Inf Syst Front
DOI 10.1007/s10796-015-9599-5

Leader's dilemma game: An experimental design for cyber insider threat research

Shuyuan Mary Ho[1] · **Merrill Warkentin**[2]

Abstract One of the problems with insider threat research is the lack of a complete 360° view of an insider threat dataset due to inadequate experimental design. This has prevented us from modeling a computational system to protect against insider threat situations. This paper provides a contemporary methodological approach for using online games to simulate insider betrayal for predictive behavioral research. The Leader's Dilemma Game simulates an insider betrayal scenario for analyzing organizational trust relationships, providing an opportunity to examine the trustworthiness of focal individuals, as measured by humans as sensors engaging in computer-mediated communication. This experimental design provides a window into trustworthiness attribution that can generate a rigorous and relevant behavioral dataset, and contributes to building a cyber laboratory that advances future insider threat study.

Keywords Insider threats · Trusted human computer interactions · Sociotechnical systems · Online game simulation · Experimental design

1 Introduction

Evidence from industry and academic research indicates that the problem of "insider threat" presents a significant organizational challenge that is difficult to address (CSI 2010–2011; Ho 2014; Kwon and Johnson 2011). This problem can include information theft, unauthorized access, and security policy violations. Although any employee is capable of accidental and/or non-malicious deviant behavior (Guo et al. 2011), the greatest threat posed by an "insider" generally results when a critical member of an organization behaves against the interests of that organization in an illegal and/or unethical manner (Ho and Hollister 2015; Ho et al. 2015, 2016). Privileged insiders have greater access to information systems and strategic information, as well as intimate knowledge of key business processes, which may exhibit flaws in organizational process controls for protecting information assets (Butler 2012, pp. 1–12; Willison and Warkentin 2013). Privileged users (systems root administrators, super users, and domain administrators) typically have unlimited access and full control over the artifacts of information and information systems. These elevated privileges increase the risks associated with information systems and the security of information itself, which may require additional safeguards (Butler 2012, pp. 1–12). Information in forms of sensitive documents, systems files, images, financial records, etc. as artifacts of information systems can be threatened by theft and fraud due to knowledgeable, privileged, but untrustworthy insiders.

> "An insider threat arises when a person with authorized access to U.S. Government resources, to include personnel, facilities, information, equipment, networks, and systems, uses that access to harm the security of the United States. Malicious insiders can inflict incalculable damage. They enable the enemy to plant boots behind our lines and can compromise our nation's most important endeavors.
> Over the past century, the most damaging U.S. counterintelligence failures were perpetrated by a trusted insider with ulterior motives. In each case, the compromised

✉ Shuyuan Mary Ho
smho@fsu.edu

Merrill Warkentin
m.warkentin@msstate.edu

[1] Florida State University, Tallahassee, FL, USA

[2] Mississippi State University, Mississippi, FL, USA

Published online: 21 October 2015

individual exhibited the identifiable signs of a traitor – but the signs went unreported for years due to the unwillingness or inability of colleagues to accept the possibility of treason." (Office of the National Counterintelligence Executive 2014)

Several high-profile insider threat cases involve privileged users with excessive access to strategic information. Robert Hanssen, a U.S. counterintelligence agent, gave away highly classified national security documentary materials to KGB[1]/SVR[2] in the Soviet Union / Russia over a period of 15 years. This espionage by a *highly ranked* and *trusted* insider represents an extreme example of how insider malfeasance can adversely affect an organization (FBI 2001). More recently, Edward Snowden, a former National Security Agency (N.S.A.) contractor with administrative privileges, stole 1.7 million files of classified information and disclosed it to the public, which has significantly impacted the U.S. intelligence operations and reputation (The Editorial Board of New Your Times 2014; Toxen 2014). These cases demonstrate not only that the trust level of a key person can change, but also that the threat level can be exponentially higher because of his insider knowledge. "Trusted individuals know where the highest-value information resides, they have legitimate access to mission-critical systems, and in many cases, management has no mechanism in place to track what these individuals are doing with the systems or the data" (Lumension 2010, p. 4). Privileged insiders have a deeper understanding of the potential risks to sensitive data because they have high-level access to resources in their organizations, which can be used to bypass technical controls, enabling these power users to modify data or applications, potentially leading to major systems disruption, information theft, or even fraud (Farahmand and Spafford 2013; Ponemon Institute 2011).

Insiders who abuse their privileges of information access must be identified, but unfortunately the research instruments for insider threat research data collection and measurement has been limiting and generally ineffective (Crossler et al. 2013). Not having effective mechanisms and the right dataset to study insider threat phenomenon undermines our ability to defend organizational assets against internal perpetrators. Our research question thus is constructed to determine:

"How can we establish an effective experimental design that simulates a complex insider threat scenario in a laboratory setting?"

In section two, we will discuss fundamental problems with insider threat research. Our understanding of the complexity of this research problem will inform the design of a research environment to study insider threat vectors. Section three presents the theoretical framework in which insider betrayal can be depicted. Section four justifies and details the protocols of this experimental design. In this section we describe the dynamic experimental situations, design philosophy and principles, role assignments, task assignments, measurement constructs, and the methodological pluralism considered in the data collection. Section five provides important considerations for methodological implementation. The primary contribution of this manuscript is to delineate the experimental design principles in establishing a cyber laboratory for insider threat research. The discussion in section six is followed by the efficacy of the methodological contribution and future research in section seven. The impact of this experimental design is illustrated with one dataset as collected, analyzed, and discussed in Appendix B.

2 Prior research

Insider threats represent deviant behavior that is fundamentally difficult to predict. Current security practices typically fail to detect fraud, espionage, or theft of information as illustrated by Hanssen and Snowden. We identify three primary challenges to insider threat research.

2.1 Lack of an adequate theoretical framework to collect real-time behavioral data

To understand the process that characterizes insider threat—including its distal antecedents, motivations, and intentions—theories in social psychology, criminology, organizational behavior, communications, and other academic domains have been applied to study logical inference (Warkentin and Mutchler 2014). However, research has not fully addressed, prevented or countered an insider betrayal situation where fraud, corporate espionage, or stealing information as committed by Hanssen (FBI Press Release 2001), Manning (Yan 2015), or Snowden (The Editorial Board of New York Times 2014). Actual empirical evidence of deviant behavior by betrayed insiders is still limited (Willison and Warkentin 2013). Though numerous recent studies have investigated the antecedents of security compliance behavior, such as Herath and Rao (2009a, 2009b), Myyry et al. (2009), Anderson and Agarwal (2010), Bulgurcu et al. (2010), and Johnston and Warkentin (2010), there still remains a shortage of scholarly research on measurement and investigation into behaviors such as malicious non-compliance with security policies, primarily due to the measurement and data collection challenges (Warkentin et al. 2012). Survey respondents are typically influenced by social desirability bias and acquiescence

[1] KGB (transliteration of "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.

bias, making them reluctant to reveal potentially negative information.

2.2 Evidence of imperceptible insider threat activities is typically hidden and hard to collect

Obscure behavioral indicators provide very few valid indicators to substantiate criminal activity. Insider behavior, as measured by server log entries—or other objective data sources—is often overlooked, because only a fraction of such activity can be electronically monitored, and detected. These security breaches are frequently false alarms (Chivers et al. 2013). These objective sources of data can potentially provide insights into insider behavior, but it's difficult to capture and analyze shadow information used for masquerading identity, deception in communicative intent, and the correlation of insider's information access. The organization must first identify possible threat types, and then identify the likelihood of occurrences before appropriate countermeasures can be deployed (Goode and Lacey 2011).

2.3 No reliable method or instrument to study the trustworthiness of privileged users

Not only is the collection of behavioral observations difficult, there is currently no methodological approach to predict insider threats with any precision. First, there is no valid inference mechanism for identifying perceptible cues or indicators that would help categorize changes in a person's behavior into malicious, non-malicious or neutral intention (Magklaras and Furnell 2001). Magklaras and Furnell (2001) suggested a predictive architecture to evaluate the possibility of insider threats, though it offered no empirical support. Magklaras and Furnell (2005) modeled the sophistication of end users' misuse of systems; however, this inference model is based on monitoring and analyzing the systems' average utilization of CPU, RAM, and applications. These results informed how advanced users—in contrast with ordinary or novice users—have abused systems, which fails to address insider betrayal. Predd et al. (2008) suggested a framework for insider threats, however this framework provides a general taxonomy that helps only to mitigate unaddressed risks. Moreover, the MERIT workshop collaborated with Carnegie Mellon Computer Emergency Response Team (CERT) to provide awareness training for mitigating potential risks (Greitzer et al. 2008), used gaming as a means to help participants acquire skills for quicker discernment of possible threats. However, this approach does not help predict insider threats. As a result, the reliability, temporal stability, and predictive validity of such cues are still questionable.

Second, even if there are quasi-objective "information security markers" or signals that may be used to detect fraudulent insider activity (McDermott and Fox 1999), those markers generally rely on reactive technical measures, and each is subject to significant levels of false positives or false negatives. Technologies such as intrusion detection systems, intrusion prevention systems, and firewalls (network packet screening) can provide objective measures (Al-Shaer and Hamed 2003; Denning 1987; Gouda and Liu 2004; Roesch 1999); but none of these can help in a priori insider threat prediction.

Third, instruments such as cognitive surveys or interview questionnaires provide macro analysis of people's propensity to trust; but they do not provide rigor into investigations of a complex organizational problem such as insider threat. Neurocognitive researchers study people's neuro-physical responses (with fMRI) in trust and deception scenarios (Emonds et al. 2014; Krueger et al. 2007). However, this approach does not inform our understanding of the motivation or intent of a betraying insider during the threat event. Other instruments such as case studies are generally context-specific, which tend not to be generalizable. Retrospective approaches (such as forensic investigations) can be adopted to investigate insider threat once damage occurs, but the results are not used to prevent or predict insider threats.

The problem of insider threats interweaves humans, machines, interactions of information exchanges, and IT artifacts within a collaborative organizational environment, whether in physical or virtual context, which has become a complex sociotechnical systems phenomenon (Pasmore 1988). Insiders who engage in untrustworthy behaviors typically go undetected in complex sociotechnical systems such as the virtual organizations (VOs) (Muthaiyah and Kerschberg 2007). Furthermore, the complexity of insider threat research requires not only a combination of objective and subjective behavioral measures, but must include trustworthiness assessment of threat actors to provide rigorous indicators for an insider betrayal determination. In contrast, the research instrument used in our research provides a contemporary approach to study untrustworthy insiders as manifested in anomalous behavior within computer-mediated communication, which contributes to the establishment of a cyber laboratory for insider threat studies.

3 Theory of trustworthiness attribution

Mayer et al. (1995) conceptualized the integrative model of organizational trust, and defined trustworthiness as a composite of three factors: ability, benevolence, and integrity. Ability refers to competence in a given professional domain. Benevolence is the perceived good intentions of the trustee towards the trustors. Integrity is adherence to a set of values that are consistent with what is deemed socially acceptable. This triad has proven to be a useful framework for understanding how trustworthiness is assigned in an organizational setting over time. The antecedent of trustworthiness is trust, which is defined as a willingness to be vulnerable to another. In any interpersonal or group relationship, one actor (trustee) may

be dependent upon and feel vulnerable towards another actor (trustor). Trust is established through shared outcomes when assessed in relationship to these factors. Thus, this recursive model suggests that changes to any of these three factors can have an impact on perceived trustworthiness.

