (ed.), *Socially Intelligent Agents*, AAAI Press Technical Report FS-97-02, American Association for Artificial Intelligence, Menlo Park, California, 1997.

We thank Kerstin Dautenhahn, the symposium participants, and the anonymous reviewers for their input. This research reported herein was supported in part by the National Science Foundation under grant IRI9633 662.

Address correspondence to Kathleen Carley, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA. E-mail: Kathleen.Carley@centro.soar.cs.cmu.edu

organizational performance (Nicholas, 1993). Without trust, work needs to be rechecked, decisions need to be reevaluated, cooperation decreases, collaboration diminishes, and organizational performance declines. In a recent survey on job performance, respondents rated "trust" and "getting along with others at work" as the two most important skills (Frazee, 1996). Similarly, Krackhardt and Hanson (1993) illustrate the significance of both advice and trust networks in an organization. But is trust always necessary for good organizational performance? And under what conditions is untrustworthiness even noticeable in an organizational environment?

We begin to explore such questions as these from a slightly different perspective and a slightly different methodology than is typically engaged for "affective" research. Specifically, we make a strategic assumption that under certain circumstances, situationally defined trust is best viewed as a primarily cognitive, rather than primarily affective, construct. Furthermore, under this assumption, trust can be articulated in computational form as a social judgment decision rule, albeit a simple one. That is, for some tasks, trust judgments may be required, but these judgments require more information than affect. From this, we explore how small variations on trust judgments, realized as simple deliberation properties of interacting computer-based agents, impact individual and organizational behavior.

Organizations are complex, dynamic systems often faced with ambiguous environments that may change rapidly and unpredictably. For organizations, factors interact in complex and often nonlinear fashions to determine performance (Carley & Svoboda, 1996). One important source of these nonlinearities is the adaptiveness both of the organizational agents and of the organization. Change in what individuals know or what they believe they know influences individual and collective performance. However, change and uncertainty go hand in hand; that is, what was true at one instant of time may not be true at another.

A critical resource for organizations in the face of change and uncertainty is information (Stinchcombe, 1990). An important source of information for an individual in an organization is that obtained by communicating with other individuals in the organization (Mintzberg, 1973). However, information from others (as indirect experience or knowledge) can be a source of uncertainty as well as information for a variety of reasons, ranging from observational difficulties of the original source, changes in the original source since observations, or even intentional communication of inaccuracies—lying. Consequently, communication from individuals during task performance (about the task) can be a source of uncertainty, thus contributing to task ambiguity. Lies can be thought of as a form of deliberate manipulation of the uncertainty of disseminated information, therefore (possibly) increasing the ambiguity of the task, as the capacity of information to influence choice decreases.

In this paper we examine the extent to which lies impact organizational performance. As we have noted, our method of exploring this topic is based on a computational model of agents with well-defined, but simplified models of social judgment and communication. We first briefly summarize the relevant literature on trust and organizational performance. We next describe a computational view of trust. We then describe the current study, in which we explore how organizational (team) size, agent benevolence, and the number of lying agents (as a source of task ambiguity) impact organizational behavior (processes) and performance in both stable and unstable task environments (another source of task ambiguity).

TRUST AND ORGANIZATIONAL BEHAVIOR

Research on organizations has generally been concerned with one of two logics of performance—the logic of the task and the logic of interaction—which can be at odds. In the logic of the task, the set of agents in the organization work collectively to solve some problem or achieve some commonly held goal. This logic assumes a set of cooperative and trustworthy agents. In such a system, ambiguity, should it occur, originates from information errors external to the individual, and ambiguity is an exogenous factor to the individual's deliberation processes. In the logic of interaction, the set of individuals communicates and exchanges information in order to create and maintain *social* norms and to achieve individual goals. This logic assumes neither cooperation nor trustworthiness, but it does presume that the task is not an overly influential basis for interaction; ambiguity may arise either from the individual or from transmission errors between individuals. We explore the range of these two logics by merging types of organizational individuals in an organizational setting. For some individuals, ambiguity is exogenous and indeed external to that individual. For others, ambiguity is endogenous and deliberately created.

Studies of organizational behavior usually assume that the logic of the task dominates the group, and task-based collective behavior becomes interpreted and reinforced by social norms of cooperation and expectations centered on trust. Empirical evidence does suggest, at the organizational level, that trust and performance are related. For example, at the inter-organizational level, Zaheer et al. (forthcoming) demonstrate that exchange performance (i.e., setting up agreements to transfer goods-services-personnel between two companies for other goods-services-personnel) was higher for those organizations where boundary spanners (i.e., those who work across organizational boundaries) reported higher levels of trust in the partner organizations. At the individual level, however, the relation between trust,

cooperatives, and performance is less clear. Cooperativeness is not always in the individual's best interest, for lying can produce an advantage, and levels of trust can fluctuate significantly (Klein, 1997).

Herein our concern is with trust in the organizational context. Our understanding of trust draws from the social and behavioral sciences, as issues of affect, in general, and trust, in particular, have long been of interest to social scientists. Research in psychology and sociology has examined the impact of trust in individuals (Barber, 1983; Rotter, 1971), in close relationships (Rempel et al., 1985; Rempel & Holmes, 1986), and in groups (Lewis & Weigert, 1985a, 1985b; Luhman, 1979; Shapiro, 1987; Zucker, 1986). Most of these studies have not been exceptionally interested in organizations as such; nevertheless, organizational researchers have recently begun to specifically examine issues of trust (Kramer & Tyler, 1996). Much of organizational trust research is in the area of strategy and focuses on interorganizational relations (Bradach & Eccles, 1989; Ring & Van de Ven, 1992; Zaheer & Venkatraman, 1995; Zaheer et al., in press).

