

# A Socio-Technical Approach to Theorizing on Perceptions of Trustworthiness in Virtual Organizations

Shuyuan Mary Ho

Syracuse University

325 Hinds, School of Information Studies

Syracuse, NY 13244-4100

(315) 443-7267

smho@syr.edu

## ABSTRACT

This paper presents a socio-technical study about perceptions of human *trustworthiness* as a key component for countering insider threats in virtual collaborative context. This study focuses on understanding how anomalous behavior can be detected by observers in a close social network. While human observations are fallible, this study adopts the concept of human-observed changes in behavior as analogous to “sensors” on a computer network. Using online team-based game-playing, this study seeks to re-create realistic situations in which human sensors have the opportunity to observe changes in the behavior of a focal individual – in this case a team leader. Four sets of experimental situations are created to test hypotheses. Results of this study may lead to the development of semi-automated or fully-automated behavioral detection systems that attempt to predict the occurrence of malfeasance.

## Categories and Subject Descriptors

H.1 [Models and Principles]. Systems and Information Theory, User / Machine System – *human factors*.

## General Terms

Theory, Design, Security, Human Factor

## Keywords

Perceived Trustworthiness, Virtual Organization, Socio-Technical Approach, Insider Threats

## 1. INTRODUCTION

Trust in organizations is critical because it enables individuals the ability to collaborate with one another. Conversely, the effects of the breakdown of trust in organizations have been well documented in current business organizations. According to the 2007 CSI Survey, financial losses caused by computer crime soared to \$67 million in 2007, up from \$52.5 million in 2006 [18]. Among those losses, nearly 37 percent of respondents attributed more than 20 percent of losses to be caused by insiders. This indicates an increase of insider abuse within network resources, from 42 to 59 percent compared with the 2006 CSI/FBI Survey. Insider misuse of authorized privileges or abuse of network access has caused significant damage and loss to corporate internal information assets. While employees are essential to the productive operation of an organization, their inside knowledge of corporate resources can also threaten corporate security, as a

result of improper use of information resources. Such improper uses are often termed by security experts as “insider threats.” When trust is violated, members can no longer collaborate well in organizations. Thus, having the ability to know an individual’s trustworthiness will enable an organization to achieve its business goals and enhance its productivity.

As the world moves towards virtual organizations, whether it is a far-flung corporate organization operating through cyber-communications or a multi-organizational collaboration to achieve pre-defined goals, the effects of trust among the individuals is both more of a problem and an opportunity. The larger problems stem from the adverse social effects of virtual communication in building trust; the larger opportunities lie in the potential that the virtual communication captures indicators or precursors of likely threats in conversations. Such precursors in conversations could lead to the detection of problems in lack of trust or in the detection of ways to build trust, and thus enhance the effectiveness of the virtual communication and productivity of virtual organizations.

This paper contains six major sections describing this socio-technical study. Since the problem gap of this research is aimed at the insider threat, my research question is raised to understand this phenomenon as stated in the *Problem-based Question*. I then synthesize the theoretical foundation of trustworthiness attribution in the *Theoretical Framework*. In the *Method* section, I describe about how the “Leader’s Dilemma” game is designed. I offer my hypotheses, and how my experiments are designed to test my hypotheses in the *Experimental Factorial Design* section. The preliminary result of this study is discussed in the *Discussion of Preliminary Results*. The *Conclusion* section summarizes this research in progress.

## 2. PROBLEM-BASED QUESTION

The phenomenon of insider threats is a social, human behavioral problem [6, 7, 13]. In the *Insider Threat Study by CERT (2004-2005)*, the US DoD<sup>1</sup>, DHS<sup>2</sup>, and Secret Service investigated various insider threat cases and discovered that embedded in a mesh of communications, a person given high social power but with insufficient trustworthiness can create a single point of trust failure [11, 16]. Thus, “insider threat” as an organizational

---

<sup>1</sup> US DOD stands for the US. Department of Defense.

<sup>2</sup> DHS stands for the Department of Homeland Security.

problem is defined as a situation where a critical member of an organization with authorized access, high social power and holding a critical job position, inflicts damage within an organization. In a way, this critical member behaves against the interests of the organization, generally in an illegal and/or unethical manner.

