Behavorial Parameters of Trustworthiness for Countering Insider Threats

Ho, Shuyuan Mary (1)
Organization(s): Syracuse University

This proposal is intended to examine human trustworthiness as a key component for countering insider threats in the arena of corporate personnel security. Employees with access and authority have the most potential to cause damage to that information, to organizational reputation, or to the operational stability of the organization. I am interested in studying the basic mechanisms of how to detect changes in the trustworthiness of an individual who holds a key position in an organization, by observing overt behavior – including communications behavior – over time. Rotter (1980) defines trust as a generalized expectancy - held by an individual or a group - which the communications of another (individual or group) can be relied upon. In this investigation, "trustworthiness" is defined as the degree of correspondence between communicated intentions and behavioral outcomes that are observed over time (Rotter, 1980 & 1967). The degree of correspondence between the target's words and actions remain reliable, ethical and consistent, which its degree of fluctuation does not exceed observer's expectations over time (Hardin, 1996). To be able to tell if the employee is trustworthy is thus determined by the subjective perceptions from individuals in his/her social network that have direct business functional connections, and thus the opportunity to repeatedly observe the correspondence between communications and behavior. This study adopts the concept of correlating data-centric attributions, as observed changes in behavior from human perceptions; as analogous to "sensors" on the network. The Attribution Theory is adopted in the experimental situations (the "leader's dilemma" game) to extract indirect perceptions of trustworthiness toward a critical worker over time in a group dynamics (Kelley, 1973). The principles of distinctiveness, consensus and consistancy are applied in these experimental situations.

See following pages for 1500-word abstract.

The Third Annual *i*Conference 2008

Research-in-Progress for Poster

Behavioral Parameters of Trustworthiness for Countering Insider Threats

(abstract)

Introduction

Since Robert Hanssen, a US counterintelligence agent, started spying and gave away highly classified national security documentary materials to KGB[1]/SVR[2] in Soviet Union / Russia in 1970, the case of a betrayal of trust by a trusted, high-ranked insider was established (FBI National Press Office, 2001). This case portrays that not only the trust level of a key person with high-level security clearance could be altered, but the danger s/he brings to corporate security is maximized as s/he knows what and where the critical corporate resources are. In the *Insider Threat Study by CERT (2004-2005),* US DoD[3], DHS[4], & 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 (Randazzo, et al., 2004; Keeney, et al., 2005). Thus, "insider threats" as an organizational problem gap is defined as executives or someone with authorized access, high social power and holding a critical job position, who is capable of inflicting high impact damage including psychological, managerial, or physical level, within an organization.

Conceptualization

Conceptualizing Trustworthiness

Figure 1 depicts the conceptual framework of my study into trustworthiness. It utilizes a multi-level analysis, including mixed "lenses" of organizational and individual norms.

---

[1] KGB (transliterationof "КГБ") is the Russian abbreviation for Committee for State Security (Комите́т Госуда́рственной Безопа́сности).

[2] SVR is the Russian abbreviation for Foreign Intelligence Service (Служба Внешней Разведки), which is Russia's primary external intelligence agency.

[3] US DOD stands for the US. Department of Defense.

[4] DHS stands for the Department of Homeland Security.

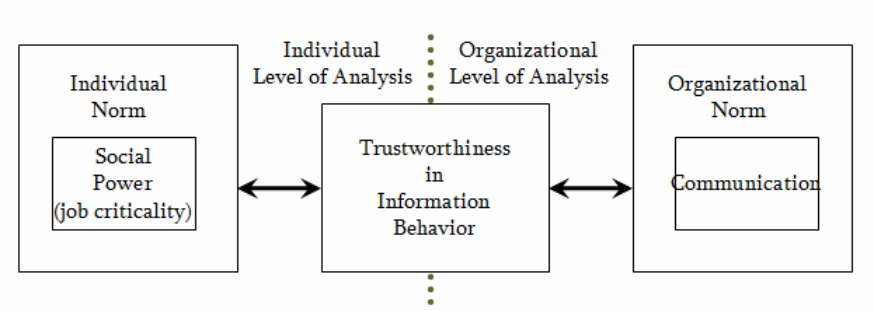

Figure 1: Conceptual Framework

Julian Rotter (1971) believes that "trust and trustworthiness are closely related." Rotter (1980) defines trust as a generalized expectancy - held by an individual or a group - that the communications of another (individual or group) can be relied upon (Rotter, 1967, p. 652) regardless of situational specificity (Rotter, 1980). Hardin (1996), on the other hand, differentiates trustworthiness from trust. He believes that trustworthiness is "a moralized account of trust" (Hardin, 1996, p. 28). For example, a criminal (A) can trust a criminal-partner (B) to conduct a joint crime and there is no moral or ethical notion involved. However, it might take much more complexity in decisions for A to let B handle A's financial accounts because A may not find B trustworthy (Hardin, 1996, p. 29). In this light, I define an employee's "trustworthiness" as the generalized expectancy, a subjective probability, toward a target's degree of correspondence between communicated intentions and behavioral outcomes that are observed and evaluated over time. In other words, the degree of correspondence between the target's words and actions remain reliable, ethical and consistent, which its degree of fluctuation does not exceed observer's expectations over time. To be able to tell if an employee is trustworthy is thus determined by the subjective perceptions of individuals in his/her social network who have direct business functional connections, and thus the opportunity to repeatedly observe correspondence of communication and behavior.