In the social and behavioral sciences, research is generally not on group trust per se, but on related concepts such as the emergence of conventions (such as truth telling) and the evolution of cooperation. Within this research, computational models are occasionally used to explore the impact of collective adaptation on the formation of group level behavior and cooperation (Axelrod, 1987; Macy, 1991b; Skvoretz & Fararo, 1995; Turner, 1993). Many of these models demonstrate that environmental and institutional factors, such as payoffs, population dynamics, and population structure, influence the evolution of cooperation in a discontinuous fashion (Axelrod & Dion, 1988). They show how simple interactions can lead to the development of and change in beliefs in the underlying culture (Ballim & Wilka, 1991; Carley, 1991a; Kaufer & Carley, 1993).

Such studies also demonstrate that when collections of individuals interact, and each individual has a simple goal (such as trying to attain the highest cumulative award), interesting and nontrivial social dynamics and norms can arise (Shoham & Tennenholtz, 1994). Macy (1991a) demonstrates the ability of social learning (i.e., learning from and about others in the group) to result in cooperation in social groups. Kephart et al. (1992) argue that social chaos (i.e., indeterminate and uninterpretable patterns of action and interactions in a group) is reduced when organizational constituents determine their actions using strategies based on observations and beliefs about others.

But what would happen if organizational constituents choose to lie? What would be the nature of the social dynamics? Following in the tradition of computer simulations of groups, we develop a computational model of organizational behavior in which the organization is modeled as a collection of adaptive social agents. We examine whether or not truthfulness, as a

prerequisite for continued cooperation, is necessary for improved performance of the organization.

From an individual agent's perspective, information about the organizational task is gained through direct observation and through a process of interacting with other agents. We take a constructural perspective on this interaction process (Carley, 1991a; Kaufer & Carley, 1993): as agents interact, they acquire information that changes the way they perceive the world, their actions, attitudes, and beliefs. These changes influence the nature and the extent of subsequent agent interactions. The dynamics underlying this change is an evaluation of the relative similarity between agents. That is, two agents are more likely to interact if they both believe that they are more similar to each other than they are to others in the group (Carley, 1991a; Carley & Krackhardt, 1996). We extend this perspective by basing these similarity judgments on the perceived truthfulness of the other. In our model, agents always see themselves as truthful, whether or not they are. Thus, as they encounter others, if they come to believe that those others are less like themselves (i.e., that they are liars), then they are less likely to engage that other in a future interaction. The likelihood of subsequent engagement is a social evaluation function incorporated in the model.

Within organization science, researchers have examined the impact of various forms of uncertainty on organizational performance (March & Weissinger-Baylon, 1986; Pfeffer et al., 1976). A subset of this work focuses on the relationship between agent uncertainty and overall organizational performance, leading to the development of organizational models that begin to address issues of uncertainty in a formal fashion. For example, Carley et al. (in press) examine how the performance of the organization is affected by the agents being incorrect information from other agents about the task state. Not surprisingly, organizational performance decreases under conditions of information uncertainty. More to the point, there is an interaction between uncertainty, the organizational structure (team versus hierarchy), and the division of labor. In particular, some organizations, solely because they have different structures, may be insulated against such forms of uncertainty (Carley, 1991b; Carley & 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 (Lin & Carley, 1993).

Researchers have begun to explore organizational issues using computational models of organizational decision makers—agents. In general, the role of affective, social, and personality-based agent attributes is of increasing interest to researchers in the areas of computational social and organizational theory (e.g., Canemero & Van de Velde, 1997), artificial intelligence (Bond & Gasser, 1988; Huhns & Singh, 1998; Dautenhahn, 1997), and social scientists in general (Picard, 1997; Reeves & Nass, 1996). For example, one

set of studies revolves around how the distribution of information and advice through a particular medium—the computer—impacts advice taking and trust judgments (Lerch et al., 1997a). They revealed a “Turing effect,” in which the characterization of advice as coming from an agent had significant effects on trust (e.g., they trusted the expert systems less than the human experts). Furthermore, evidence revealed possible sources for the Turing effect and demonstrated that manipulations of how an agent is characterized can significantly influence trust judgments (Lerch et al., 1997b). Trust, in this experiment, was a function not of the message content, but of expectations about the nature of the agent delivering the information.

COMPUTATIONAL APPROACH TO TRUST

As we have suggested, one way of understanding organizations, and developing theories of them, is by crafting and analyzing computational models of organizations as collections of agents. Computational modeling (and theorizing) is an invaluable asset to the organizational researcher, as it helps to lay bare the structure and implications of agent properties, adaptation mechanisms, task elements, and organizational characteristics on performance. Computational theorizing helps to systematically reason through the consequences of multiple interacting factors within organizations contributing to (or inhibiting) performance. Furthermore, as both the social interaction and the social capabilities of an agent can be defined (or approximated) computationally (Carley & Newell, 1994; Carley & Prietula, 1994), it is meaningful to conduct computational experiments that address social interaction as well as (associated) sociocognitive agent properties in order to explicate their individual and collective behaviors (Carley & Prietula, in press; Carley et al., 1992; Prietula & Carley, 1994).

Computational theorizing is also clearly useful when theorizing about adaptive or dynamic systems. In such systems, the level of complexity, lack of critical simplifying assumptions, and existence of significant nonlinearities dictate performance and may render the system mathematically intractable. In these situations, computational theorizing acts as an aid by enabling the theorist to systemically explore the space of possibilities by examining the behavior and performance of such systems under varying parametric conditions. Computational theorizing thus can facilitate the study of how human characteristics (like trust) and human-like interaction (communication based exchanges) impact organizational performance under different task and structural constraints.

In the abstracted organizational setting we describe, trust is primarily cognitive and only secondarily affective; therefore it is fundamentally computational (Newell, 1990). The view of trust as primarily a cognitive con-

struct is consistent with empirical findings on the nature of trust in certain organizational settings (e.g., McAllister, 1995).

In addition, we also view trust as a multidimensional construct (Corazzini, 1977). Trust encompasses a number of attributes, including belief in the predictability of others and an expectation that others will act with goodwill. Predictability of others implies that each agent thinks that others will behave as they did in the past. Hence it is reasonable for each agent to predicate their actions on both their own knowledge and on the past actions of others. The expectation of goodwill means that each agent initially presumes that all others are honest. Truthfulness, in our agents, is the default. Further, since goodwill is expected, even if agents think that others have lied, they are going to discount the event. Agents may learn that others lie and so respond to others as liars rather than truth tellers, but it will take a number of instances of lying for that change to occur.