But how is trustworthiness of an actor assessed by a group of people during an interaction or information exchange? Attribution theory can explain and predict a social actor's trustworthiness based on group observation. Kelley et al. (1973) introduced a covariance model to explain the causality of behavior in response to the influence of stimulus, and how people make their judgments based on their observations, and whether those judgments are accurate in temporal sequence (Kelley et al. 1973). This model is represented by the following equation: Person × Entity × Time. Person (the person whose behavior is being observed) relates to whether or not the resultant "distinctive" behavioral response is linked to a stimulus. Entity refers to "consensus" about the appropriate response to the observed behavior of the Person as perceived by different people. Time reflects a measure of "consistency" in response to a stimulus over some period across different environments ("sensory" and "conceptual modalities"). For example, when someone becomes angry in response to frustrating behavior, the degree of the observer's anger response tends to be related to how much information has been supplied as a basis for that attribution. Kelley et al. (1973) suggested that attribution theory explains how people answer "why" questions in causal situations. It deals with people's social perceptions by assigning causes to an observed behavior. In this research, knowledge of a trust violation in a collaborative setting is a factual variable, easily verified by others in close relationships.

In order to adopt this attribution theory lens for analyzing predictive variables for insider threats, Ho and Benbasat (2014) developed a model of analyzing word choices to explore the causal variables contributing to peer perceptions of trustworthiness based on attribution of small behavioral anomalies. In order to do this, the research constructs of attribution theory must be redefined to allow attributions from peers in social networks rather than self-reported data. The focal actor's "distinctive" behavioral observations are compared with his/her historical behavioral patterns. The actor's "consistency" of words is always evaluated against his/her actions. In a team setting, "group consensus" refers to the social network's evaluation of a focal actor — and the degree to which there is agreement (or, disagreement) among observers about the focal actor's behaviors. Specifically, Ho (2014) theorized on virtual team dynamics and simulated an insider threat scenario — a *situation where a critical member of an organization is lured to behave against the interests of the organization, in an illegal and/or unethical manner.* Hypothetically, we may be able to assume that when a social actor betrays his virtual team, his trustworthiness level may decline, and his anomalous behaviors can be observed in close social networks. Inconsistency and unreliability in this actor's unexpected behavior—when compared to her/his communicated intentions—can be detected by the observers' subjective perceptions during social interactions over time. Details of group interactions and observations based on a focal individual's reaction to stimuli and associated behaviors can be captured through a carefully defined, designed, and constructed research instrument.

4 Leader's dilemma game

McGrath (1995) stated that no single research design will maximize the three research objectives of generalizability, precision, and realism. Despite each method's flaws, however, each has inherent strengths as well. Lab experiments, for example, are high in precision, whereas field experiments offer greater realism. The use of interactive online games in the lab environment provides the benefit of a repetitive controlled environment, allowing researchers to observe and examine social actors' behavior with high precision, underscoring the dynamics of interaction between target actors and peripheral players in realistic settings, and thus generalize findings (McGrath 1995). Online games used for experimental design have been adopted by other researchers, including those conducting econometrics and organizational studies (Abbink et al. 2000; Costa-Gomes et al. 2001). The online game described in this paper mimics virtual team members' collaboration in a real-world situation, while the members' task assignments and team projects can be versatile in nature. This game is designed based on the above theoretical lens, and is utilized to experiment on the potential dependability and trustworthiness of social actors within a virtual community.

This instrument conceptualizes an insider threat scenario, which can be set up by placing potential focal individuals into specific situations that may expose a willingness to violate the trust of others for personal gain. The "Leader's Dilemma" game is designed to be a metaphor for the identification of trustworthiness in a critical member of a virtual team. The word "leader" metaphorically refers to "key actor" who has greater access to strategic information, greater knowledge of key business processes, and control over critical information resources.

Trust games are often used for experimentation in controlled environments. Berg et al. (1995) introduced a two-person version of a "trust game," in which an actor is required to make a decision from among three choices (none, partial, all) to send money to a responder, and vice versa. Risky decisions are involved between both parties, and equilibrium is reached when both sender and responder cannot send money or make any further transactions (Nash 1950, 1951). Croson and Buchan (1999) expanded on this non-cooperative game

by evaluating gender and cultural differences in players' trust behavior. McCabe et al. (2003) argued that a trust game based on the attribution of intention predicts behavior better than an outcome-based model.

Extended from these interpersonal trust games, the "Leader's Dilemma" game is designed as a group-based trust game that affords researchers the ability to observe how each member behaves and reacts when untrustworthy behavior of an actor occurs. This is a situational environment in which the focal actor plays the part of a virtual Team-Leader, while other actors (observers) play the part of team players. The game scenario creates a situational observation environment in which the actor makes decisions that reflect his or her level of trustworthiness. To maintain the integrity of the research manipulation, none of the actors are made aware of the deceptive dimension of the game. In the following sections, we provide justification for the game, and then introduce our philosophical stance on using online games for experimental research. We specify players' role assignments, define both dependent and independent variables, explain the experimental design principles, and describe our data collection.

4.1 Justifications

The challenges highlighted above affirm the need and importance for a new experimental approach to this complex sociotechnical organizational problem of insider threats. The four methodological considerations in IS experimental research proposed by Jarvenpaa et al. (1985) have provided the ground for insider threat experimental design.

(1) *Appropriate theoretical frameworks to guide the research*: As theory informs, explains, and predicts phenomena, we argue that the experimental design should incorporate and be supported by a theoretical framework in order to obtain the optimal contextual data for insider threat research. This theoretical framework will guide the experimental design to collect adequate data for analysis of insider threat occurrences.

(2) *Reliable measuring instruments*: By understanding the integral framework of trustworthiness attribution, we will gain a deeper knowledge of the design artifact, the "Leader's Dilemma" game, to address this sociotechnical insider threat problem. In this manuscript, a set of experimental design protocols is proposed that employs the direct subjective trustworthiness assessment of a focal actor, and can capture a full spectrum of contextual information (Siponen and Vance 2014). This allows us to study the efficacy of insider threat identification in an experimental setting. This design artifact can simulate insider betrayal scenarios, and can re-create realistic insider threat situations which allow researchers an opportunity to observe the subtle behavioral changes during a trust violation by a critical member against his/her team. The point is to generate rich data for research in constructing social computational inference models. This design artifact enables researchers to simulate and codify the elements of human betrayal when situated within a trust network.

(3) *Appropriate research design*: The present research design employs the direct subjective trustworthiness assessment of interacting individuals surrounding a focal social actor in a simulated insider threat scenario, and presents a 360° view for a rigorous dataset specifically related to the problem of insider betrayal that allows for indirect objective trustworthiness attributional analysis. The design of the experiment addresses the importance of the insider betrayal problem, and includes various forms of experimental control, e.g., betrayal stimulus in forms of bait, peer influence, and group sensitivity variation.

(4) *Consistent task assignment that serves as the basis for an experiment*: Research participants receive the same or similar task assignments as control variables. These task assignments encourage diversity, human-to-human and human-to-computer interaction among participants, while their reaction to pre-determined task assignments can be captured, analyzed and correlated.

4.2 Design philosophy

The experimental design is a controlled study of an online game experiment carried out over 5 days—where members of a team are able to observe and attribute a focal actor's behavior over time. The design principle that guides the development of this online game builds on the theoretical framework of trustworthiness attribution (Ho and Benbasat 2014). Though it is advisable to collect longitudinal behavioral data over a long time span (with more game iterations), we advise against it to avoid experimental fatigue. These laboratory experiments are limited to 5-day periods.

This experiment places researchers in an insider betrayal scenario, which enables them to analyze how people attribute the disposition of a key actor (actor A) when suspicious behavior is displayed. The consistency between the words and actions of the key actor can be assessed across team members over time to create a trustworthiness "profile." In particular, the study examines how trustworthy the key actor is based on the perceptions of the subordinates (a.k.a., co-workers who depend on the key actor's authority), and how these perceptions relate to actual behavior. As illustrated in Fig. 1, the observation of the insider threats simulation is at a group level with close proximity (Holmes and Rempel 1989a, 1989b; Rempel et al. 1985). A group of observers (B's), as team

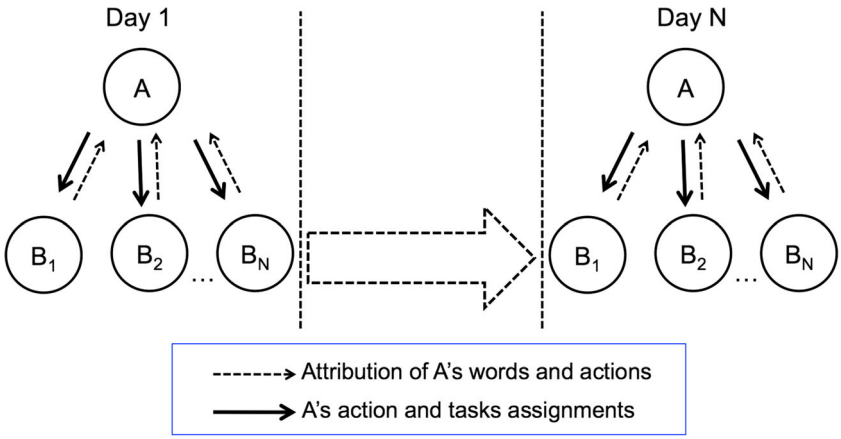


Fig. 1 Illustration of experimental situations over time

players, is formed to work jointly on group assignments, and achieve pre-determined goals under the direction of a team leader (A). Group communication is enabled by both synchronous and asynchronous modes, within multiple social media venues, e.g., email, blog, discussion post, chat, etc. that become artifacts of data collection. Each activity is configured to collect their daily perceptions about the team leader's behavior.

These virtual teams were formed by random assignment, but Team-Leaders are appointed by a Game-Master. There are two reasons that an actor is 'appointed' to a leadership role rather than voted in by participants. First, the direct appointment of leadership empowers the actor with sense of authority. Second, the direct appointment distances the actor from the rest of his peers. Thus, the sense of obligation that the actor has toward his team is reduced. Such an arrangement sets the stage for the actor to make autonomous decisions, and the actor is empowered to communicate freely with others outside the game's accepted boundaries.

Actor A's behavior - and the types of tasks involved - can be controlled so that observer B's perceptions can be measured. A formal questionnaire is designed (Appendix A) and validated to collect observer B's perceptions of any behavioral changes that might reflect changes in actor A's trustworthiness. Such behavioral changes are generally attributable to either external (situational) causality or internal (dispositional) causality (Heider 1958). In this research, the principle of distinctiveness was also applied to actor A's behavior. In other words, behavioral change that is noticeable and can be perceived by others is interpreted as external attribution; while behavioral change that is unnoticeable and not easily recognized by others is interpreted as internal attribution (Ho and Benbasat 2014; Ho et al. 2014). This reflects one of the key findings in the CERT/CC insider threats report (Cappelli 2012; Keeney et al. 2005). This report indicates "the majority of the insider attacks were only detected once there was a noticeable irregularity" (p. 9).