This study, based on the above definition, examines basic mechanisms for detecting changes in the trustworthiness of an individual who holds a key position in an organization, by observing overt behavior – including communication behavior – over time. Since Steinke [21] suggests that it is possible to detect cheating behavior without directly observing the individual, the overarching question is: *What changes of behaviors can reflect a downward<sup>3</sup> shift in the trustworthiness of a critical member in a virtual or physical organization which might signal possible insider threats?* My hypothesis is that the downward shift in a person’s trustworthiness can be reflected in his or her behavior. And, the inconsistency and unreliability in this actor’s unexpected behaviors when compared to his or her communicated intentions can be detected by the observers’ subjective perceptions over time. The observers refers to the members of his or her close social network.

### 3. THEORETICAL FRAMEWORK

In order to understand how people observe a target individual’s behavior over time, make inferences about changes in behavior that signify something abnormal, and be able to predict a likelihood of a downward shift of the target’s intention as reflected in his or her behavior [1, 2], attribution theory is adopted to look at an aspect of a basic human relationship, trust. The observer assigns the target’s behavior to a cause and it may suggest a possible threat if the attributed cause is abnormal. The trustworthiness of the target’s intention is perceived, attributed and assigned with meaning by his or her social network [8, 9, 17]. The theoretical framework of this research is introduced by reviewing trustworthiness and differentiating trustworthiness from trust. Additionally, attribution theory is discussed within the context of a trust relationship. Then, the framework sets the foundation of my hypotheses and research design.

Rotter [20] asserted that “trust and trustworthiness are closely related,” but trust depicts a relationship among two or multiple parties or actors while trustworthiness is an attribute or a quality of a person. Trustworthiness<sup>4</sup> is defined as a generalized expectancy concerning a target person’s degree of correspondence between communicated intentions and behavioral outcomes that are observed<sup>5</sup> and evaluated, which remain reliable, ethical and consistent, and any fluctuation between target’s intentions and actions does not exceed the observer’s expectations over time [3, 10, 19].

As Figure 1 depicted, Mayer, Davis and Schoorman [15] further defined three factors of perceived trustworthiness to be ability (competence), benevolence (kindness) and integrity

(goodwill/ethics). Mayer and Davis [14] found a significant impact of the appraisal system’s acceptability on trust for management, which was mediated by the factors of perceived trustworthiness. The implication of this finding was that trust is constantly influenced by the combination of competence, benevolence and integrity.

Attribution theory is adopted to understand how people attribute (or assign) the causes of others’ behaviors [4, 5]. The attribution of the target’s behavior by observers is determined by observers’ judgment that the target intentionally or unintentionally [4] behaves in a way that is attributable to either external (situational) causality or internal (dispositional) causality [12]. Because all human beings are of the same species and born with similar types of features and functions, man should “know” and be able to “sense” from his own perceptions and with his judgment of how the world operates [12, 4] despite the fact that sometimes those attributions may not be accurate or valid. Moreover, an individual’s observed behavior can be interpreted and perceived in a single observation – or through multiple observations over time. The theoretical framework of this study adopts these principles in multiple observations: *distinctiveness, consensus, and consistency*.

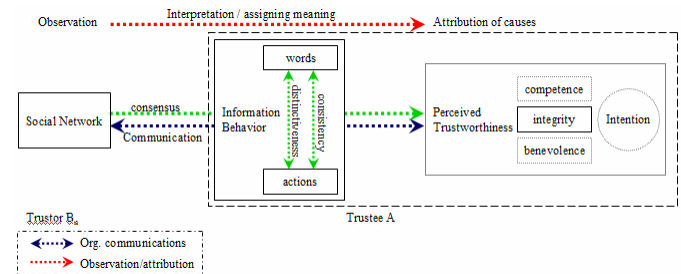


Figure 1: Theoretical Framework

In Figure 1, the relationship of the constructs is represented by arrows. Blue arrows represent communication between the target and the social network. Red arrows represent the observation and attribution by members of the social network regarding the target’s behavior. Green arrows represent three attribution principles depicted by Kelley [12]. In this framework, the communications among the target’s social network (including communications to and from the target) sheds light on the target’s perceived trustworthiness. In other words, members of the social network attribute (or assign) meaning to the target’s trustworthiness level based on their observations of the target’s behavior.