COMPUTATIONAL STUDY OF TRUST

A simulation study was conducted by defining a task to be accomplished by teams of computational agents who can communicate regarding aspects of the task. The task is a generically defined search task that could be specified in a variety of isomorphic forms, such as a search for items in a warehouse (Prietula & Carley, 1994) or as a search for specific resources on the Internet (Carley & Prietula, in press), as depicted in Figure 1. In this study, we used the warehouse interpretation of the task, whereby each agent must proceed to specific order-stack location to obtain a single item-order. The agent must then search the warehouse for the item, moving from "stack to stack," then retrieve the item, deliver the item to a specific delivery location, and return to the order-stack for the next order-item. Agent queues may develop at any stack location.

The simulation was based on a discretized event cycle, where each agent could simultaneously execute a move on each cycle (unless blocked in a queue). The stacks are one move away from each other in a linearly increasing distance from the order-stack, at zero, to the farthest item stack, in this case at ten. An agent can detect an item in a stack only if the agent is at a specific stack location queue, and agent queues do not interfere with this detection. However, an agent can only retrieve an item (or an order) if that agent is at the front of the queue.

Agents were modeled in the following manner. Each agent has an *item memory* (it recalls where it has encountered items), *communication capability* (it can ask and receive advice concerning order-item locations), and a *social memory* (it recalls the accuracy of the advice obtained from other agents). In addition, each agent had an *honesty* construct (whether it would lie or be honest in providing advice) and a *benevolence* construct (a forgiveness algo-

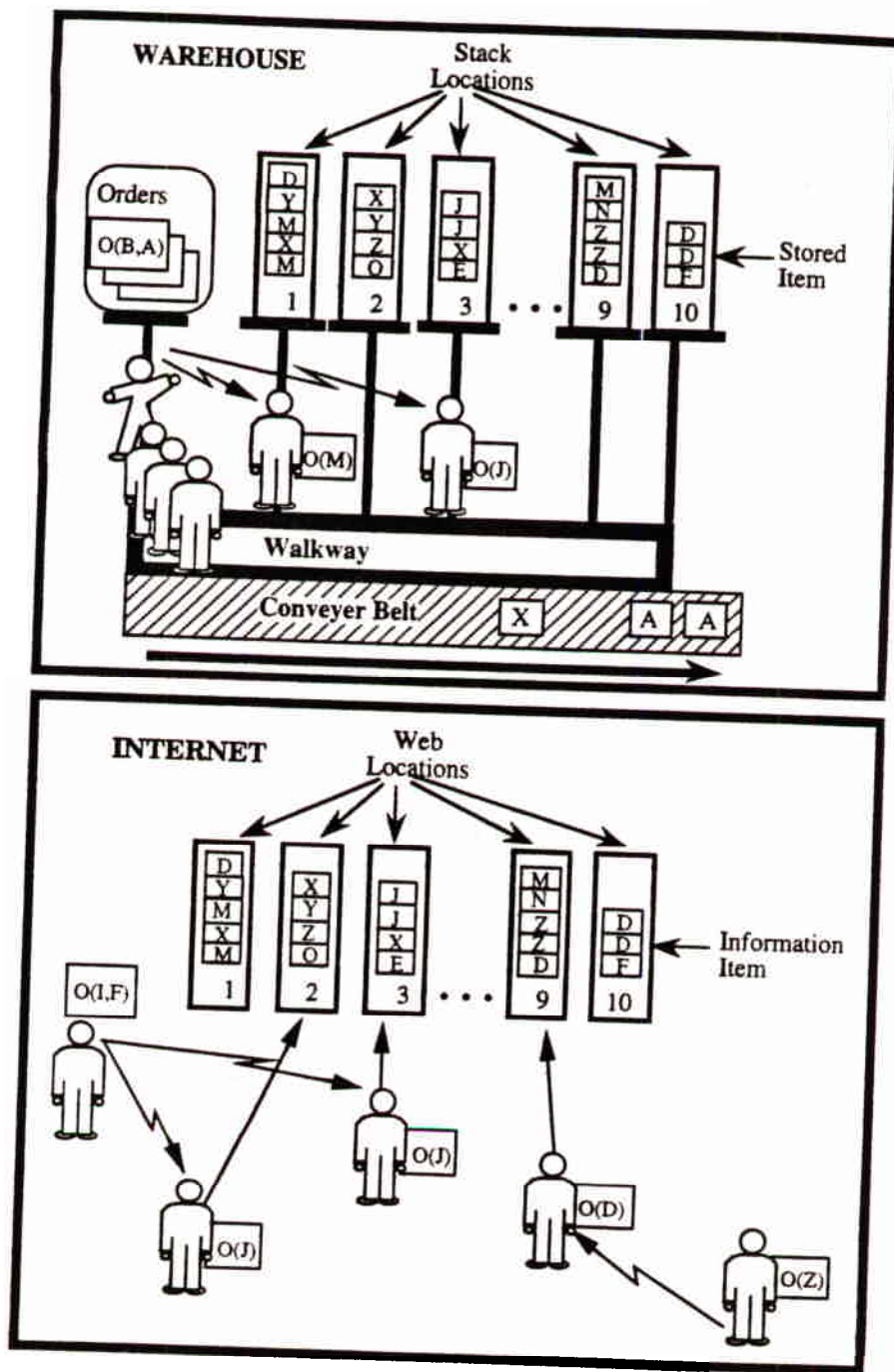


FIGURE 1. Illustrative stylized search task.

rithm for bad advice). All of these elements work together to define the social interaction among agents.