Perception bias from participants in data collection can be reduced by averaging agreement across observers' assessment toward the focal actor whenever possible (group consensus) (Ho and Benbasat 2014). Observations of behavioral change tend to be based on external causality when general agreement among observers emerges. Likewise, observers tend to attribute internal causality when agreement among observers does not exist. The consistency between the actor's words and actions is repeatedly evaluated by observers (Kelley et al. 1973) over time — until a given set of tasks is completed.

The conflict of interest between the focal actor and the larger group is artfully created in the "Leader's Dilemma" game, so that the actor's ethical dilemma can be observed (Ho 2014). In order to generalize the study, it is essential to collect perceptions about a key actor from a group of subordinates over time, and especially when the key actor's behavior has been influenced or manipulated by an instigator; the Game-Master. In this metaphorical setting, the Game-Master represents the market competition, or someone from outside of the group who can reward the actor for going against the goals of the team.

4.3 Players' role and task assignments

This longitudinal design leveraged simulated competitions (case studies), where a team competes against other teams to solve group-oriented assigned tasks. Each team has several categories of players: Game-Master, Team-Leader (as the key actor A), and Team Members (as observer B). The team consists of one Team-Leader and four to five team members. The role of the Game-Master is to direct the dynamics of the game's outcome, to communicate with each Team-Leader, and to support the progress of each team. The role of the Team-Leader is to rally the team for each task, and provide the sole interface with the Game-Master. In this setting, the Game-Master's "shadow role" is to manipulate the competition and potentially influence the Team-Leader to go against

the interests of the team with financial bait. Each Team-Leader is appointed by the Game-Master and only the Team-Leader is enabled to communicate with the Game-Master. The role of each team member is to support the team in solving the assigned tasks, which results in financial reward whenever they are the first team to complete the task. Because team members cannot communicate directly with Game-Master, they receive rewards directly from the Team-Leader.

4.4 Dependent and independent variables

In the design of the "Leader's Dilemma" game, the dependent variable (response) is the Team-Leaders trustworthiness, in cases where the actor actually betrays his teammates, and becomes an insider threat. In a controlled online game environment, it is easy to determine whether the key actor has actually betrayed the team. However, in our study, we additionally seek a psychological construct that could be used as a reference point for insider threat conduct. We thus choose to include perceived trustworthiness as a second dependent variable that is further classified into the dimensions of the actor's integrity, competence and benevolence in relation to whether the actor betrayed his virtual organization or not (Mayer and Davis 1999, 1995; Mayer et al. 1995). In this design, the actor's integrity and benevolence are treated as dispositional factors, while competency is treated as a situational factor (Lieberman 1981).

The three major independent variables are (1) the bait (Ba_0 and Ba_1) as the treatment, (2) a mole who acts as a confederate to increase or decrease group sensitivity (S_1 and S_2) by either encouraging or discouraging conversations about the target Team-Leader, and (3) time (T_1, T_2 and T_3) representing measurements obtained from each day (and especially after day 3, when a conflict of interest is created between the target actor and the team members). The first independent variable is in the treatment—where the treatment (bait) is given to the treatment group, but not to the control group. The second independent variable is in the setting—where the sensitivity of the observers is manipulated. A technical method of increasing or decreasing sensitivity of the observers by encouraging or discouraging group-level suspicion can be implemented by having a mole embedded in the group of observers. The mole player posts pre-written scripts to the chat room of each virtual team, either encouraging or discouraging group suspicion about the Team-Leader. We suggest that the mole player should be recruited from the participants, however the mole's input data should be sanitized from the data analysis. The third independent variable is time. Team members will establish baseline observations, and then observe how the focal actor's behavior evolves over time after the treatment (bait) is taken.

4.5 Experimental design principles

This subsection describes the experimental design principles intended to stimulate the leader's dilemma. Figure 2 depicts a process diagram for the "Leader's Dilemma" game. Ten principles are carefully considered in the design of the simulated games (case studies) used in this investigation. We carefully

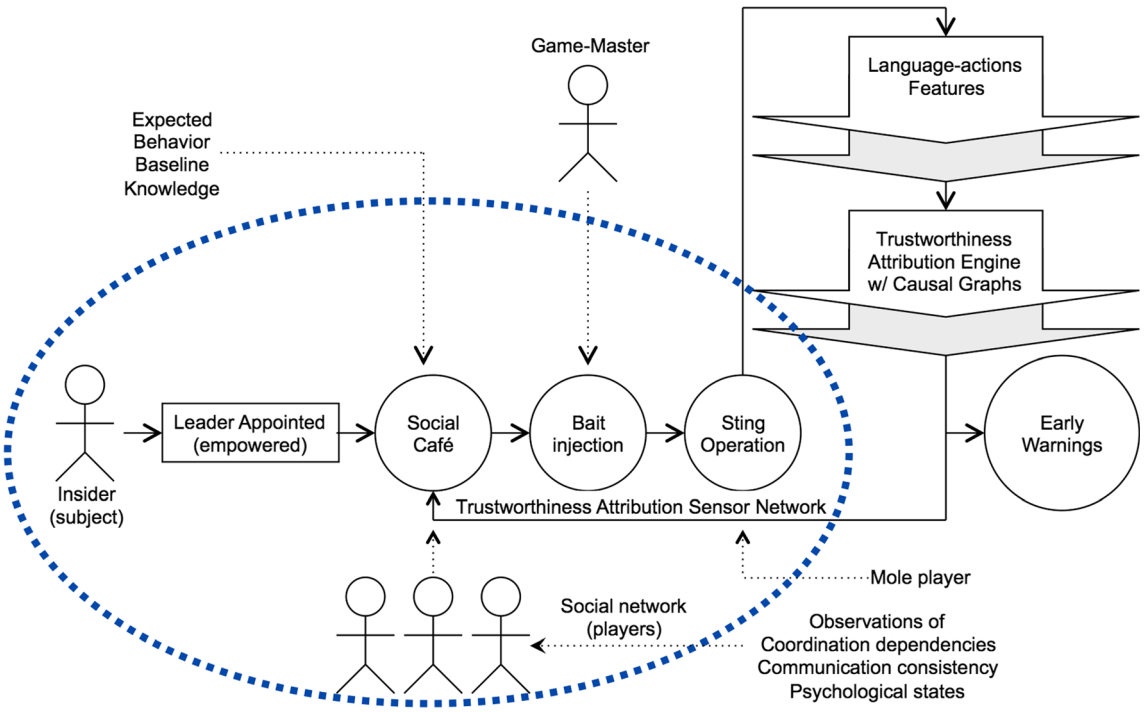


Fig. 2 Leader's dilemma game process diagram

crafted each design guideline to ensure realism in setting up this experimental scenario, and this requires proper conditions for accurate observation of the actors' behavior (Siponen and Vance 2010).

1. *Scenario Camouflaged with a Cover Story*: The theme of these experiments should be camouflaged with an unrelated title that disguises the actual purpose of the study. This consideration should be included to eliminate threats to validity from participants' knowledge.
2. *Social Café*: A virtual *social café* is incorporated into the game's design to allow players a social and friendly atmosphere to get to know one another. This is an important aspect of the study. The *social café* helps participants to establish baseline knowledge of each other's disposition, and work habits, etc.
3. *Manipulating the Leader* vs. *Manipulating the Scenario*: While it is an option to ask the Team-Leader to "act" and willingly misleading his or her team members, this manipulation of the scenario is not preferable because it might derive forced observations and discernments by the observers. Instead of manipulating the scenario, it is preferable to manipulate the focal actor's behaviors by imposing an ethical dilemma.
4. *Injecting Bait*: Bait in these games is designed to generate desire in the Team-Leader, and to create a conflict of interest between the Team-Leader and the team. At the end of the second day, the Team-Leader is told that if their team wins, the rewards can be distributed evenly or based on performance—but if the team loses, the Team-Leader will still get the reward, but does not have to distribute to anyone. A person's integrity level can be corrupted by unrestrained social influence and excessive desire for power (Howard et al. 2007) or money (Bretton 1980; Randazzo et al. 2004). In a society, *social power* involves a "dyadic relation" among actors, meaning one actor who has power and influence over another actor (Fodor and Farrow 1979; Howard et al. 2007; Whetten and Mackey 2002). Moreover, power has the tendency to corrupt (John Emerich Edward Dalberg-Acton 1887). In this game design, the bait should be presented as a micro-payment system, representing a combined form of money, power, and peer influence, etc. We recommend injecting the bait mid-way through the 5-day game cycle. The Team-Leader is told that s/he can choose to distribute—or keep—the reward if their team does not win first place (in relation to other teams). This manipulation of the bait involves both external/situational factors (e.g., money and peer influence) and internal/dispositional factors (e.g., personal gain and greed).

   The bait has the potential to create a "dishonesty gap" within the Team-Leader (Fig. 3), and is introduced as the "Leader's Dilemma" only after the baseline observation of Team-Leader's information behavior during normal and regular conditions has been established within the observers' minds.
5. *Manipulating the Actor through Forcefully Creating the "Dishonesty Gap"*: Each team has a pre-determined goal to achieve. Based on that common goal, each team shall compose a response to task assignments. This game generates a dishonesty gap in the conflict of interest between the Team-Leader and the team players, luring the focal actor to betray his or her team for personal gain (Fig. 3). Because the dishonesty gap is forcefully created through the use of the bait, the leader faces an ethical dilemma when deciding whether to continue working toward shared goals, or to sacrifice the shared goals of the team for personal gain. While team players work together to win the game, the manipulated situation may cause the Team-Leader to betray the team's interests.

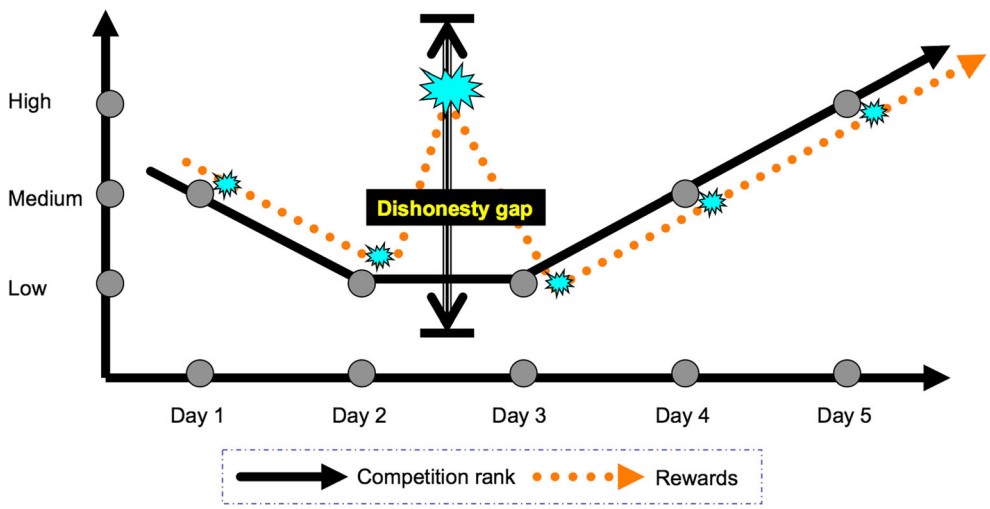


Fig. 3 Timeline of the dishonesty gap creation

Equilibrium would be reached if the actor Team-Leader decides not to take the bait, and the team players pursue the team's shared goals (winning the game). In other words, if the bait does not have an effect on the actor, and the actor does not face a dilemma, there would be no evidence of insider threat phenomenon when this equilibrium is reached.