### 4. METHOD

This study adopts a positivist view to identify the indicators of abnormal behavior and the basic criteria of trustworthiness assessment. The leader’s Dilemma game was a simulated, controlled situation created to test how leader’s trustworthiness was perceived by his team members. In my definition, a virtual organization (VO) refers to a group of individuals whose members and resources may be dispersed geographically, but function as a coherent unit through the use of cyber-infrastructure. This group of individuals is team-based and goal-oriented, where leaders and subordinates work together to achieve pre-determined goals. A design of this experimental setting is depicted in Figure 2, where a virtual contest was launched. In these experimental

<sup>3</sup> Same as *unethical*, or *illegal*.

<sup>4</sup> Trustworthiness is portrayed and defined as “reliable, dependable, responsible, loyal, honorable, ethical, moral and incorruptible.”

<sup>5</sup> Same as *perceived*.

settings, Game-Master (G) is the role to direct the dynamics of the virtual competition. Experimenter (M) takes on the role of a judge in these online games. In these experiments, Team-Leader (A) is the target, who is appointed from among the team participants by the Game-Master. Team members are observers ( $B_n$ ), whom work with Team-Leader in achieving their pre-determined goals.

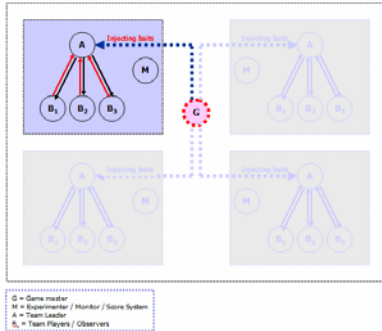


Figure 2: Experimental Control Room Design

A pilot study<sup>6</sup> (n=5) and a full-scale experiment<sup>7</sup> (n=26) were simulated based on the above definition of virtual organization. The goal of this experiment was for a team to solve brain teasers, as their task assignments, in a given timeframe. The virtual contest was manipulated in an online game environment<sup>8</sup> to reach its climax when a dishonesty gap was forcefully created by offering “bait” to the Team-Leader. A conflict of interest between the Team-Leader and the team members forcefully causes the Team-Leader to face “ethical dilemma” in making decisions (Figure 3). Bait was used in a form of a micro-payment system, which connects a monetary value to the real-world rewards. The concept of sting operation is implemented in the game to enhance the group sensitivity. A mole player is embedded in the team to question the leader, raise tensions and stir up discussions in the team. This will enhance awareness within team members of what the leader is doing or thinking. Moreover, peer influence could be enhanced by having a third Team-Leader chatting and persuading this target to accept the bait and betray his team. With the awareness of knowing that this is the critical point in determining whether this experiment is successful or not, the bait given to the leader has to be invisible to the teams, and sufficiently tempting that the leader will risk taking the bait.

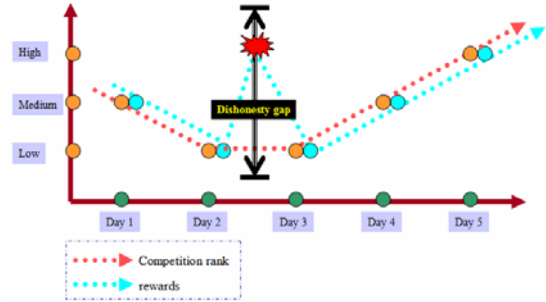


Figure 3: Logics for Virtually Controlled Contest

## 5. EXPERIMENTAL FACTORIAL DESIGN

The same research settings designed in the pilot is simulated in the full-scale experiment. A  $2 \times 2 \times 3$  factorial design is developed to generalize the findings. The dependent variable (response) is target’s perceived trustworthiness (Y) in terms of taking the bait. Factor 1 is the observers’ attribution ( $X_1$ ) in terms of group sensitivity toward target’s words (target’s communicated intentions), factor 2 is the observers’ attribution ( $X_2$ ) in terms of group sensitivity toward target’s actions (target’s information behavior). Factor 3 is the time ( $X_3$ ).