The nature of social interaction was as follows. With respect to the honesty construct, agents are either Liars or they are Honest; thus this construct defines a situationally independent agent propensity. Essentially, this describes how an agent responds to any request for advice. An Honest agent will respond directly to the questioning agent only if it knows the location of the item in question. On the other hand, a Liar agent will respond to *any* request for advice, supplying incorrect location information.

An agent recalls the advice provided by an agent and engages a simple social judgment model of advice acceptance and benevolence, based on three agent judgment states: trustworthy, risky, and untrustworthy. Good/bad advice moves the judgments up/down. Advice from trusty/risky agents is accepted, and questions from them are answered if the agent knows the correct answer (unless the answering agent is a Liar, then no knowledge of the true location is required). Questions from agents judged as untrustworthy are ignored. With respect to the benevolence construct, agents are either Forgiving or Nonforgiving. If an agent is Nonforgiving, the untrustworthy state is absorbing, and an agent deemed untrustworthy remains so judged, so all advice is ignored and no advice is provided to that agent. If an agent is Forgiving, received good advice from a risky agent can alter the judgment of that agent to a trusty state, and received good advice from an untrustworthy agent can alter the judgment of that agent to a risky state.

In addition, we varied the stability of the task environment (i.e., to be Stable or Unstable) in the following manner. In the Unstable condition, as each agent retrieves an item from a particular stack, that agent will disrupt the task environment by restacking interfering items from that stack to another stack. In the Stable condition, an agent simply removes the target item with no subsequent repositioning of interfering items.

The first set of issues we address concerns the nature of various baseline behaviors for the task. For the Stable and Unstable task environment conditions, the task was run with one agent each. This resulted in total moves of 610 (for the Stable) and 617 (for the Unstable) to complete a 20-item task. In this task, a total of 20 orders (no duplicates) were to be filled by searching the warehouse locations. Over those 10 locations were distributed the 20 items (no duplicates) as 2, items per stack, with items 1 and 11 on stack, items 2 and 12 on stack 2, and so forth.

We then examined the effect of group size on homogeneous groups varying agent honesty (Honest, Liar) and agent benevolence (Forgiving, Nonforgiving), while varying agent group sizes as 2, 4, 6, 8, and 10 agents. These were crossed with Stable and Unstable task environment conditions. The results are presented in Figure 2. In Figure 2 the conditions are identified by the 3-tuple, $i-j-k$, where $i = \{L/Liar, H/Honest\}$, $j = \{F/$

tored
item

ormation
1

2
O(Z)

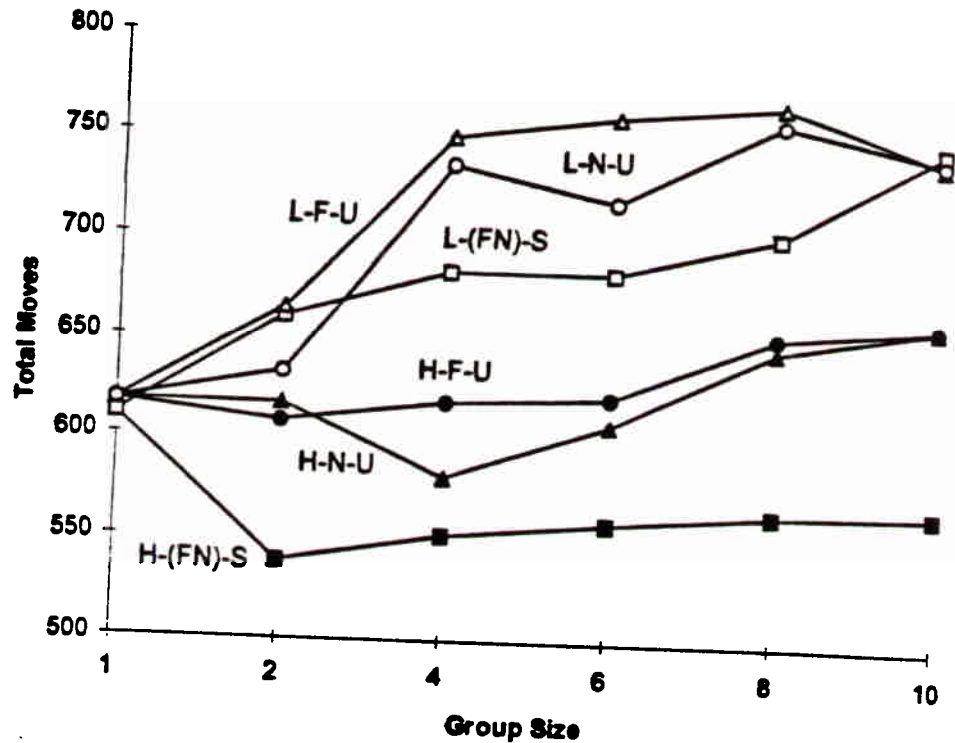


FIGURE 2. Impact of group size on total moves by type of agent.

forgiving, N/nonforgiving agents}, and $k = \{S/\text{stable}, U/\text{unstable environment}\}$. Thus each 3-tuple describes a group of agents, all of which possess the particular characteristics and task situation.

As can be seen from Figure 2, the best performing groups, as may be expected, are groups of Honest agents in Stable environments. In such situations, the benevolence of the agents is not relevant, as no "bad advice" is generated, so no forgiveness is required. In addition, the optimal size for the task is actually the 2-agent group. Adding additional agents does not improve the task, but rather incurs costs resulting from wait times in the queues. However, the additional agents still yield effort (as total moves) savings over a single agent.

For Honest agents, Unstable environments and group size do have differential effects. Unlike the previous condition, a 2-agent group is essentially equivalent to the single-agent situation. The reason for this is the costs beginning to result from bad advice originating from changing conditions in an Unstable task environment. That is, an agent recalls an item from some particular stack and communicates its knowledge, but it does not know that another agent has subsequently moved that item. From Figure 2, we can see that "honest mistakes" can generally be addressed by assuming a non-

forgiving stance. Forgiveness can be a group liability in midsized teams under Unstable task environments. Nonforgiveness has relative improvements up to the 10-agent group. Note that the overall effort exceeds that of a single agent once 8 agents are involved, regardless of their benevolence.