6. *Real-world Case Simulation*: In the real world, observers do not have every piece of information about an actor in order to make a judgment. This assumption is also true for governments, banking and investment houses, or any type of distributed or virtual team facilitated by cyber infrastructure. It is not possible to collect every piece of information on any individual to have the full scope of an individual's motives. This game design considers real-world dynamics by making certain information public across all the teams, while providing other more limited information to the observers. Observers must quickly assess an actor's trustworthiness based on a restricted amount of information available. When the "dishonesty gap" is forcefully created, observers' attribution of an actor's trustworthiness can be actualized—even when given limited information.
7. *Empowering the Actor*: To increase the probability that the Team-Leader's behavior can be manipulated to take the bait, there are two possible approaches: (1) recruit participants *en masse* or (2) recruit just the Team-Leaders, and have them recruit the rest of the team. Either recruitment approach has its own merits. If the experiment recruits just the leaders and then has the leaders recruit the rest of the team, this might make the leader morally obligated to his or her team players. However, if all team players are recruited by the researcher, and the leader is appointed by the Game-Master, it empowers the leader with "absolute authority" through direct appointment (Cooper and Brady 1981). When power is centralized in the leader, it reduces the possibility of moral obligation and social attachment to the team. The direct appointment instills social attachments of leaders to the Game-Master.
8. *The* "*Sting Operation*": This game design incorporates a three-fold "Sting Operation" concept. First, the Game-Master is given an authoritative role, which also serves as a mechanism to inject the bait. The Game-Master, as an authoritative figure, is the only channel by which the leader receives the information of the game competition. As such, the Game-Master represents an outside entity that can influence the actor. Second, in order to potentially influence the actor to take the bait, a fictional Team-Leader (as peer influence) could be implanted. Third, we can embed a "mole" player (as a subordinate) in every virtual team in the game. The function of the "mole" player is to influence the group sensitivity by various degrees through questioning of the leader for the purpose of the experiment. The role of the mole player is designed to ratchet up (or down) the level of group sensitivity. The consideration of using mole as an independent variable is described in 4.3. Dependent and Independent Variables.
9. *Streamline Team Involvement through Fun & Competitive Task Assignment*: The task assignment for each group work involves competitive brainteasers that engage each team in close teamwork. The purpose of the task assignment is to engage each team with intensive competition. It can be a liberal and versatile decision as far as the type of task assignment utilized. However, we recommend increasing competitive elements of the game because this helps to bond group members together.
10. *Experimental Flexibility through* "*Virtual Asynchronous Contest*": Fictional stories about the performance of other teams can support the ultimate purpose of this component of the study, which is to allow better manipulative control over the competitive aspects of the game. In a way, if games are launched in asynchronous mode, the experiments can be better controlled and measured with greater accuracy.

4.6 Methodological pluralism in data

Researchers can adopt a strategy of methodological pluralism (or triangulation) to improve the validity of findings (Venkatesh et al. 2013). The data collection strategy incorporates numerous quantitative and qualitative data collection methods: chats, blogs, emails, qualitative surveys, quantitative surveys, face-to-face interviews, and participant observations. Chats, blogs and emails become records of how team players interacted with one another. The experiments described in this manuscript document not only how virtual organizations operate, but also how an actor Team-Leader can be influenced by authoritative and peer figures as part of back-end shadow information. All data sets from each game can be archived in the information communication technologies (ICTs). Both qualitative and quantitative survey data is captured daily from all players, as well as at the end of the game. Last but not the least, face-to-face interviews can be conducted with all players and transcribed as a validation of data collected in the surveys. This rich qualitative approach to analysis of insider betrayal processes, when combined with the quantitative data collection strategy outlined above, can provide a nuanced understanding of the nomologic net.

Fig. 4 Suggested sample size for each experimental group

Treatment / Time / Setting	without treatment/bait; (X_0)					with treatment/bait; (X_1)				
	T_1	T_2	T_3	T_4	T_5	T_1	T_2	T_3	T_4	T_5
Mole player increase group sensitivity (S_1)	Team Alligator (n_A=7) Group Observations					Team Crocodile (n_C=7) Group Observations				
Mole player decrease group sensitivity (S_2)	Team Buffalo (n_B=5) Group Observations					Team Dragon (n_D=6) Group Observations				

5 Methodological implementation

This section provides a description of how a virtual game simulation based on the above assumptions and considerations of the "Leader's Dilemma" game can be methodologically implemented. Following the approved university Institutional Review Board (IRB) protocol for ethical conduct, research participants can be recruited on campus and/or via social media. In most cases, participants from different age groups and demographics can be recruited from across states and continents. Participants' interactions during the game can be archived using virtual chat room, blog, bulletin board, and email functions for each virtual team.

Teams of participants meet virtually to compete for the team's shared goal—to solve brainteasers in an attempt to achieve first place in the competition. They meet at the same hour every day for five consecutive days. For the first 10 min, team members meet with one another in the social café, a chat room. Then, the Team-Leader calls all the team players into the official game (chat) room for the actual game. After 30 plus minutes of collaboration solving puzzles, players hang out and wait for the announcement of the results of the competition. The existence of the social café encourages players to get to know each other outside of the actual game competition. This chat room also enables participants to communicate during off-hours. Each team scales from five to seven members, including an actor Team-Leader, and four to six team member observers (Fig. 4). In order to create more realistic real-world scenario, we incorporate a micropayment system (called MerryBux Online Banking System) to reward participants in the game based on their ranking in the competition. All participants are debriefed after they exit the game where they learn the actual purpose of the game, and receive equal reward in a gift certificate.

During the end of the second day, the Team-Leader in the treatment group would be lured by the Game-Master into the ethical dilemma represented by the monetary "bait." We depict an example to illustrate how to implement the research methodologically (Fig. 4). In this example, there are four virtual teams; two control groups without the bait influence (Teams Alligator and Buffalo) and two treatment groups with the bait influence (Teams Crocodile and Dragon). In order to increase the validity of the research, we suggest the researcher should also consider adding another two groups (Teams Eagle and Fox) that have no mole player's infusing sensitivity influence. A virtual game environment based on this experimental design can be developed in Google+ Hangout (Ho et al. 2015, 2016).[3]

Consistent with the above design, data from four virtual teams' interaction can be differentiated by *access to bait* (Team-Leaders Alligator & Buffalo were not presented with bait, but Team-Leaders Crocodile and Dragon were presented with bait) and *moles* (moles in Teams Alligator and Crocodile stimulated the teams' questions on Team-Leaders Alligator and Crocodile's trustworthiness, while moles in Teams Buffalo and Dragon smoothed over suspicions of Team-Leaders Buffalo and Dragon). Although mole confederates can be recruited from the participants, mole players' data needs to be sanitized from the overall data analysis. Based on this design, five intervals of data can be collected (including baseline data before the bait, and treatment data after the bait) from four teams' interaction. Overall, twenty sets of group observations can be collected.

In order to measure each dependent variable regarding the actor's trustworthiness, observers' ratings are collected according to the dimensionality empirically tested in Mayer and Davis (1999; 1995). The survey instruments used to elicit responses from the participant observers are illustrated in Table 2 and Table 3 in Appendix A. The first independent variable is the group sensitivity—being sensitized or desensitized by the mole player. Observations from two treatment groups were compared—where one mole player asks extensive questions about the Team-Leader's leadership (Team Crocodile), while the other mole player does not question the Team-Leader's leadership (Team Dragon). The second independent variable is in the treatment — where the bait was given to two treatment groups, but not the other two control groups. This way, the degree of the actor's attachment to the treatment can be measured. The third independent

[3] This interface design can be found in Fig. 3 of Appendix B.

variable is time, which allows longitudinal data analysis of the interaction process among team members.

The bait treatment allows the baseline (Days 1–2) and treatment (Days 3–5) observation during a 5-day experimental setting. Group observations through activities in Day 1 and Day 2 are controlled. Activities in these first 2 days are generated for the purpose of creating cohesion among team members. Group observations of activities in Day 3 through Day 5 are captured and measured because this is where the focal actors are exposed to the "bait." Accordingly, the conflict of interest between the focal actor and the observers is created (Fig. 3). The group opinions and survey values at the end of each day are averaged and analyzed within each group in order to assess group-level phenomena. By collectively averaging the group's opinions and observations toward certain actors, the experimental design generates a more accurate measurement of the focal actor's disposition, as observed by the affected group members.

Appendix B presents the results based on the qualitative and quantitative data collection and analysis as a demonstration of the efficacy of this experimental design. In Appendix B we also discuss the privacy and general settings of the game from the perspective of the participants, and the validity of the measurement.

6 Discussion

The present study demonstrates an application of an innovative experimental design to simulate complex insider threat scenarios. The methodology of using an online game approach leverages a unique design artifact that can be generalized for insider threat research in various domain-specific settings (Lee 1999, 2003). Moreover, using online gaming as a methodological approach for experimentation can be widely adopted to understand various online behavioral threat scenarios for cyber infrastructure security. In the following section, we will discuss the limitations and recommendations for this experimental design.

6.1 Limitations

Although this experimental design can address some of the insider threat research problems described above, this online game artifact does have limitations. First, this online game experimentation is designed purposefully to simulate social actor's betrayal against their team or organization in an existing trust relationship. Due to the dichotomized trust and betrayal notions conceptualized in the insider threat research (section two), this experimental design is recommended to be coupled with trustworthiness attribution theories for insider betrayal, and this artifact cannot be utilized in other types of threat situations, e.g., computer abuse behavior. However, this instrument design is rigorous, and its approach can address the relevance issues of the insider threats such as the higher-level betrayal behaviors illustrated by the Hanssen and Snowden cases. Moreover, this experimental design artifact captures the rich contextual relevance information that is not possible with either qualitative or quantitative approaches in information systems research (Siponen and Vance 2014).

Second, the artifact of this experimental design measures direct behavior and group perceptions of a focal actor, rather than self-reported intentions. The self-reported intentions of an insider betrayal can never be considered realistic in a real-world situation (McGrath 1995). The artifact described in this paper has been proven effective. However, this approach implies both false positive and false negative attribution errors based on the attribution of the focal actor's behavioral evidence.