Four sets of simulated case studies of online games were planned to be conducted (Figure 4). In other words, 12 sets of group observations were obtained. While the dependent variable (response) is target’s perceived trustworthiness, major independent variables (factors) include: the bait ( $B_0$  and  $B_1$ ) as the treatment, a mole that increases or decreases group sensibility ( $S_1$  and  $S_2$ ) by either encouraging or discouraging conversations about the team-leader, and time ( $T_1$ ,  $T_2$  and  $T_3$ ) representing measurement obtained from each day, in particular, after conflict of interest between the team-leader and the team members is created.

Treatment	without treatment/bait; ( $B_0$ )			with treatment/bait; ( $B_1$ )		
	Time ( $T_1$ )	Time ( $T_2$ )	Time ( $T_3$ )	Time ( $T_1$ )	Time ( $T_2$ )	Time ( $T_3$ )
Increase group sensibility ( $S_1$ )	Group 1 Average			Group 3 Average		
Decrease group sensibility ( $S_2$ )	Group 2 Average			Group 4 Average		

Figure 4:  $2 \times 2 \times 3$  factorial design

These hypothesized situations can be explained in the following. There is no bait given to the team-leader in Group 1, but the group sensitivity is enhanced through encouraging discussion about the leader. I hypothesize that attribution from Group 1 toward their team-leader’s perceived trustworthiness will rise. The perceived trustworthiness of the leader remains the same or slightly dropped. Likewise, there is no bait given to the team-leader in Group 2, but the group sensitivity is decreased through discouraging discussion about the target. I hypothesize that the attribution from Group 2 toward the target’s perceived trustworthiness should remain the same or relatively higher. As for Group 3, there is a bait given to the team-leader and the group sensitivity is enhanced through encouraging discussion about the target. I hypothesize that the attribution from Group 3 toward the perceived trustworthiness should drop significantly. Finally, the

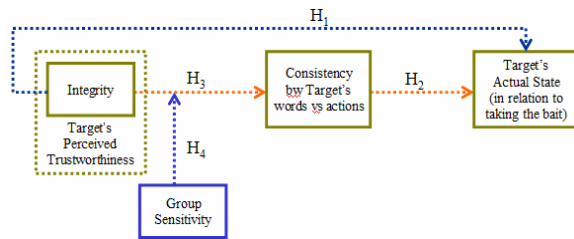
<sup>6</sup> Syracuse University IRB#07-276, conducted in Fall 2007.

<sup>7</sup> Syracuse University IRB#07-276, conducted in Fall 2008.

<sup>8</sup> The Learning Management System (<http://ischool.syr.edu/learn/>) is hosted by School of Information Studies, Syracuse University.

target team-leader in Group 4 is given a bait but the group sensitivity is reduced through discouraging questioning about the target. I hypothesize that the attribution toward the target's perceived trustworthiness in Group 4 might also be relatively dropped.

In this research design, I tested the following hypotheses. My main hypothesis is that the downward shift in a person's trustworthiness can be reflected in his or her behavior. And, the inconsistency and unreliability in this actor's unexpected behaviors when compared to his or her communicated intentions can be detected by the observers' subjective perceptions over time. There are four hypotheses, which support my main hypothesis (Figure 5).



**Figure 5: Hypotheses**

**Hypothesis 1 (H<sub>1</sub>):** There is a positive relationship between the target's actual state and the group observation of the target's perceived trustworthiness, in terms of his or her integrity. This means that if the target has taken the bait, it can be successfully attributed by the observers over time. If the target has not taken the bait, it will not trigger any suspicion in observers' attribution.

**Hypothesis 2 (H<sub>2</sub>):** When target's actual state is positive (meaning that he has taken the bait), the group can reach consensus about target's inconsistency between communicated intentions (words) and information behavior (actions).

**Hypothesis 3 (H<sub>3</sub>):** When target's perceived trustworthiness is relatively low, observers tend to attribute inconsistency in his or her words and actions.

**Hypothesis 4 (H<sub>4</sub>):** The group sensitivity has a significant influence on the perceived trustworthiness, in terms of integrity, of the target. The higher the group sensitivity is, the more likely for the group to detect inconsistency between target's words and actions.

A detail discussion of these hypotheses are planned to be discussed in the future work.