As expected, groups consisting of Liar agents do much worse than groups of Honest agents. Paralleling the prior results, a Stable environment affords relatively lower costs (except for the 2-agent group) under a group of Liar agents, regardless of benevolence. The reason for this, in part, is that although agents generate false information regarding the location of an item, a subsequent search can systematically locate it in a series of steps that are bounded by the starting location (of the agent searching) and the actual location of the item. Thus, for this version of the task (number of items in the task, number of stacks in the task environment), moderately sized groups (2-agent through 8-agent) can exploit task stability over benevolence.

On the other hand, the worst performing groups are agents that are both Forgiving and Liars, closely followed by Nonforgiving agents, in Unstable environments. The Unstable environment makes sequential search (the agents' default mechanism when no acceptable advice is forthcoming) difficult, as the search space is constantly changing. Similarly, Forgiving agents in this Unstable environment is a liability. Again, as the group size approaches 10, the differential effects disappear.

Finally, with 10-agent homogeneous groups, the overall effects converge to three categories: Liar agents (regardless of benevolence or task stability), Honest agents (regardless of benevolence) in Unstable environments, and Honest agents (regardless of benevolence) in Stable environments. Thus, for this task, homogeneous groups are sensitive to the effects of benevolence, lying, and task stability with relatively small group sizes, but certain effects systematically disappear as group size approaches 10. Consequently, we then focused on groups of 10 agents.

In general, and as we have found, it is expected that organizations of Liar agents will do worse than organizations of Honest agents. What is perhaps less obvious is the impact Liar agents would have in a Benevolent group. This was explored as follows. The organization's size was held constant at 10 agents, while the number of Liar agents in the group was varied according to the schedule: 0, 1, 2, 4, 6, 8, and 10 agents. Therefore the extremes considered are the prior reported conditions: all Liar agents, all Honest agents. It is the mix that is of interest.

Additionally, we examined the mix under both Stable and Unstable task situations. The results are presented in Figure 3. In Figure 3 the conditions are identified by the 2-tuple, $i-j$, where $i = \{F/forgiving, N/nonforgiving\}$ and $j = \{S/stable, U/unstable\}$. Included in Figure 3 are control lines for the total moves taken by a single agent (1-Agent), the 10-agent group of Honest agents in a Stable task environment (10-Honest/S),



U/unstable
all of which

10

os, as may be
In such situ-
oad advice" is
al size for the
ents does not
t times in the
total moves)

we do have dif-
p is essentially
is is the costs
g conditions in
em from some
not know that
e 2, we can see
suming a non-

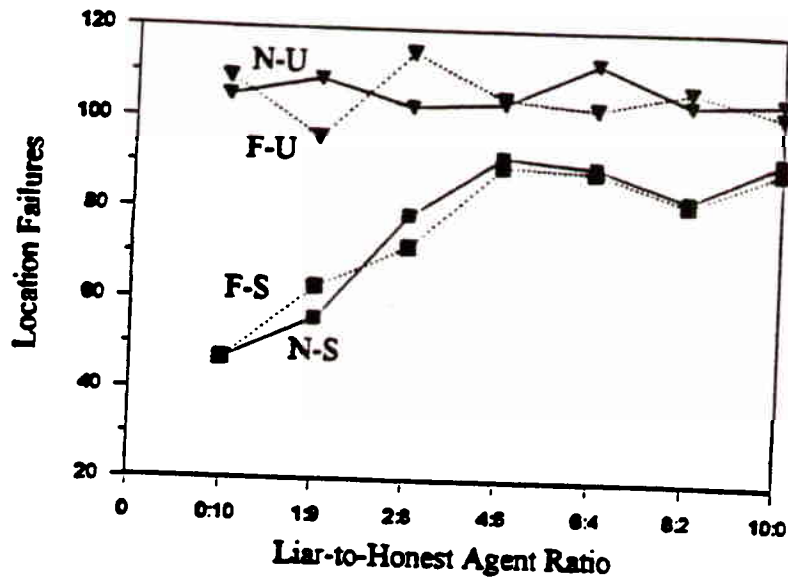


FIGURE 3. Impact of agent ratios on failures by environment.

the 10-agent group of Honest agents in an Unstable environment (10-Honest/U), and the 10-agent group of Liar agents (10-Liars).

As can be seen from Figure 3, the inclusion of a single Liar agent in a group of Honest agents can almost entirely mitigate the effects of environmental stability (i.e., lines F-S and N-S). For Honest agents in an Unstable environment, the inclusion of two Liar agents can achieve levels of effort similar to those of groups made up of 10 Liar agents. Four Liar agents in the Stable task environment achieve such results.

Insight into this can be found by examining the number of location failures or faults that occur. A location failure is when an agent perceives a sought order-item at a particular stack location, is then queued at the stack, and then fails to find the target item when moved to the head of the queue. Whether or not the agent was advised to go there, the effect is that such increases in general location failures result in overall increases in total effort. Figure 4 depicts the number of location failures for the groups. In Figure 4 the conditions are identified by the 2-tuple, $i-j$, where $i = \{F/\text{forgiving, N}/\text{nonforgiving agents}\}$ and $j = \{S/\text{stable, U}/\text{unstable environment}\}$.

Figure 4 reveals the effects of location failures across Stable and Unstable environments. In Unstable environments there seems to be little effect of Liar agents on the number of encountered location failures, while there is a pronounced increase of location failures within the Stable task environment situation. Interestingly, the amount of location failures begins to converge with 4 Liar agents (in the Stable condition) to a value somewhat lower than the Unstable case. What would account for the difference?

