Demystifying Insider Threat: Language-Action Cues in Group Dynamics

Shuyuan Mary Ho
Florida State University
smho@fsu.edu

Jeffrey T. Hancock
Stanford University
jeff.hancock@stanford.edu

Cheryl Booth
Florida State University
clb14h@my.fsu.edu

Mike Burmester
Florida State University
burmeste@cs.fsu.edu

Xiuwen Liu
Florida State University
liux@cs.fsu.edu

Shashanka S. Timmarajus
Florida State University
st13f@my.fsu.edu

Abstract

Written language as a symbolic medium of expression plays an important role in communications. In particular, written words communicated online can provide indications of an actor's behavioral intent. This paper describes an ongoing investigation into the interconnectivity between words and actions for a deceptive insider on group dynamics in virtual team collaboration. An experiment using an online game environment was conducted in 2014. Our findings support the hypothesis that language-action cues of group interactions will change significantly after an insider has been compromised. Deceptive actors tend to use more cognition, inclusivity and exclusivity words when interacting with group members. Future work will employ finely tuned complex Linguistic Inquiry and Word Count (LIWC) dictionaries to identify additional language-action cues for deception in various experimental conditions.

1. Introduction

Communicative intent in the physical world is informed by a number of physical cues. Subtle facial expressions, appearance, body language, and even pitch and tone of voice [7] all serve to provide insights into a communicator's intent. In a virtual/online environment, however, many of these cues are unavailable. Written language, such as in e-mails, instant messages/texts, blogs, e-documents etc., becomes the only available means of assessing communicative intent and providing insight into an actor's intimate "feelings," "thoughts," and "intentions" [1, p. 40]. Thus, in an online environment, the communicator's implicit and explicit communicative intent is expressed through written language [10].

This expressed communicative intent can lead to the establishment of trust when given effect with consistent action. This holds true for communications and interactions by and between two individuals, as well as communications and interactions between and among members of a virtual group. While trust is important in both interpersonal and group-based communication, Jones and Marsh [22] asserted that group trust is particularly important—even critical—to effective virtual group interactions. Group members often rely on each others' actions and information in performing certain tasks in order to achieve the specific required or desired outcomes collectively. When group trust is lost or undermined, group interaction likewise can and will become ineffective. The degree to which the loss of group trust can impact group efficacy varies by specific situation and circumstance. In some cases, the cause(s) of a loss of group trust may be superficial and relatively minor (e.g., personality clashes between individual group members), and have only a small overall impact on group trust and group efficacy (i.e. the group collectively may be able to partially compensate for these types of rifts). In other cases, however, the cause for the loss of group trust is much deeper and more significant, resulting in a much larger breach of trust within the group that may have a significant impact on group efficacy. One prime example of this is the insider threat, a situation in which a group member engages in deceptive behavior or practices—particularly when doing so in order to benefit him/ herself at the expense of the group. This paper discusses a sociotechnical study investigating this insider threat problem, simulated in an online game environment, through the lens of language-action cues present in the deceptive actors' communications between and among members of the impacted group. The study seeks to address the research question: *How do language-action cues in group interactions change when a compromised actor engages in a deceptive act*?

In terms of analytical approach, there have been studies examining the insider threat problem from psycholinguistic analysis perspectives [2, 3, 33] as well as psychosocial behavioral perspectives [8, 9]. Ho, Fu et al. [15] designed and developed an online

1530-1605/16 $31.00 © 2016 IEEE
DOI 10.1109/HICSS.2016.343

multiplayer interactive game to simulate group interactions in an insider threat betrayal scenario, which was used to collect the data reported herein. The novelty of the present study and its contribution to the field consists primarily in its specific (and direct) approach to applying linguistic analysis tools to examine the insider-threat problem. Moreover, the data collected consists of both group-level interaction data, and individual-level interaction data. This allows a side-by-side comparison of the statistical significance of group-level cues versus individual-level cues, and provides insight as to the availability of cues to deception in each type of interaction.

This paper will first discuss the general framework of this study, and then go on to describe the specific research design employed and the data collected. Thereafter, the analysis of the data will be discussed, with the results of the study presented and compared to the specified hypotheses. Finally, the conclusions drawn from the study will be presented, along with a brief discussion of possible future work.

2. Study Framework

The framework for this study draws from the literature in several areas. Specifically, it leverages existing works in trust, deception, insider threat, linguistic analysis, and group dynamics. The following section discusses each of these areas.