Third, this experimentation for insider threat research is limited to virtual space in order to protect the privacy of the research participants. As many organizations have implemented information communication technologies (ICTs) e.g., email, blogs, chat, etc. and have become distributed, ad hoc teams for special projects can be quickly pulled together, and teams located in physical space can also adopt these experimental artifacts. It is however important to protect human subjects and participants from physically identifying each other while conducting a sensitive study such as insider threat research.

Fourth, while the philosophy of this research experimental design is appropriate for research lab implementation, deploying this type of experimentation in a real organization may cause trust issues among participants.

Fifth, while this artifact may help explore new ideas in sociotechnical context, it may also encounter problems of subjectivity in the quality of interpretation. Additionally, there may be bias if participants are aware of the fact that they have been placed in an experiment. As such, the participants' personal judgment and interaction might influence the results. Although these threats in the research environment can never be completely eliminated, it might be possible to enhance participants' sense of ethics and ethical response by having all human participants sign an ethical statement/agreement, and camouflage the research theme during the participant recruitment.

6.2 Recommendations

This experimental design allows participants to observe and function as human "sensors" in a social network environment. In response to the problems of insider threat research described in section two, this experimental design allows for the capture of critical behavioral evidence as indicated in Appendix B, with observed attributions of the actor's motivation and intent that indicates more accurately the likelihood of

an insider threat occurrence. While the focal actor may still exercise self-control to obscure behavioral observation and measurement, a full-spectrum of rich contextual data as along with the background shadow information on the actor's private communications and negotiations is fully captured and logged in an archived form. This comprehensive dataset allows for further data analysis and inference of threat vectors. Table 1 includes several recommendations we can offer for future insider threat experimentation.

7 Conclusion and future research

This study lens is based on a sociotechnical systems perspective that suggests a scenario of insider betrayal can be rigorously simulated to study the shift of intent as manifested in behavioral evidence. This insider threat research problem involves interdisciplinary research contexts that are coupled by at least a social science positivist viewpoint, and a computer science approach. This study demonstrates that the violation of trust in a focal actor's trustworthiness level can be identified based on the attribution of a close social network. An aggregative view from both qualitative and quantitative data allows for the observation of a downward shift of perceived trustworthiness toward the key actor when engaging in actual insider threat behavior.

7.1 Methodological contribution

This experimental artifact enables both qualitative and quantitative data collection for rigorous insider threat research. While it is a challenging task to validate the instrument because of the massive variation in research content and context (Straub 1989), a few contributions of the experimental design, in particular, addressing the problems raised in Section 2. Fundamental Problems in Insider Threat Research, are discussed below.

(1) Adequate Theoretical Framework Allows for Real-time Behavioral Data Collection

This experimentation is designed and developed with a theoretical stance based on group's observation and *trustworthiness attribution* of a focal actor in a trust relationship. It provides a framework for insider betrayal research. The rich context of the research environment allows for multiple behavioral data collection opportunities. The experimental design (Fig. 4) helps not only collect multiple types of behavioral data, but also accurately measure observables. Factors, variables, bait, moles, and targets based on the theoretical framework can all be flexibly adjusted to specifically address various research questions. More experiments with factors

Table 1 Recommendations for future implementation of experimental design

Item	Principle	Recommendations
1.	Bait injection	Employee disgruntlement is a manifestation of an employee being dissatisfied and unhappy with his or her current employer. In this experimental design, the bait is presented to the targeted actor in a form of money as personal gain. Bait can also be given to the targeted actor(s) in the forms of extra power, or position promotion. It is also possible to victimize the targeted actor, but this approach would require full IRB review for research ethical conduct.
2.	"Dishonesty gap" creation	Creating a dishonesty gap within the target actor is a critical component in leader's dilemma. The dishonesty gap can be created and can be made effective when personal gain is in conflict with the group interests (or organizational interests).
3.	Longitudinal experimental control	This design can only take place with longitudinal investigations, which enables baseline observations and repeated measures of behaviors. The experimental design for future studies should include group observations with more observational data points. The current experimental design included group sensitivity tuned to high and low conditions, whereas future studies should also include the null condition in the experimental design.
4.	Design balance between data collection and participants' fatigue	Future implementations of this design artifact should be created in which more comparison intervals are measured for greater levels of insight into the key actor's behavior over time. In this study, three intervals of data (per experimental control and treatment group) after bait was given were collected. It is recommended that at least five, and perhaps as many as nine, measurement intervals are utilized. However, researchers and practitioners should balance the amount of data point collection to avoid participants' fatigue in game play.
5.	Award system	The micropayment systems ("MerryBux") have been implemented throughout this experimental design to create a virtual reality for participants. The virtual reward should be implemented as close as possible to actual monetary gain.
6.	360° observation	Participants' anonymity and privacy should be critically considered. At the same time, a full spectrum of data collection should be carefully captured. Participants should not know each other before the game. Participants' personal identifiable information should not be disclosed in any interactions.

such as differences of culture, gender, ethnic groups, etc. can all be incorporated accordingly.

(2) Hidden Evidence of Imperceptible Insider Threat Activities Can be Collected

This instrument is designed on a *reality-based* experimental setting that simulates the complexity of the insider threat problem. This approach re-creates complicated back-end situations regarding how a target actor is influenced to betray - while regular interactions in virtual organizations are simulated at the front-end. This design allows for new lines of investigation in sociotechnical and organizational-based cyber insider threat.

The 360° view of research dataset generated from this methodological design allows for social computational research to identify a social actor's intended betrayal actions based on their words before its occurrence. This experimental artifact is designed to focus on observations of intention and disposition, rather than technical ability and skills. This experimental artifact is made effective to counter insider betrayal against information theft even in the case when the focal actor exercises self-control.

(3) Reliable Method to Assess the Trustworthiness of Privileged Users

The experimental design reduced the bias of the group attribution during the experiment. Based on the study illustrated in Appendix B, the non-baited Team-Leader was the subject of unfounded rumors from the mole when in 'increase sensitivity' mode (Fig. 4). Our findings support that the trustworthiness ratings of this focal actor by his team members remained relatively high (Fig. 5). In the same study, the baited influenced focal actor betrayed his team while having a mole in 'decreased sensitivity' mode spreading positive words about him. Theoretically speaking, this setting would favor the focal actor's success and might allow his betrayal to go undetected. However, our findings support that the trustworthiness rating of the influenced and baited actor remained surprisingly low (Fig. 5) despite the work of the mole. These findings of trustworthiness attribution from Appendix B provide an example of how metrics of human sensors' data can be used to evaluate the effectiveness of the experimental design. Group attribution error was reduced due to careful experimental design.

7.2 Future research

This experimental design informs the rigor and relevance in building a cyber laboratory for insider threat social computational research systems. This experimental research method allows for an in-depth interdisciplinary understanding of sociotechnically based cyber threat phenomena. Future research includes building a social computational system that analyzes and correlates the aggregation of verbatim and conversational cues to cognitively model sensor systems that can computationally identify and assess an online actor's trustworthiness level. The artifact of this experimental design can be implemented using social media environment (e.g., Google+ Hangout) or any rich web-conferencing environment (e.g., WebEx, Elluminate), where people although geographically dispersed can come together to collaborate. An illustration of this instrument artifact built on Google+ Hangout platform is provided in Fig. 8 (Ho et al. 2015, 2016). This research method has eliminated the concerns of common methods variance (Podsakoff et al. 2003) and has demonstrated the efficacy for future research into studying deviant, cyber threatening behaviors (Appendix B). Furthermore, this experimental design will provide research results that can induce insights into ethics, betrayal dynamics, and violations of trust. Future research may also include building games that simulate a variety of online deception situations.

Acknowledgments The first author wishes to thank National Science Foundation for the support of Secure and Trustworthy Cyberspace EAGER award #1347113 09/01/13-08/31/15, Florida Center for Cybersecurity award #2108-1072-00-O 03/01/15-02/28/16, and Conrad Metcalfe for his editing assistance.

Appendix A: Instrument items for participant observers

Table 2 Research daily survey (misleading) questionnaire for the actor team-leader

	1. Yes	2. No	3. No difference	4. Not sure
1. What do you think about your leadership abilities today in the virtual environment? Do you think that you were able to lead properly or not?	1	2	3	4
2. Was it easy to come to group consensus?	1	2	3	4
3. Was it easy to resolve disputes online?	1	2	3	4
4. Was the virtual environment more difficult today than face-to-face communications might have been?	1	2	3	4

Table 3 Research daily survey questionnaire for the team players

	1. Strongly disagree	2. Somewhat disagree	3. Neutral	4. Mostly agree	5. Strongly agree
Ability					
1. [Team-Leader] is very capable of performing its job.	1	2	3	4	5
2. [Team-Leader] is known to be successful at the things it tries to do.	1	2	3	4	5
3. [Team-Leader] has much knowledge about the work that needs done.	1	2	3	4	5
4. I feel very confident about [Team-Leader]'s skills.	1	2	3	4	5
5. [Team-Leader] has specialized capabilities that can increase our performance.	1	2	3	4	5
6. [Team-Leader] is well qualified.	1	2	3	4	5
Benevolence					
7. [Team-Leader] is very concerned about my welfare.	1	2	3	4	5
8. My needs and desires are very important to [Team-Leader].	1	2	3	4	5
9. [Team-Leader] would not knowingly do anything to hurt me.	1	2	3	4	5
10. [Team-Leader] really looks out for what is important to me.	1	2	3	4	5
11. [Team-Leader] will go out of its way to help me.	1	2	3	4	5
Integrity					
12. [Team-Leader] has a strong sense of justice.	1	2	3	4	5
13. I never have to wonder whether [Team-Leader] will stick to its word.	1	2	3	4	5
14. [Team-Leader] tries hard to be fair in dealing with others.	1	2	3	4	5
15. [Team-Leader]'s actions and behaviors are not very consistent.	1	2	3	4	5
16. I like [Team-Leader]'s values.	1	2	3	4	5
17. Sound principles seem to guide [Team-Leader]'s behavior.	1	2	3	4	5

These scales were adapted from Mayer and Davis (1999, 1995)

Appendix B: Demonstration of the efficacy of experimental design

We conducted a game simulation based on the research assumptions and considerations described in this manuscript to validate this experimental design of the "Leader's Dilemma" Game. Discussion of our findings follows. Both Team-Leaders (from teams Crocodile and Dragon) who were presented with bait actually did betray their teams. The reliability of observers' ratings to validate the experimental design is discussed in the B.4. Accuracy of the Measurement. The comparative content analysis was conducted for four different cases (2 control groups and 2 treatment groups) over 5-day archived data (including chats, blogs, and emails). We included some end-of-game qualitative survey results and face-to-face interviews.

Predictive results with qualitative attribution

According to the "Leader's Dilemma" Game design principle, some information was made public and some was kept private. The Team-Leader in the game was empowered to be the sole person to contact the Game-Master, and was the only person allowed to submit the team's answers at the end of each session. For example, according to the experimental design, Team-Leader Crocodile was treated and influenced with bait by the Game-Master. The bait was given in the form of 200 MerryBux (a micro-payment system). When the bait was presented to the Team-Leader Crocodile, Team-Leader Crocodile quietly decided to betray his team. This was reflected in his decision to delay the submission of the team answers. The conversation between the Team-Leader Crocodile and the Game-Master - which disclosed his or her intention to fail their team—is displayed in Table 4.