The difference can be accounted for (i.e., given that the total moves con-

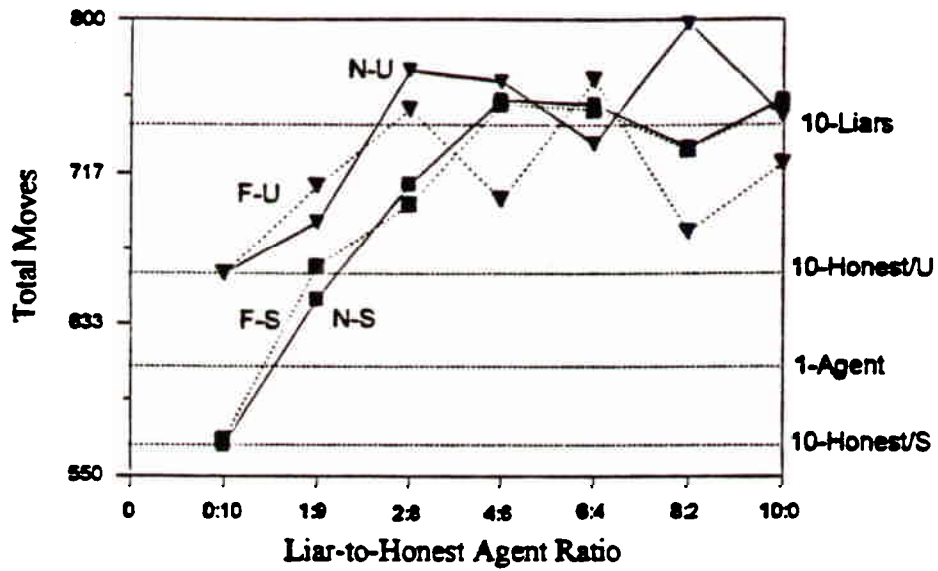


FIGURE 4. Impact of agent ratio on moves by environment.

verge, in Figure 3) by examining the effects of bad advice. In Figure 5 the differences between communication events (Good Advice – Bad Advice) are shown. In Figure 5 the conditions are identified by the 2-tuple, $i-j$, where $i = \{F/\text{forgiving}, N/\text{nonforgiving agents}\}$ and $j = \{S/\text{stable}, U/\text{unstable environment}\}$.

Stack failures may or may not be the result of advice (i.e., a failure could

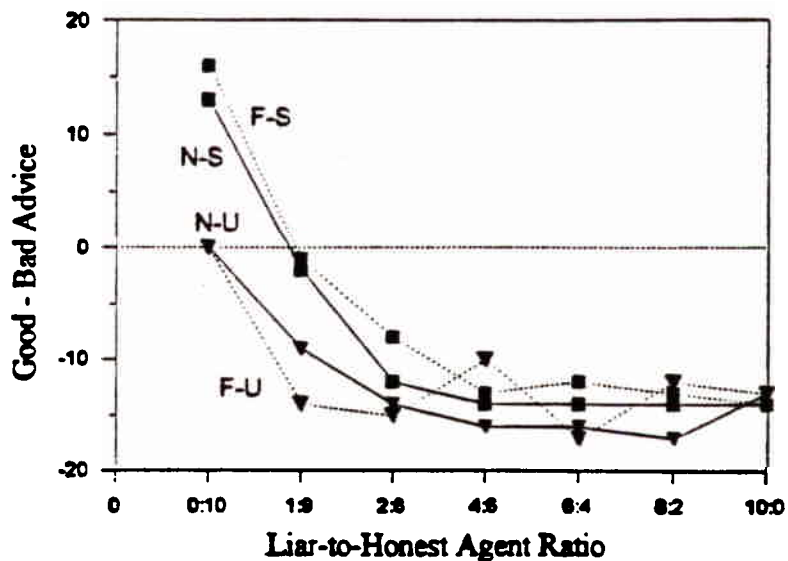


FIGURE 5. Impact of agent ratio on advice quality by environment.



ronment (10-
 liar agent in a
 ts of environ-
 an Unstable
 vels of effort
 Liar agents in

 f location fail-
 nt perceives a
 d at the stack,
 of the queue.
 t is that such
 in total effort.
 s. In Figure 4
 /forgiving, N/
 nt}.
 ss Stable and
 ms to be little
 failures, while
 he Stable task
 failures begins
 alse somewhat
 erence?
 tal moves con-

occur during a sequential search without advice). What advice indicates is that there was a deliberate sequence of moves (i.e., effort) to arrive at a particular location. Bad advice suggests that those were extraneous moves, while good advice suggests that those moves were necessary. Consequently, bad advice often results in increased effort and stack failures. In Figure 5, note that in Stable environments with all Honest agents, good advice greatly facilitates search. However, even with Honest agents, an Unstable environment negates the effect of advice. The addition of Liar agents increases the amount of extraneous search (and therefore effort).

Thus the amount of bad advice generated quickly dominates the organizational communication. At low Liar-to-Honest agent ratio levels, a small number of Liar agents can be extremely disruptive to organizational performance. How this occurs is twofold.

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. Liar agents do not search their memories for the correct answer, nor do they require it. As designed, it is the preference for all agents to initially prefer and trust communication. Without a given advised location, they will engage in a systematic search. However, information provided will cause them to proceed directly to the suggested location (without investigating interim locations, as that would require extraneous moves). This has the effect, for Liar agents, of sending agents to false locations, causing location faults—going to a location without success.

DISCUSSION

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 and determine the extent to which they are amenable to computational form and examination. 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. However, as even our very simple models illustrate, there are interactions between agent-based sources of uncertainty (such as lies) and environmental uncertainty (movement of items) that can alter the effectiveness of the coping mechanisms (benevolence models). Further, in a volatile and uncertain environment, a certain amount of rigidity and inflexibility at the individual level may be beneficial at the groups or team level. These kinds of trade-offs need more exploration.

One of the major limitations of the two models described herein is the lack of a model of motivation. In the models we have described, the agents either are, or are not, liars. If they are liars, then they do not choose to whom they lie or how often they lie. Rather, given the opportunity to provide information, they lie. Other than severe pathological cases, humans do not behave in this fashion, and the situation within which we placed these agents was specifically circumscribed. Nonetheless, even with this very simplified view, we find that it is not always possible for other agents to recognize a lie when they "hear" it. Other types of uncertainty can mask lies and make Honest agents appear to be Liars, and vice versa. Organizational theorists have been quick to point out that structures, processes, and routines can stabilize the task environment to the extent that it is possible for trust to develop and persist (Heide & Miner, 1992). Our argument is truth may persist even in the absence of a stable environment if the various sources of uncertainty serve to mitigate each other.