Attribution Theory Review

Attribution Theory intends to understand how people attribute (or assign) causes to another's behavior (Heider, 1944 & 1958). It's a cognitive perception. Attribution theory dichotomizes behavioral causes to both internal and external. If the causes of behavior are attributed to the person, it is called the internal attribution. Such causality of behavior is both internal and dispositional. If the causes of behavior are attributed to the situation, it is considered external attribution. The basic observational "setting" contains three major variables: the observer, the target and the situation. Perception can vary if the observations (or interpretation) are from different observers, or if the target being studied is different, or if the situation is different (Heider, 1958; Weiner, 2006, p. xvii).

Kelley (1973) suggests a causal attribution theory where an observer has multiple sources of relevant information, and is likely to infer and detect the causes of observed behavior. The observed behavior can be interpreted and perceived in a single observation – or through multiple observations over time. Kelley (1973) suggests three principles, distinctiveness, consensus and consistency, be applied in these multiple observations.

Research Question

I am interested in studying the 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 (see, for example, Steinke, 1975). Since Steinke suggests that it is possible to detect cheating behavior without directly observing the individual, my overarching question with regards to insider threat phenomenon is: *Why are the clues to a critical worker's future behavior so difficult to detect by members of a community?* Specifically, my research questions can be rephrased to three continuous sub-questions: With regards to personnel in authority positions, is it possible to detect changes in trustworthiness from subjective reflections and indirect perceptions of his/her social networks (peers, subordinates or associates)? If yes, what are the basic mechanisms of detecting changes in an individual's trustworthiness level? Would it then be possible and reliable to predict the potential for insider threats in advance using these subjective reflections?

Methodology

I intend to design experiments to answer my research questions. The perception given by observers can vary depending on the interpretations of different observers, the target, and different situational settings (Heider, 1958; Weiner, 2006). The attribution of the target's (A's) behavior by observers (B's) is believed and determined by B's judgment that A intentionally or unintentionally (Heider, 1958) behaves in a way that the cause of behavior is attributable to either external (situational) causality or internal (dispositional) causality.

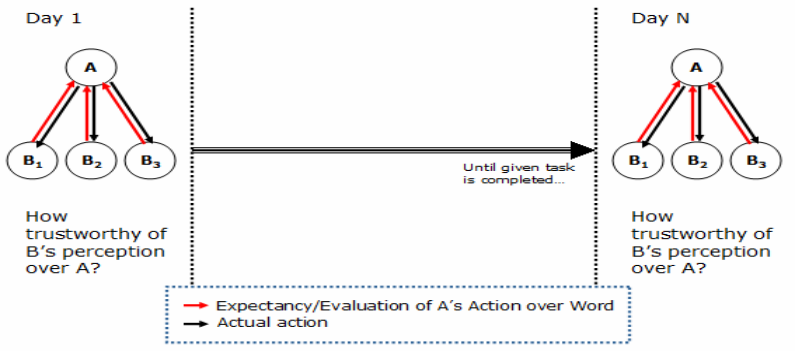

Figure 2: Illustration of Experimental Observation Over Time

The principle of distinctiveness shall be applied to the target's behavior. In another words, the target's behavioral change has to be noticeable for others to perceive it. In order to eliminate participant bias, consensus among observers should be obtained whenever possible. The consistency between the target's words and actions shall be evaluated by these observers (Kelley, 1973) over time (Day 1 through Day N) until a given task is completed (Figure 2).

Operationalization Rationale

Consistency between the target's (A's) words and actions is an important indicator to properly measure generalized expectancy of this target's social communication to and from his/her peers. The real focus of the study is to see how other people characterize the attributes of this leader during slightly suspicious behavior. Some challenging considerations and innovative ideas are utilized in the design of this experiment. 1.) Scenario encapsulation by a shielded title so that the unwanted participant acting bias can be filtered out. 2.) Manipulating the target through forcefully creating a dishonesty gap (Figure 4) between the interests of the leader and of the team players. 3.) A real-world case is simulated through feeding information to the participants in both public and private settings. 4.) The target is empowered through direct appointment. 5.) The art of "fishing dynamics" is introduced in the game scene while an authoritative figure is manipulating the target with an ethical dilemma. 6.) Team involvement is quickened through fun and competitive brain teasers. 7.) Experiment is designed to be flexible and can be repeated through "virtual asynchronous contest."