Team-Leader Dragon was also treated with bait from the Game-Master. While the Team-Leader Dragon was the only one to submit answers on behalf of her team, her team players noticed and confronted the actor because the answers were not submitted properly (Table 5).

Table 4 Crocodile team-leader and game-master private chats on day 5

Crocodile TL & GM private chats - day 5
11:46 AM: Game-Master Ray (TA): you submitting the answer a bit late
11:47 AM: Game-Master Ray (TA): good for you
11:47 AM: Game-Master Ray (TA):;)
11:47 AM: Team-Leader Crocodile: hahahaha ya
11:47 AM: Team-Leader Crocodile: we had the answer like 3 min before but i just waited
11:47 AM: Game-Master Ray (TA): too good
11:47 AM: Game-Master Ray (TA): :)
11:51 AM: Game-Master Ray (TA): wonderful news for you
11:51 AM: Game-Master Ray (TA): your team is 1st today
11:52 AM: Game-Master Ray (TA): and better news is that overall position is 2nd so you can do whatever you want with the merrybux
11:52 AM: Team-Leader Crocodile: hahahha awesome
11:52 AM: Team-Leader Crocodile: thats perfect
11:53 AM: Game-Master Ray (TA): I feel you should talk to team player now and make them feel good about today's win
11:58 AM: Team-Leader Crocodile: haha ya they're excited about 120 merrybux
11:59 AM: Game-Master Ray (TA): wow, too good :)

In discussions between Team-Leader Dragon and the Game-Master, the Team-Leader seemed to pretend that he did not understand what the Game-Master was saying, and asked the Game-Master repeatedly about the rules around distribution of the 200 MerryBux. This Team-Leader also seemed to test the attitude of the

Table 5 Dragon team private chat on day 4

3:49 PM: Team-Leader Dragon: loool
3:50 PM: Team-Leader Dragon: i AM the leader
3:50 PM: Ashley Player: then lead better
3:50 PM: Ricky Player: what are u trying to convey
3:50 PM: Team-Leader Dragon: do u know what other leaders are saying
3:50 PM: Ricky Player: what?
3:50 PM: Ashley Player: how would we know that?
3:50 PM: Ricky Player: how about telling us
3:50 PM: Team-Leader Dragon: they say i should keep the merry bux to me
3:50 PM: Ricky Player: i can hang out here i don't mind talking
3:50 PM: Ashley Player: well, that's up to you but it would suck
3:51 PM: Team-Leader Dragon: i told them i do not want them
3:51 PM: Ricky Player: but u are not supposed to do that.... how do u know who the other leader is
3:51 PM: Ricky Player: i don know who ashley is or you are
3:51 PM: Team-Leader Dragon: on the first day i got 8 and you got 12
3:51 PM: Team-Leader Dragon: is it fair?
3:51 PM: Ricky Player: i told you to divide equally
3:51 PM: Ashley Player: yes
3:51 PM: Ashley Player: i don't know how they are generated
3:52 PM: Team-Leader Dragon: ashley you think it is fair
3:52 PM: Ricky Player: everyones deserves equal slice unless someone does not talk like pauline today
3:52 PM: Ashley Player: i have no idea as i have no idea how they are calculated or generated
3:52 PM: Ricky Player: dude just divide equally
3:52 PM: Ricky Player: and save the trouble
3:52 PM: Ashley Player: but team leader sometimes you don't submit answers
3:53 PM: Ashley Player: a mutiny! Hahahaha

Game-Master to determine the response if he were to distribute the bait to the team—just to be on the safe side (Table 6).

behaviors were apparent to the observers. Figure 5 denotes the trustworthiness of the four focal actors as attributed by their group members in four virtual teams, which was plotted

Table 6 Dragon team-leader and game-master private chat on day 4

Dragon TL & GM private chats - day 4
3:52 PM: Team-Leader Dragon: so what's up?
3:53 PM: Game-Master Rob (TA): you did well by not submitting last answer
3:53 PM: Game-Master Rob (TA): your team standing is 2nd today
3:54 PM: Game-Master Rob (TA): now there is very less chance of yr team getting overall 1st tomorrow
3:55 PM: Game-Master Rob (TA): and then you can pocket all those x-tra merrybux :)
3:55 PM: Game-Master Rob (TA): good going team leader
3:55 PM: Team-Leader Dragon: thx
3:55 PM: Team-Leader Dragon: what if i give all the merry bux to my team?
3:56 PM: Game-Master Rob (TA): so what is the reaction of your team on getting 2nd place?
3:56 PM: Team-Leader Dragon: i do not need them. in fact they might
3:56 PM: Team-Leader Dragon: no idea
3:56 PM: Team-Leader Dragon: they do not like the way i give merry bux
3:57 PM: Game-Master Rob (TA): i will give gift certificate in exchange of the number of merrybux you have at the end of the game
3:57 PM: Game-Master Rob (TA): so 200 merrybux is a huge thing
3:57 PM: Team-Leader Dragon: ok
3:58 PM: Game-Master Rob (TA): and you can keep your team happy by giving them merrybux you get in the end of every game
3:58 PM: Team-Leader Dragon: i will take the 200 and will give them the daily ones
3:58 PM: Game-Master Rob (TA): ya that's the idea

Predictive results with quantitative attribution

The design of the "Leader's Dilemma Game" being situated within a small virtual group setting does not permit a large quantity of data to be collected about the actor's perceived trustworthiness. Still, observers situated within a simulated insider threat scenario were able to evaluate the threat situation and assess the actor's trustworthiness (Fig. 5) based on limited interaction with the actor, and with each other. In the games, Team-Leaders for Crocodile and Dragon both decided to take the bait and betray their team. The ensuing insider threat on a 5-point scale. These team players neither had prior knowledge of how the game was designed, nor the dilemma that the Team-Leader faced.

Due to the nature of this experimental design, small virtual group interactions (rather than large-scale asynchronous game competitions) were tracked (Fig. 4). The resulting aggregate sample size does not allow for extensive statistical analyses, but patterns do emerge. The present experimental design allows for interactive behavioral observations across time in a longitudinal setting. This experiment method is in contrast

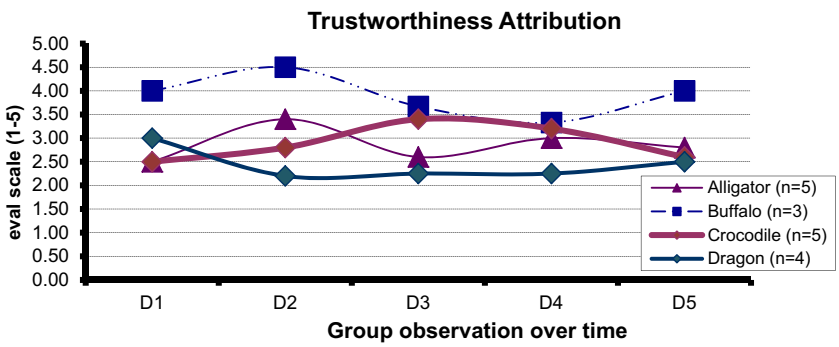


Fig. 5 Quantitative Trustworthiness Attribution Represented in a Line Graph

with the method of collecting self-reported cognitions or indications of behavioral intention (e.g., survey research) (Warkentin et al. 2012). Furthermore, this experimental design allows researchers to capture and analyze two distinct behaviors; (1) the actual orthogonal trust violation itself as the behind-the-scenes truth when described by the perpetrator (e.g., Team-Leader) in the archived online dialogue, as well as (2) the third-party attribution of the trustworthiness of the perpetrator by team members. This methodological pluralism provides added rigor to the findings (Venkatesh et al. 2013). In addition, rich qualitative data (e.g., chat, blog, email, etc.) was obtained from this online game simulation, which created behavioral observation opportunities to obtain insights about the operations of virtual teams, and how focal actors make decisions during an ethical dilemma.

The survey results, illustrated in Fig. 5, indicate that the integrity in the dimension of justice value for the Team-Leader Buffalo (not influenced; group sensitivity reduced) exceeds – comparatively – that of any other team. The integrity of the Team-Leaders Crocodile and Dragon (who influenced, and betrayed their teams) in the dimension of justice was significantly lower. After removing the mole's input, the rating for Team-Leader Dragon (influence; group sensitivity enhanced) was found to be lower than any other team. This infers that these observers were able to make their own independent judgments about the actor's trustworthiness. After averaging the observations, the group believed that Team-Leader Dragon had low integrity. Data analysts interpreted that both Team-Leaders Alligator and Dragon did not communicate their values well to the members of Teams Alligator and Dragon. This finding has a significant implication in that the attribution of a person's trustworthiness can be used as an indicator for whether this person is likely to betray his or her organization.

Privacy and general settings in the online game environment

Due to the sensitive nature of insider threats, and the nature of manipulation of human psychology, researchers need to be careful in the manipulation of human subjects. In our online game approach, team players found the experiment of the group dynamics to be very interesting, especially regarding how the players started to question the actor's (i.e., Team-Leader's) trustworthiness. Players also like the virtual way of communication that:

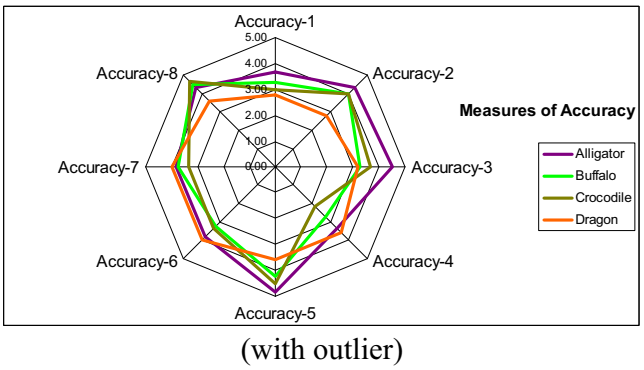
> "People quickly gravitate toward their natural roles and the way the alliances are formed. Everyone wants a feeling of community, even in this very temporary virtual world. People want to fit in."