Incorporating additional motives, although increasing model verticality, will also increase the complexity of the results. Imagine, for example, what might happen if agents were motivated to lie only if that lie was likely to increase their own relative performance. In this case, Liar agents would reserve their lies for others who were high performers. If the Liar happened to be the highest performer, the agent might never lie. In group settings, we might find it behooves a group to have some group members who are liars so that the group as a whole can outperform the other group—an interesting ethical dilemma.

Finally, Barnes (1981) argued that excessive reliance on trust can result in exclusive and dysfunctional reliance on "soft data" over "hard data" in making decisions. For example, one such type of soft data is the opinion of a local expert. In our study, if the environment is volatile, the opinion of any one agent is unlikely to consistently match the underlying reality. Thus trust in a single agent may result in major errors even when that agent is inherently honest. An important extension of the work we have provided would be to examine how much worse or better organizational decisions would be if they were made on the basis of expert opinion or data when the environment is uncertain and one source of that uncertainty is the proportion of agents in the organization that are lying.

CONCLUSION

Turner (1993) suggests that one of the basic problems in multiagent systems is that of resource consumption, and argues that this problem exists for both artificial and human societies. Thus, computational studies about how to avoid this problem will have value to both societal types. We note that issues of uncertainty, and trust, are also basic multiagent problems and

that such issues will impact resource consumption. To the extent that Turner's argument holds, then understanding the impact of truthfulness on organizational behavior is also important for organizations of both humans and artificial agents.

Our argument is simple. In certain situations, trust is first and foremost a cognitive construct. Hence trust is a computational construct. We can therefore explore in a meaningful way the relation between trust and organizational behavior using multiagent computational models. The results described in this paper are a small step in this direction. Although limited, these results show that, at the organizational level, lies are just one source of uncertainty and the mechanisms for reducing uncertainty, whether or not they are aimed at liars, may be performance enhancing. Furthermore, since the impact of lying on organizational performance is a function of the size of the organization, then the cost of tolerating liars may vary with organizational size.

We conclude with a small observation. In our worlds of work, the thought of a set of individuals intentionally and consistently lying may be difficult to accept. We therefore offer a more modest interpretation. Imagine someone at work who consistently offers advice that is unintentionally incorrect (either in whole or in part), as skill is both fleeting and fragile in volatile knowledge environments. Would the organizational effects and responses be substantially different? Perhaps the road to suboptimality is also paved with good intentions.

REFERENCES

- Axelrod, R. 1987. The evolution of strategies in the iterated prisoner's dilemma. In *Genetic algorithms and simulated annealing*, ed. L. Davis. London: Pitman.
- Axelrod, R., and D. Dion. 1988. The further evolution of cooperation. *Science* 242:1385-1390.
- Ballim, A., and Y. Wilks. 1991. *Artificial believers: The ascription of belief*. Hillsdale, NJ: Lawrence Erlbaum.
- Barber, B. 1983. *The logic of trust*. New Brunswick, NJ: Rutgers University Press.
- Barnes, L. B. 1981. Managing the paradox of organizational trust. *Harvard Business Review* 59:107-116.
- Bond, A., and L. Gasser (eds). 1988. *Readings in distributed artificial intelligence*. San Mateo, Calif.: Morgan Kaufmann.
- Bradach, J., and R. Eccles. 1989. Price, authority, and trust. From ideal types to plural forms. *Annual Review of Sociology* 15:97-118.
- Canamero, D., and W. Van de Velde. 1997. Socially emotional: Using emotions to ground social interaction. In *Socially intelligent agents: Papers from the 1997 AAAI fall symposium*, ed. K. Dautenhahn. Technical Report FS-97-02. Menlo Park, Calif.: AAAI Press.
- Carley, K. 1991a. A theory of group stability. *American Sociological Review* 56:331-354.
- Carley, K. 1991b. Designing organizational structures to cope with communication breakdowns: A simulation model. *Industrial Crisis Quarterly* 5:19-57.
- Carley, K., and D. Krackhardt. 1996. Cognitive inconsistencies and non-symmetric friendship. *Social Networks* 18:1-27.
- Carley, K., J. Kjaer-Hansen, A. Newell, and M. Prietula. 1992. Plural-soar. A prolegomenon to artificial agents and organizational behavior. In *Artificial intelligence in organization and management theory*, eds. M. Masuch and M. Warglien. Amsterdam: North-Holland.

