A Computational Model of Trust and Rumor

M. J. Prietula
Fisher School of Accounting
267 Business Building
University of Florida
Gainesville FL 32611
prietula@ufl.edu

K. M. Carley
Social and Decision Sciences
Carnegie Mellon University
Pittsburgh PA 15213
Kathleen.Carley@CS.CMU.EDU

Abstract

In this paper we explored the impact of various Agent trust models and propensity to lie, within the context of a rumor augmented communication structure and environmental volatility. Results demonstrate main effects of volatility and lying on increasing organizational cost, and a main effect of rumor on decreasing organizational cost. Interactions revealed that rumor attenuates effect of liars. Some effects altered as group size increased.

Introduction

Our work focuses on understanding the relationship between information and knowledge exchange, and organization performance by crafting computational models of individual agents (Carley & Prietula, 1994a; Prietula, Carley & Gasser, 1998; Zhu, Prietula & Hsu, 1997). Building upon a series of research efforts that explicate and model trust and its effect on individual and group behaviors and performance, we describe an expansion to this work that takes into account emotional influences of trust as agent constructs in a rumor network under varying group sizes and task stability.

First, a few definitions. As one might well imagine, there are a host of definitions for "emotion." However, as our theory is based on melding human cognition and agent interaction, our view of *emotion* is most similar to that articulated by Ortony, Core and Collins (1988) as "valenced reactions to events, agents, or objects, with their particular nature being determined by the way in which the eliciting situation is construed (p. 13)." In our model an Agent reacts to events (i.e., advice and rumor) and the Agents involved in those events. The valenced reactions are the varying degrees to which an Agent responds to those events. Emotions, from this vantage, have a strong cognitive component as the individual's knowledge affects the response.

Likewise, *trust* has a host of definitions. In our work we generically define (interpersonal) trust as the Agent's ability, given attributional information, to act on predictions and make predictions that other Agents will act in a cooperative, helpful manner. Trust is primarily a cognitive construct and secondarily an emotional one (Prietula & Carley, in press). Cognition and emotion work together in effecting the three components of trust (Remple, Holmes & Zanna, 1985): predictability, dependability, and faith. Predictability refers to the most concrete and observable dimension of trust – predicable agent behavior over time in a stable environment. Dependability refers to attributions made about the agent (personal characteristics) that contribute to the observed behaviors. The Agent's emotional state impacts this attribution by affecting the perception of dependability.. Finally, Faith is predominantly an emotional response and is less rooted in available evidence and dispositional attributions than either Predictability or Dependability.

Finally, this effort is part of a broader examination of ACTS theory (Carley & Prietula, 1994b). ACTS theory is an extension of bounded rationality that views organizations as collections of intelligent agents who are cognitively restricted, task oriented, and socially situated. Furthermore, ACTS theory is consistent with (subsumes) the argument that the vast complexity of task and social situations is necessary for many characteristically human (social) behaviors to emerge (Carley & Newell, 1994). ACTS theory proposes that cognitive-based deliberations (individual and social) are renderable in computable forms.

Trust, Advice, and Rumor

Emotion and trust are highly contextual and may have both individual and social consequences. The contextuality is (properly) suggested by the definition provided by Ortony et al. (1988). Thus, the exact nature of how emotion and trust unfold depends on the task. Furthermore, data from empirical studies on that task (i.e., advice) served as a basis for subsequent crafting of the emotional and attributional forms in the simulation.

In Lerch, Prietula and Kulik (1997), a series of human-computer interaction experiments revealed a "Turing Effect" in which the characterization of advice had significant effects on trust. Differentially characterizing the source of the computer-based advice (expert, novice, and

To appear in Proceedings of the 1998 AAAI Fall Symposium Series, Emotional and Intelligent: The Tangled Knot of Cognition, October 23-25, 1998.

expert system) affected how the receiver trusted the advice. Furthermore, evidence revealed that the cause for this Turing Effect is based on how specific attributions were made regarding the Situation, Knowledge, and Reasoning and certain effects are time-dependent (Lerch, Prietula, Kim & Buzas, 1998). Predictability impacts Dependability, and Dependability impacts Faith. Both direct experience and information about the source of advice affect trust.