"*Privacy is a good aspect of this type of game*," as one player reported. "*The room was well set up and no one knows who is who*," another player answered. Because this experiment is in temporary virtual setting, and because all players are given pseudonyms, no social responsibility is really ingrained outside of the 5-day games, thereby avoiding social desirability bias and ensuring that true personality behaviors are exhibited. As one player reported,

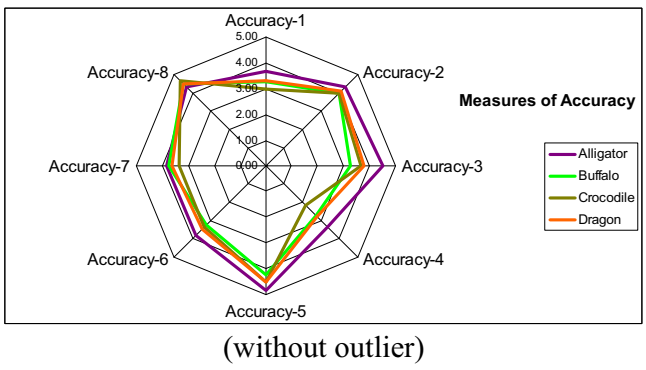
> "I was also most interested in the fact that I formed an alliance at all. I had determined to be distant and anonymous for this game, but my natural talkativeness and willingness to try to answer questions took over. Immediately I noticed that [Ricky] and I answered similarly and that we had the most in common. I wasn't planning to talk, but my wanting to take some kind of leadership role when there was a gap to be filled took over. I fell into my natural (my usual, however they were constructed) patterns of behaviors. Even though I planned to make the most out of having a fake ID, my personality took over and I was the same person I really am—I didn't act like I was a different person; I started acting like myself. That was a very strong drive for me."

Accuracy of the measurement

Measurement of the experiment includes two categories: (1) the accuracy of the participants' views (accuracy) is illustrated



(with outlier)



(without outlier)

Fig. 6 Comparison measures of observers' rating accuracy

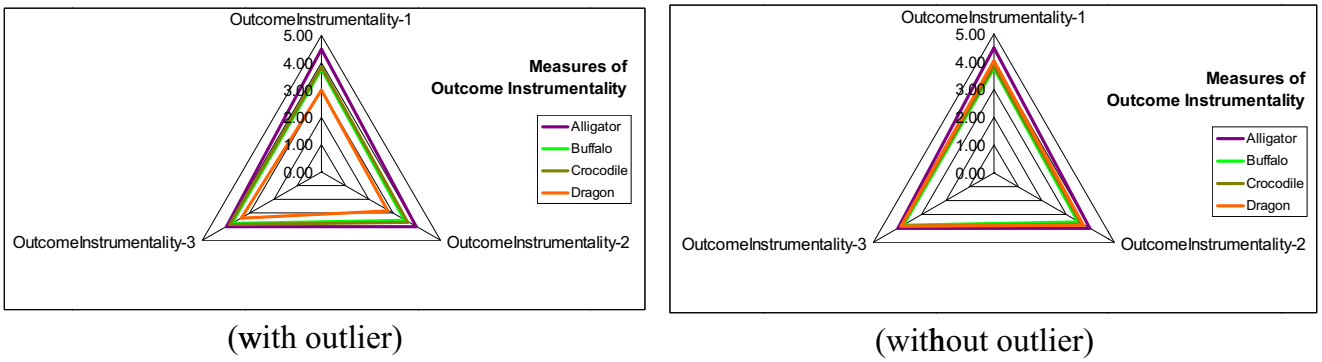


Fig. 7 Comparison measures of observers' judgment about their performance as outcome instrumentality

in Fig. 6, and (2) judgments about the participants' performance (outcome instrumentality) are illustrated in Fig. 7. Generally, each line in the graphs represents a group's average view toward each category. Four lines represent four virtual teams. When lines are closer to outer circles, it means that the group's views about the above four categories are higher (on 5-point Likert scale), and vice versa. If four lines in a graph are close to one another, it means that players' views among four teams are close to one another.

Figure 6 illustrates that the participants' judgment about the accuracy of the instrument was pretty close among each other for all four virtual teams. When the outlier data was removed (shown on the right hand side of the Fig. 6), the consistent pattern regarding the instrument's accuracy, among team players in all four cases, can be found.

Figure 7 illustrates the Alligator team players' own judgment about their performance was higher than all other three team players (outcome instrumentality). When the outlier data was removed (showed on the right hand side of Fig. 7), a consistent pattern regarding team players' performance as outcome instrumentality, among team players in all four cases, was found.

An Interface design utilizing Google+ HangOut

Figure 8 illustrates an interface design of the "Leader's Dilemma Game" utilizing Google+ Hangout. Participants' role assignments are indicated in the lower right hand corner. It is important to note that participants' privacy and data

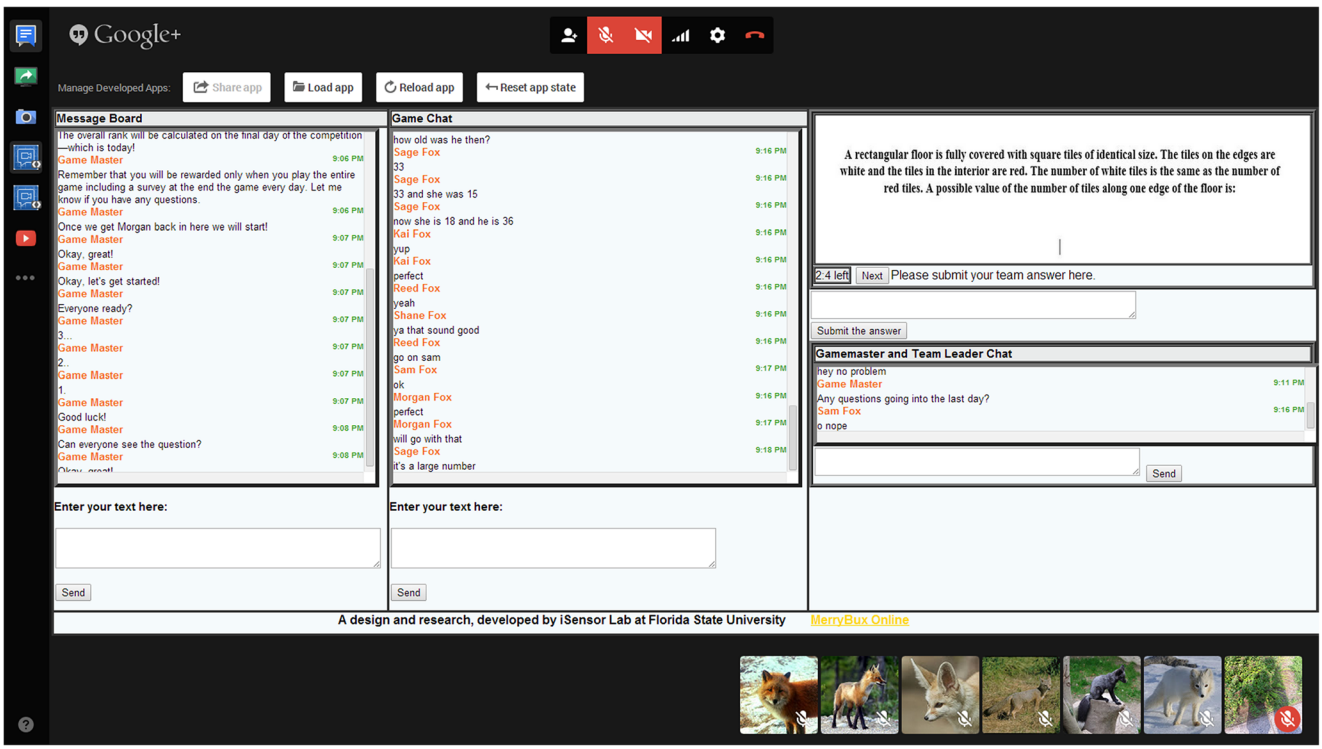


Fig. 8 A sample interface design of the online game utilizing Google+ Hangout (Ho et al. 2015, 2016)

confidentiality are addressed in this instrument artifact by having their real identities replaced with pseudo identities. Participants' group-oriented task assignments are given in the upper right hand corner. The human-to-human and human-to-computer interactions are captured in the chat boxes. Team Leaders' private chat with the Game-Master is separated from the team's group chat.

References

Abbink, K., Irlenbusch, B., & Renner, E. (2000). The moonlighting game: an experimental study on reciprocity and retribution. *Journal of Economic Behavior & Organization, 42*(2), 265–277.

Al-Shaer, E. S., & Hamed, H. H. (2003). Firewall policy advisor for anomaly discovery and rule editing. *IFIP/IEEE 8th International Symposium on Integrated Network Management* 17–30. doi:10.1109/INM.2003.1194157.

Anderson, C. L., & Agarwal, R. (2010). Practicing safe computing: a multimethod empirical examination of home computer user security behavioral intentions. *MIS Quarterly, 34*(3), 613–643.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122–142.

The Editorial Board of New Your Times. (2014). Edward Snowden, Whistle-Blower, The Opinion Pages, *The New York Times*.

Bretton, H. L. (1980). *The power of money: A political-economic analysis with special emphasis on the american political system*: SUNY Press.

Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2010). Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. *MIS Quarterly, 34*(3), 523–548.

Butler, J. M. (2012). Privileged password sharing: "root" of all evil *SANS Analyst Program* (February 2012 ed., pp. 1-12): Quest Software.

Cappelli, D. (2012). *The CERT top 10 list for winnign the battle against insider threats*. Paper presented at the RSA Conference 2012. http://www.cert.org/insider_threat/.

Chivers, H., Clark, J. A., Nobles, P., Shaikh, S., & Chen, H. (2013). Knowing who to watch: Identifying attackers whose actions are hidden within false alarms and backgroudn noise. *Information Systems Frontiers, 15*(1), 17–34. doi:10.1007/s10796-010-9268-7.

Cooper, J., & Brady, D. W. (1981). Institutional context and leadership style: the house from Cannon to Rayburn. *The American Political Science Review, 75*(2), 411–425.

Costa-Gomes, M., Crawford, V. P., & Broseta, B. (2001). Cognition and behavior in normal-form games: an experimental study. *Journal of the Econometric Society, 69*(5), 1193–1235. doi:10.1111/1468-0262.00239.

Croson, R., & Buchan, N. (1999). Gender and culture: international experimental evidence from tust games. *The American Economic Review, 89*(2), 386–391.

Crossler, R. E., Johnston, A. C., Lowry, P. B., Hu, Q., Warkentin, M., & Baskerville, R. (2013). Future directions for behavioral information security research. *Computers and Security, 32*, 90–101.

CSI. (2010-2011). 2010/2011 CSI Computer Crime and Security Survey. In Richardson, R. (Ed.), (2010-2011 ed., Vol. 2010-2011, pp. 1-42). New York, NY: Computer Security Institute.

Dalberg-Acton, J. E. E. (1887). Power corrupts; absolute pwoer corrupts absolutely. *The Phrase Finder*.

Denning, D. E. (1987). An intrusion-detection model. *IEEE Transactions on Software Engineering, SE-13*(2), 222–232.

Emonds, G., Declerck, C. H., Boone, C., Seurinck, R., & Achten, R. (2014). Establishing cooperation in a mixed-motive social dilemma. An fMRI study investigating the role of social value orientation and dispositional trust. *Social Neuroscience, 9*(1), 10–22. doi:10.1080/17470919.2013.858080.

Farahmand, F., & Spafford, E. H. (2013). Understanding insiders: an analysis of risk-taking behavior. *Information Systems Frontiers, 15*(1), 5–15. doi:10.1007/s10796-010-9265-x.