## 6. DISCUSSION OF PRELIMINARY RESULTS

During the 5-Day game, the participants did not know that the target Team-Leader was manipulated by the Game-Master – which occurred in the background. Since insufficient evidence existed regarding the target, the resulting perceptions depended on whether the target's behavior was generally reliable or ethical, and the outcome itself. The perception of the target's behavior was positive from Day 1 through Day 3 during the experiment. The observers' attribution of the target was seen as being trustworthy. This inferred that the target's competence in leading the team was found to be satisfactory, and his communicated intention was found to be consistent in terms of his information behavior. In this sense, the “anomalous” behavior was not found

to be significant. However, the outcome of the target's behavior showed negative on Day 4 and 5, and the level of the target's integrity dropped as a result of taking the bait despite internal ethical struggles. Thus, the target's anomalous behavior was found to be significant.

The swift trust developed amongst the team players towards the target was caused by the leadership halo effect - until Day 3 or Day 4. At this point, the target showed signs of dishonesty on a couple of occasions. In addition to the micro-currency given to the target, the Game-Master used a negative team evaluation appraisal strategy on the target. This negative appraisal evaluation on the target from his team members stirred up his disgruntled feelings about his team. However, the Game-Master showed understanding to the human weaknesses. He won the target's trust.

Interestingly, the outcome of the target's behavior went negative on the last day of this pilot, when the level of the target's integrity dropped as a result of taking the bait despite internal ethical struggles. The target made four significant misleading statements. First, the target denied or never disclosed that an overall reward existed for the team. Second, the target intentionally misled his team members about administrative processes. For example, the target was ambiguous about team answers. Third, the target lied or refused to reveal his real identity. Fourth, significant fabrications occurred in discussions relating to monetary rewards. Resulting comments indicated that the target's anomalous behavior was noticed by the team. The results showed that target took the bait, and behaved in a way that was *defensive* and *didactic/pedantic*. In another situation, the target, who took the bait, would get upset with his team members.

## 7. CONCLUSION

Human perception is not fully reliable due to the fact that not all information is made transparent to the perceivers. Humans attribute their perception of people's trustworthiness based on limited social interactions. Most of the attributions are context-specific, time-dependent and are combined with judgment regarding the target's capability to hold responsibility and accountability for achieving external goals. Basic struggles of personal gain, selfishness and greediness remain - not only in physical environment - but in virtual organizations, as well as in an online community. The ethical values and moral standards are vaguely defined by the society and therefore vaguely adapted by individuals.

This theoretical framework utilizes a non-conventional social psychological approach to address this gap. While data collection for the insider threats problem is a challenge, the “Leader's Dilemma” game creates and simulates this complicated situation in an online environment. Not only front-end data concerning how a target interacts with his or her organization is generated, but this online game also captures shadow data concerning how a target is influenced. The contribution of this theory lies in its utilization of attribution theory in a basic human trust relationship within a workplace to understand a complicated organizational problem, insider threats.

The findings demonstrate hope that it is possible to trace and detect anomalous information behavior of an insider leader, although the leader's change in behavior is subtle. Nevertheless, an internal attribution needs to be measured from the target's

social network in the future work. It is believed that this framework of trustworthiness attribution can be generalized through understanding human conversational acts and logics, and which can be formalized to build a socio-technical system for insider threat prediction.

## 8. ACKNOWLEDGMENTS

I thank Jeffrey M. Stanton, my advisor, for his constant support and insight. I thank Conrad Metcalfe for his helpful comments and editing assistance.