- Carley, K., and Z. Lin. 1995. Organizational designs suited to high performance under stress, *IEEE Transactions on Systems, Man, and Cybernetics* 251:221-230.
- Carley, K., and A. Newell. 1994. The nature of the social agent. *Journal of Mathematical Sociology* 194:221-262.
- Carley, K., and M. Prietula. 1994. ACTS theory: Extending the model of bounded rationality. In *Computational organization theory*, eds. K. Carley and M. Prietula. Hillsdale, NJ.: Lawrence Erlbaum.
- Carley, K., and M. Prietula. In press. WebBots, trust and organizational science. In *Simulating societies: Computational models of institutions and groups*, eds. M. Prietula, K. Carley, and L. Gasser. Cambridge, Mass.: AAAI/MIT Press.
- Carley, K., M. Prietula, and J. Lin. In press. Design versus cognition: The interaction of agent cognition and organizational design on organizational performance. In *Evolving societies: The computer simulation of social systems*, eds. R. Conte and E. Chattoe.
- Carley, K., and D. Svoboda. 1996. Modeling organizational adaptation as a simulated annealing process. *Sociological Methods and Research* 251:138-168.
- Corazzini, R. 1977. Trust as a complex multi-dimensional construct. *Psychological Reports* 40:75-80.
- Dautenhahn, K. (ed.). 1997. *Socially intelligent agents: Papers from the 1997 fall symposium*. Technical Report FS-97-02. Menlo Park, Calif.: AAAI Press.
- Frazer, V. 1996. Employees value workplace relationships. *Personnel Journal* 756:25.
- Heide, J., and A. Miner. 1992. The shadow of the future: Effects of anticipated interaction frequency of contact on buyer-seller cooperation. *Academy of Management Journal* 352:256-291.
- Huhns, M., and M. Singh (eds.). 1998. *Readings in agents*. San Francisco, Calif.: Morgan Kaufmann.
- Kaufer, D., and K. Carley. 1993. *Communication at a distance: The effect of print on socio-cultural organization and change*. Hillsdale, NJ.: Lawrence Erlbaum.
- Kepphart, J., B. Huberman, and T. Hogg. 1992. Can predictive agents prevent chaos? In *Economics and cognitive science*, eds. P. Bourguin and B. Walliser. Oxford, England: Pergamon Press.
- Klein, D. (ed.). 1997. *Reputation: Studies in the voluntary elicitation of good conduct*. Ann Arbor, Mich.: University of Michigan Press.
- Krackhardt, D., and R. Hanson. 1993. Informal networks: The company behind the chart. *Harvard Business Review* (July-August):104-111.
- Kramer, R., and T. Tyler (eds.). 1996. *Trust in organizations: Frontiers of theory and research*. Thousand Oaks, Calif.: Sage.
- Lerch, J., M. Prietula, and C. Kulik. 1997a. The Turing effect: The nature of trust in machine advice. In *Expertise in context: Human and machine*, eds. P. Feltoovich, K. Ford, and R. Hoffman. Cambridge, Mass.: AAAI/MIT Press.
- Lerch, J., M. Prietula, J. Kim, and T. Buzas. 1997b. Unraveling the Turing effect: Measuring trust in machine advice. Working Paper, Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, Pa.
- Lewis, J., and A. Weigert. 1985a. Trust as a social reality. *Social Forces* 63:966-985.
- Lewis, J., and A. Weigert. 1985b. Social atomism, holism, and trust. *Sociological Quarterly* 26:455-471.
- Lin, Z., and K. Carley. 1993. Proactive or reactive: An analysis of the effect of agent style on organizational decision making performance. *International Journal of Intelligent Systems in Accounting, Finance and Management* 24:271-288.
- Luhmann, N. 1979. *Trust and power*. New York: Wiley.
- Macy, M. 1991a. Learning to cooperate: Stochastic and tacit collusion in social exchange. *American Journal of Sociology* 973:808-843.
- Macy, M. 1991b. Chains of cooperation: Threshold effects in collective action. *American Sociological Review* 56:730-747.
- March, J., and R. Weissinger-Baylon (eds.). 1986. *Ambiguity and command: Organizational perspectives on military decision-making*. Marshfield, Mass.: Pitman.
- McAllister, D. 1995. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal* 381:24-59.
- Mintzberg, H. 1973. *The nature of managerial work*. New York: Harper and Row.
- Newell, A. 1990. *Unified theories of cognition*. Cambridge, Mass.: Harvard University.
- Nicholas, T. 1993. *Secrets of entrepreneurial leadership: Building top performance through trust and teamwork*. Chicago, Ill.: Enterprise Dearborn.

- Pfeffer, J., G. Salancik, and H. Leblebici. 1976. The effect of uncertainty on the use of social influence in organizational decision-making. *Administrative Science Quarterly* 21:227-245.
- Picard, R. 1997. *Affective computing*. Cambridge, Mass.: MIT Press.
- Prietula, M., and K. Carley. 1994. Computational organization theory: Autonomous agents and emergent behavior. *Journal of Organizational Computing* 41:41-83.
- Reeves, B., and C. Nass. 1996. *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge, England: Cambridge University Press.
- Rempel, J., and J. Holmes. 1986. How do I trust thee? *Psychology Today* (February):28-34.
- Rempel, J., J. Holmes, and M. Zanna. 1985. Trust in close relationships. *Journal of Personality and Social Psychology* 49:95-112.
- Ring, P., and A. Van de Ven. 1992. Structuring cooperative relationships between organizations. *Strategic Management Journal* 13:483-498.
- Rotter, J. 1971. Generalized expectancies of interpersonal trust. *American Psychologist* 26:443-452.
- Shapiro, S. 1987. The social control of impersonal trust. *American Journal of Sociology* 93:623-658.
- Shoham, Y., and M. Tennenholtz. 1994. Co-learning and the evolution of social activity. Stanford University, Department of Computer Science, Technical Report STAN-CS-TR-94-1511. Stanford, Calif.
- Skvoretz, J., and T. Fararo. 1995. The evolution of systems of social interaction, In *Current perspectives in social theory*, ed. B. Agger, vol. 15. Greenwich, Conn.: JAI Press.
- Stinchcombe, A. 1990. *Information and organizations*. Los Angeles, Calif.: University of California Press.
- Turner, R. 1993. The tragedy of the commons and distributed AI systems. Technical Report 93-01. University of New Hampshire, Department of Computer Science, Durham, N. H.
- Zaheer, A., and N. Venkatraman. 1995. Relational governance as an interorganizational strategy: An empirical test of the role of trust in economic exchange. *Strategic Management Journal* 19:373-392.
- Zaheer, A., B. McEvily, and V. Perrone. In press. Does trust matter? Exploring the effects of inter-organizational and interpersonal trust on performance. *Organizational Science*.
- Zucker, L. 1986. Production of trust: Institutional sources of economic structure, 1840-1920. In *Research in organizational behaviour* 8, L. L. Cummings and B. M. Staw, eds. Greenwich, Conn.: JAI Press.

State
ANDR
UK for
47 Run
Philadel
Surfson
Chestnu
Philadel
percent
Gunpow
precedin
dealers
circulatio
other me
distribute
actual nu
printed: 7
NONE; C
(E) Free
(G) Total
from new
complete