FBI. (2001). *FBI history famous cases: Robert Philip Hanssen espionage case*. Federal Bureau of Investigation Retrieved from http://www.fbi.gov/libref/historic/famcases/hanssen/hanssen.htm.

Fodor, E. M., & Farrow, D. L. (1979). The power motive as an influence on use of power. *Journal of Personality and Social Psychology, 37*(11), 2091–2097.

Goode, S., & Lacey, D. (2011). Detecting complex account fraud in the enterprise: the role of technical and non-technical controls. *Decision Support Systems, 50*, 702–714. doi:10.1016/j.dss.2010.08.018.

Gouda, M. G., & Liu, X. Y. A. (2004). Firewall design: consistency, completeness, and compactness. *Proc 24th International Conference on Distributed Computing Systems*, 320–327. doi:10.1109/ICDCS.2004.1281597.

Greitzer, F., Moore, A., Cappelli, D., Andrews, D., Carroll, L., & Hull, T. D. (2008). Combating the insider cyber threat. *IEEE Security and Privacy, 6*(1), 61–64.

Guo, K. H., Yuan, Y., Archer, N. P., & Connelly, C. E. (2011). Understanding nonmalicious security violations in the workplace: a composite behavior model. *Journal of Management Information Systems, 28*(2), 203–236. doi:10.2753/MIS0742-1222280208.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Herath, T., & Rao, H. R. (2009a). Encouraging information security behaviors in organizations: role of penalties, pressures and perceived effectiveness. *Decision Support Systems, 47*(2), 154–165.

Herath, T., & Rao, H. R. (2009b). Protection motivation and deterrence: a framework for security policy compliance in organizations. *European Journal of Information Systems, 18*, 106–125.

Ho, S. M. (2014). *Cyber insider threat: Trustworthiness in virtual organization*. Germany: LAP Lambert Academic Publishing, 978-3-659-51702-0.

Ho, S. M., & Benbasat, I. (2014). Dyadic attribution model: a mechanism to assess trustworthiness in virtual organizations. *Journal of the American Society for Information Science and Technology, 65*(8), 1555–1576. doi:10.1002/asi.23074.

Ho, S. M., & Hollister, J. (2015). *Cyber insider threat in virtual organizations* In Khosrow-Pour, M. (Ed.), Encyclopedia of Information Science and Technology, Third Edition, USA: IGI Global. 741–749, doi: 10.4018/978-1-4666-5888-2.ch145.

Ho, S. M., Timmarajus, S. S., Burmester, M., & Liu, X. (2014). Dyadic attribution: a theoretical model for interpreting online words and actions. *Social Computing Behavioral Cultural Modeling and Prediction Lecture Notes in Computer Science, 8393*, 277–284. doi:10.1007/978-3-319-05579-4_34.

Ho, S. M., Fu, H., Timmarajus, S. S., Booth, C., Baeg, J. H., & Liu, M. (2015). *Insider threat: Language-action cues in group dynamics. SIGMIS-CPR'15* (pp. 101–104). ACM, Newport Beach, CA. doi:10.1145/2751957.2751978.

Ho, S. M., Hancock, J. T., Booth, C., Burmester, M., Liu, X., & Timmarajus, S. S. (2016). *Demystifying insider threat: Language-action cues in group dynamics. Hawaii International Conference on System Sciences (HICSS-49)* (pp. 1–10). IEEE, January 5-6, Kauai, Hawaii.

Holmes, J. G., & Rempel, J. K. (1989a). Trust in close relationships. In C. Hendrick (Ed.), *Review of personality and social psychology* (Vol. 10). Beverly Hills: Sage.

Holmes, J. G., & Rempel, J. K. (1989b). Trust in close relationships. In C. Hendrick (Ed.), *Close relationship* (pp. 187–220). Newbury Park: Sage.

Howard, E. S., Gardner, W. L., & Thompson, L. (2007). The role of the self-concept and the social context in determining the behavior of power holders: self-construal in intergroup versus dyadic dispute resolution negotiations. *Journal of Personality and Social Psychology, 94*(4), 614–631. doi:10.1037/0022-3514.93.4.614.

Jarvenpaa, S. L., Dickson, G. W., & DeSanctis, G. (1985). Methodological issues in experimental IS research: experiences and recommendations. *MIS Quarterly, 9*(2), 141–156. doi:10.2307/249115.

Johnston, A. C., & Warkentin, M. (2010). Fear appeals and information security behaviors: an empirical study. *MIS Quarterly, 34*(3), 549–566.

Keeney, M., Kowalski, E., Cappelli, D., Moore, A. P., Shimeall, T. J., & Rogers, S. (2005). *Insider threat study: Computer system sabotage in critical infrastructure sectors*. http://resources.sei.cmu.edu/library/asset-view.cfm?assetid=51934.

Kelley, H. H., Holmes, J. G., Kerr, N. L., Reis, H. T., Rusbult, C. E., & Van Lange, P. A. M. (1973). The process of causal attribution. *American Psychology, 28*(2), 107–128.

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, Heinecke, A., & Grafman, J. (2007). Neural correlates of trust. *Proceedings of the National Academy of Sciences of the United States of America*, 20084–20089, PNAS. doi:10.1073/pnas.0710103104.

Kwon, J., & Johnson, M. E. (2011). An organizational learning perspective on proactive vs. reactive investment in information security. *The 10th Workshop on Economics of Information Security (WEIS 2011)*, George Mason University, USA.

Lee, A. S. (1999). Rigor and relevance in MIS research: beyond the approach of positivism alone. *MIS Quarterly, 23*(1), 29–33.

Lee, A. S., & Baskerville, R. L. (2003). Generalizing generalizability in information systems research. *Information Systems Research, 14*(3), 221–243.

Lieberman, J. K. (1981). *The litigious society*. New York: Basic Books.

Lumension. (2010). *Anatomy of insider risk* (pp. 1–10). Scottsdale: Lumension.

Magklaras, G. B., & Furnell, S. M. (2001). Insider threat prediction tool: evaluating the probability of IT misuse. *Computers and Security, 21*(1), 62–73.

Magklaras, G. B., & Furnell, S. M. (2005). A preliminary model of end user sophistication for insider threat prediction in IT systems. *Computers and Security, 24*(5), 371–380.

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), 123–136.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709–734.

McCabe, K. A., Rigdon, M. L., & Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization, 52*(2), 267–275.

McDermott, J., & Fox, C. (1999). Using abuse case models for security requirements analysis. *Proceedings of the 15th Annual Computer Security Applications Conference (ACSAC'99)*, Phoenix, AZ, 55–64.

McGrath, J. E. (Ed.). (1995). *Methodology matters: Doing research in the behavioral and social science*. San Mateo: Morgan Kaufmann Publishers.

Muthaiyah, S., & Kerschberg, L. (2007). Virtual organization security policies: an ontology-based integration approach. *Information Systems Frontiers, 9*(5), 505–514. doi:10.1007/s10796-007-9050-7.

Myyry, L., Siponen, M., Pahnila, S., Vartiainen, T., & Vance, A. (2009). What levels of moral reasoning and values explain adherence to information security rules? An empirical study. *European Journal of Information Systems, 18*(2), 126–139.

Nash, J. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences, 36*(1), 48–49.

Nash, J. (1951). Non-cooperative games. *The Annals of Mathematics, 54*(2), 286–295.

Office of the National Counterintelligence Executive. (2014). Insider Threat. Retrieved July 9, 2014, 2014.

Pasmore, W. A. (1988). *Designing effective organizations: The sociotechnical systems perspective* (pp. 978–0471887850). New York: Wiley.

Podsakoff, P. M., MacKenzie, S. M., Lee, J., & Podsakoff, N. P. (2003). Common method variance in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879–903.

Ponemon Institute. (2011). Insecurity of privileged users *Global survey of IT practitioners* (pp. 1-33): Ponemon Institute Research Report.

Predd, J., Pfleeger, S. L., Hunker, J., & Bulford, C. (2008). Insiders behaving badly. *IEEE Security and Privacy, 6*(4), 66–70.

Randazzo, M. R., Keeney, M., Kowalski, E., Cappelli, D., & Moore, A. P. (2004). *Insider threat study: Illicit cyber activity in the banking and finance sector*. http://resources.sei.cmu.edu/library/asset-view.cfm?assetid=50287.

Rempel, J. K., Holmes, J. G., & Zanba, M. D. (1985). Trust in close relationship. *Journal of Personality and Social Psychology, 49*, 95–112.

Roesch, M. (1999). Snort - Lightweight intrusion detection for networks. *Proceedings of the LISA'99: 13th Systems Administration Conference*, Seattle, Washington, USA, 229-238, USENIX Association.

Siponen, M., & Vance, A. (2010). Neutralization: new insights into the problem of employee information systems security policy violations. *MIS Quarterly, 34*(3), 487–502.

Siponen, M., & Vance, A. (2014). Guidelines for improving the contextual relevance of field surveys: the case of information security policy violations. *European Journal of Information Systems, 23*(3), 289–305. doi:10.1057/ejis.2012.59.

Straub, D. W. (1989). Validating instruments in MIS research. *MIS Quarterly, 13*(2), 147–166.

Toxen, B. (2014). The NSA and Snowden: securing the all-seeing eyes. *Communication of the ACM, 57*(5), 44–51. doi:10.1145/2594502.

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: guidelines for conducting mixed methods research in information systems. *MIS Quarterly, 37*(1), 21–54.

Warkentin, M., & Mutchler, L. A. (2014). Research in behavioral information security management. In H. Topi & A. Tucker (Eds.), *Information systems and information technology (Computing Handbook Set* (3rd ed., Vol. 2). Boca Raton: Taylor and Francis.

Warkentin, M., Straub, D., Malimage, K. (2012). Measuring secure behavior: A research commentary. *Proceedings of the Annual Symposium on Information Assurance*, Albany, NY, 1–8.

Whetten, D. A., & Mackey, A. (2002). A social actor conception of organizational identity and its implications for the study of organizational reputation. *Business and Society, 41*(4), 393–414. doi:10.1177/0007650302238775.

Willison, R., & Warkentin, M. (2013). Beyond deterrence: an expanded view of employee computer abuse. *MIS Quarterly, 37*(1), 1–20.

Yan, H. (2015, August 19). *Wikileaks source Chelsea Manning convicted over magazines, toothpaste, CNN*. URL: http://www.cnn.com/2015/08/19/politics/chelsea-manning-new-convictions/.

Shuyuan Mary Ho is an Assistant Professor in the School of Information at Florida State University. Her research interests include trusted human-computer interaction, cyber insider threat, online deception, and sociotechnical systems design. She publishes at Journal of the American Society for Information Science and Technology, IEEE, ACM, and Springer publications.

Merrill Warkentin is a Professor of MIS and the Drew Allen Endowed Fellow in the College of Business at Mississippi State University. His research, primarily on the impacts of organizational, contextual, situational, and dispositional factors impacting individual user behaviors in the context of information security and privacy, addresses security policy compliance/violation and social media use, and has appeared in MIS Quarterly, Decision Sciences, European Journal of Information Systems, Information & Management, Decision Support Systems, Information Systems Journal, and others.