Experiment Design

This experiment design is called the "leader's dilemma" game. This experiment implements the concept of a "virtual asynchronous contest," and is designed to recruit one real team, and share fictitious scores of three other teams (Figure 3).

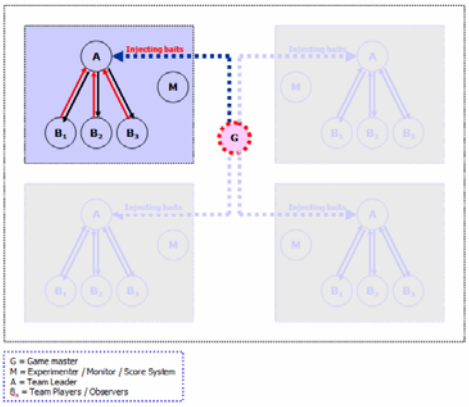

Figure 3: Controlled Room Design of Pilot Study

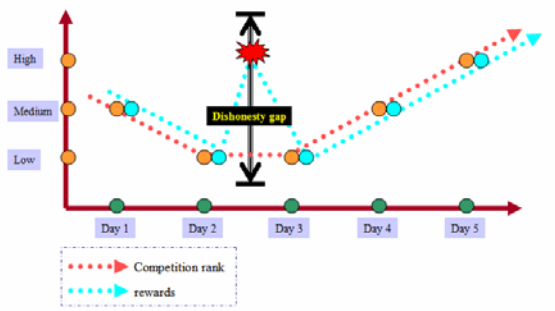

Figure 4: Logics for Virtually Controlled Contest

The bright blue area is the actual team recruited - the gray areas represent fictitious teams. I plan to recruit 4 or 5 participants each experiment. One participant will be appointed as the leader (A). The remaining participants will be team players (B_n). The experimenter (M) will

be monitoring the game play and collecting observations. M's role in this controlled environment will be that of a positive judge. The game-master (G) is in an authoritative position, but has a slightly negative role. G has knowledge of the competition and has the power to award the wining team. G's job is to inject the bait (in the form of micro-currency) to entrap A. While A is entrapped, B_n perceives A's behavior. The perceptions from B_n will be collected in various ways: 1) daily survey from B_n, 2) participant observation from M, and 3) the individual semi-structured interview at the end of the experiment of both A and B_n by the experimenter, M. This virtual contest reaches its climax when a dishonesty gap is forcefully created by feeding the bait to the team leader alone (Figure 4).

Study Contribution and Limitation

In traditional information security strategy, in order to effectively quarantine polymorphic virus codes, it has been necessary to study how codes change. Likewise, we may be able to detect suspicious behavior and counter insider threats by studying how human behavior changes and how his/her trustworthiness is altered through indirect perception of others within an organization. This study provides basic mechanisms, clues, and early warning signs to investigate and detect fluctuating personnel trustworthiness; however it is important to note that these by no means possess full assurance to convict crimes. While these basic mechanisms could be adopted in cognitive modeling of detection systems, human intervention is still necessary in this loop.

References

FBI National Press Office. (2001). *Federal Beaureu of Investigation Story: Robert Philip Hanssen Espionage Case*. Released on Feb 20, 2001. Obtained from http://www.fbi.gov/libref/historic/famcases/hanssen/hanssen.htm.

Hardin, R. (1996). Trustworthiness. *Ethics*, Vol. 107, No. 1. (Oct., 1996), pp. 26-42.

Heider, F. (1944). *Social perception and phenomenal causality*, Psychological Review, 51, 358-374.

Heider, F. (1958). The psychology of interpersonal relations. New York: John Wiley & Sons. ISBN: 0-471-36833-4.

Ho, S.M. (2008). *Towards a Deeper Understanding of Personnel Anomaly Detection*. Encyclopedia of Cyber Warfare and Cyber Terrorism, 2008 IGI Global Publications, Hershey, PA.

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.

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 on July 5th, 2007.

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. http://www.secretservice.gov/ntac/its_report_040820.pdf.

Rotter, J.B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, 35 (4), 651–665.

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.

Rotter, J.B. (1980). Interpersonal Trust, Trustworthiness, and Gullibility. *American Psychologist*, Jan 1980, 35(1), 1–7.

Steinke, G.D. (1975). The prediction of untrustworthy behavior and the Interpersonal Trust Scale. Unpublished doctoral dissertation, University of Connecticut, 1975.

Weiner, B. (2006). Social Motivation, Justice, and the Moral Emotions: An Attributional Approach. Lawrence Erlbaum Associates, inc., Mahwah, New Jersey.