## 9. REFERENCES

- [1] Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, 50(2), December 1991, 179-211.
- [2] Beck, L., & Ajzen, I. (1991). Predicting Dishonest Actions Using the Theory of Planned Behavior, *Journal of Research in Personality*, 25(3), September 1991, 285-301.
- [3] Hardin, R. (2003). Gaming trust. In E. Ostrom & J. Walker (Eds.), *Trust and reciprocity: Interdisciplinary lessons from experimental research* (pp. 80-101). New York: Russell Sage Foundation.
- [4] Heider, F. (1958). The psychology of interpersonal relations. New York: John Wiley & Sons.
- [5] Heider, F. (1944). Social perception and phenomenal causality. *Psychological Review*, 51, 358-374.
- [6] Ho, S. M. (2008a). Attribution-based Anomaly-Detection: Trustworthiness in an Online Community. *Social Computing, Behavioral Modeling, and Prediction*. Springer: January 2008, 129-140.
- [7] Ho, S. M. (2008b). *Towards a Deeper Understanding of Personnel Anomaly Detection*. Encyclopedia of Cyber Warfare and Cyber Terrorism, 2008 IGI Global Publications, Hershey, PA, 206-215.
- [8] Holmes, J. G., and Rempel, J. K. (1989a). "Trust in Close Relationships." In *Review of Personality and Social Psychology*, Vol. 10, ed. C. Hendrick. Beverly Hills, CA: Sage Publications.
- [9] Holmes, J. G., & Rempel, J. K. (1989b). Trust in close relationship. In C. Hendrick. (Ed.), *Close relationship* (pp. 187-220). Newbury Park: CA: Sage.
- [10] Hosmer, L. T. (1995). Trust: The Connecting Link between Organizational Theory and Philosophical Ethics. *Academy of Management Review*, 20(2), Apr., 1995, 379-403.
- [11] Keeney, M., Kowalski, E., Cappelli, D., Moore, A., Shimeall, T., and Rogers, S. (2005). "Insider Threat Study: Computer System Sabotage in Critical Infrastructure Sectors." National Threat Assessment Center, U.S. Secret Service, and CERT® Coordination Center/Software Engineering Institute, Carnegie Mellon, May 2005, pp.21-34. Obtained from <http://www.cert.org/archive/pdf/insidercross051105.pdf> on April 10, 2007.
- [12] Kelley, H.H. (1973). The Process of Causal Attribution, *American Psychologist*, Feb 1973, 107-128. Obtained from [http://faculty.babson.edu/krollag/org\\_site/soc\\_psych/kelly\\_attrib.html](http://faculty.babson.edu/krollag/org_site/soc_psych/kelly_attrib.html) on July 5th, 2007.
- [13] Martinez-Moyano, I. J., Rich, E. H., Conrad, S. H., & Andersen, D. F. (2006). *Modeling the Emergence of Insider Threat Vulnerabilities*. In Perrone, L. F., Wieland, F. P., Liu, J., Lawson, B. G., Nicol, D. M., & Fujimoto, R. M. (eds.), IEEE, Proceedings of the 2006 Winter Simulation Conference, 562-568. Obtain from [http://www.dis.anl.gov/publications/articles/Martinez-Moyano\\_et\\_al\\_2006\\_WSC.pdf](http://www.dis.anl.gov/publications/articles/Martinez-Moyano_et_al_2006_WSC.pdf) on March 25, 2008.
- [14] Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*, 84(1), Feb. 1999, 123-136.
- [15] Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709-734.
- [16] Randazzo, M. R., Keeney, M., Kowalski, E., Cappelli, D., and Moore, A. (2004). Insider Threat Study: Illicit Cyber Activity in the Banking and Finance Sector. National Threat Assessment Center, U.S. Secret Service, and CERT® Coordination Center/Software Engineering Institute, Carnegie Mellon, August 2004. Obtained from [http://www.secretservice.gov/ntac/its\\_report\\_040820.pdf](http://www.secretservice.gov/ntac/its_report_040820.pdf) n April 10, 2007.
- [17] Rempel, J.K., Holmes, J.G., & Zanba, M.D. (1985). Trust in close relationship. *Journal of Personality and Social Psychology*, Vol. 49, p. 95-112.
- [18] Richardson, R. (2007). 2007 *CSI Computer Crime and Security Survey*. Computer Security Institute.
- [19] Rotter, J.B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, 35 (4), 651-665.
- [20] Rotter, J.B. and Stein, D.K. (1971). Public Attitudes Toward the Trustworthiness, Competence, and Altruism of Twenty Selected Occupations. *Journal of Applied Social Psychology*, Dec 1971, 1(4), 334-343.
- [21] Steinke, G. D. (1975). The prediction of untrustworthy behavior and the Interpersonal Trust Scale. Unpublished doctoral dissertation, University of Connecticut, 1975.