Rumor (or gossip) generally involves the provision of information by one agent to another about a third (Wittek & Wielers, 1998). How is trust related to rumor? Information impacts trust, and rumor spreads information. Thus, both the content and timing of information dissemination can potentially impact trust. The content and timing of rumor is based on the results of Turing Effect experiments. Emotional components impact trust, and trust has both and individual and social consequences. In our model, a rumor about an Agent (here, always a negative attribution) can result in changes in cooperation.

The Initial Study

The basic task involved a set of five agents, where each agent acquires an order for an item at the order stack, seeks the item in a warehouse, retrieves that item, then proceeds back to the order stack for the next request. Manipulations were made at the group level (i.e., homogeneous Agents within manipulation): Rumor (all assert, none assert – a rumor is a truthful opinion based on experience that an Agent is a Liar), Disruption (all disrupt item locations when retrieving, none disrupt – defining environmental stability), and Benevolence (3 emotion-based models). The Benevolence models defined emotional sensitivity to bad advice, resulting in adjustment to trust levels for an Agent. The 3 models were: (1) always trusty, (2) trusty or untrusty, (3) trusty, risky, untrusty. Agents would have a preference to ask (broadcast) if any Agent knows the location of an item. In addition, the prior three manipulations were crossed with 5 levels of Liar ratios (no Liar agents to all Liar agents). Liar agents, when asked, would lie about the location of a sought item. Dependent variables were total Time (maximum number of moves for any Agent in the group) and organizational Effort (total number of moves by the group).

Results and Conclusion

Main effects were found for Rumor on Effort (Rumor reduced organizational Effort but not total Time); Disruption on Time and Effort (Disruptive Agents increased Effort and Time); Liars on Time and Effort (Liars increased Effort and Time). Additionally, Rumor interacted with the Liar manipulation (Rumor reduced effect of Liar agents on Effort and Time). No main effects were found for Benevolence.

Sometimes task environments are unstable and Agents may easily (though not maliciously) disrupt it. In such cases, communication capability can actually inhibit organizational performance. Liars in an organization reflect a "false disruption" and similarly impact organizational performance. However, Rumor reduces the organizational cost of Liars. The lack of strong Benevolence effects was seen to be a problem with valence: all models were too forgiving and the task could not tease out the emotional impact on trust. Subsequent research will add less forgiving Benevolence models, extend the time-frame for the task to allow emotional impact to occur, and adjust Rumor sensitivity.

References

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

Carley, K. & M. Prietula (Eds.) (1994a). *Computational Organization Theory.* Hillsdale, NJ: Lawrence Erlbaum.

Carley, K. & M. Prietula (1994b). ACTS Theory: Extending the Model of Bounded Rationality. In K. Carley and M. Prietula (Eds.), *Computational Organization Theory* (pp. 55-87). Hillsdale, NJ: Lawrence Erlbaum.

Lerch, J., M. Prietula, & C. Kulik (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* (pp. 417-448). Cambridge, MA: AAAI/MIT Press.

Lerch, J., M. Prietula, J. Kim & T. Buzas (1998). *Unraveling the Turing Effect: Measuring Trust in Machine Advice.* Working Paper, Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh PA.

Ortony, A., G. Clore & A. Collins (1988). *The Cognitive Structure of Emotions*. Cambridge, England: Cambridge University Press.

Prietula, M. & K. Carley (in press). Exploring the effects of agent trust and benevolence in a simulated organizational task. *Applied Artificial Intelligence.*

Prietula, M., K. Carley & L. Gasser (Eds.) (1998). *Simulating Organizations: Computational Models of Institutions and Groups.* Cambridge,MA: AAAI/MIT Press.

Remple, J., J. Holmes & M. Zanna (1985). Trust in close relationships. *Journal of Personality and Social Psychology,* 49(1), 95-112.

Wittek, R. & R. Wielers (1998). Gossip in organizations. *Computational and Mathematical Organization Theory*, 4(2), 189-204.

Zhu, D., M. Prietula, M. & W. Hsu (1997). When processes learn: Steps toward crafting an intelligent organization. *Information Systems Research*, 8(3), 302-317.