2.1. Trust and Action

Trust is a multi-dimensional concept, which impacts and influences interpersonal relationships on the individual level, as well as intra- and inter-group interactions and relations [18, 31]. Trust is fundamental to relationships of all kinds. In particular, trust operates to set expectations within relationships. Rotter [31] defined trust as "…a generalized expectancy…that the word, promise, oral or written statement of another individual or group can be relied on" (p. 1). That is, one individual or a group (i.e., trustor) expects that another individual or group (i.e., trustee) making the promise or statement can be relied upon to act or behave in accordance with that promise or statement regardless of the specific situation or circumstance.

While the foregoing speaks to the practical/ functional dimension of trust, there is also an ethical dimension that must be considered and is particular germane to any discussion of the breach of trust involved in insider threats. To address this dimension, Hosmer [18] proposed a different definition:

> "*Trust is the expectation by one person, group or firm of ethically justifiable behavior – that is, morally correct decisions and actions based upon ethical principles of analysis – on the part of the other person, group or firm in a joint endeavor or economic exchange*." (p. 399)

From the perspective of a trustor, trust involves assessing and taking risks [26]; that is, a trustor is willing to take a risky action or make a risky decision with uncertain outcomes after he or she calculates the risks, and has determined the trustee to be trustworthy and/or otherwise that the potential benefits outweigh the potential risks and costs in trusting a trustee. This decision and action taken by the trustor makes him or her vulnerable to the trustee. Such trust evaluation involves trustor's internal knowledge of the trustee, as well as external factors such as his or her interactions with the trustee in a particular situational context and over time.

2.2. Collective Trust in Virtual Team Context

Trust between two individuals is complex. Establishing, developing and maintaining trust between and among group members is even more so. However, as Jones and Marsh [22] so succinctly stated, group trust is critical to effective group interactions (whether within a single group, or between and among groups). It should be noted here that trust in the context of this discussion is not just a "…global feeling of warmth or affection…", but refers to the conscious decision of one actor to be dependent on another to various degrees based on circumstances [37]. Individuals in groups are expected to fulfill their respective roles and responsibilities and contribute to the group's collective work in order to achieve collective goals [25]. Implicit in this concept is that each individual's ability to fulfill his/her responsibilities is dependent upon another individual fulfilling their own responsibilities, each individual in a group or team must rely upon (i.e. trust) each other to some degree. In this group dynamic, each individual as a decision maker will constantly calculate risks and calibrate their expectations and behaviors in uncertain group dynamics. Klimoski and Karol [24] discussed the impact of interpersonal trust (i.e. the trust between and among individual members) on the group as a whole and the group's ability collectively to successfully accomplish assigned tasks. The results of their study indicate that if one member of the group has low trust in one (or more) of the others, that member will be less forthcoming in sharing information, suggestions or opinions. Particularly

where the task is problem solving by brainstorming and similar idea-sharing techniques, this illustrates how a lack of (or low) collective trust can negatively impact the group's effectiveness [19-22].

Several factors could potentially foster or undermine the development and/or maintenance of group trust—including opportunities for and extent of intra-group interaction [25], and the individual personality (communication style and behavior) characteristics of its members. While many of these are common both to groups that are physically co-located and virtual teams (i.e. a group of individuals whose members and resources may be dispersed geographically, but which still functions as a coherent unit through the use of cyber-infrastructure [14]), group trust in virtual settings presents specific and unique challenges. Of particular concern in the context of the insider threat problem is that in virtual teams, the fundamental decision to trust (or not) is largely based on a risk/reward analysis rather than shared history, experience and knowledge, and is informed by and directly dependent upon the alignment of the words and actions of the trustee with the group [23]. The trust developed within virtual teams therefore tends to be superficial, and capable of being undermined relatively easily. Moreover, when access to sensitive data is given to member(s) of a virtual team, the team becomes vulnerable to human behaviors such as deception that threaten the confidentiality and integrity of the shared information. As successful teams rely on sustainable trust relationships, the ability to understand deceptive cues in a trust-enabled computer-mediated communication will better enable virtual teams to collaborate effectively.

2.3. Insider Threat Research

Modern information systems (IS) are exposed to a variety of threats—both internal and external. While external attacks (attempted hacking, phishing, viruses and the like) are well-known and publicized phenomena, in fact, insider threats pose a significantly greater level of risk [17]. For clarity and purposes of common understanding around this concept, an *insider threat* is any situation in which a member of an organization (team, group) behaves against the interests of the organization, whether in an illegal or unethical manner [14]. As may be imagined, a malicious insider has a distinct advantage over "outsiders" in that s/he has an intimate understanding of the information assets, processes and infrastructure of the organization. In addition, s/he has legitimate and often privileged access to potentially sensitive resources and information—such as knowledge of other team members and work in progress—allowing him/her to target specific information and resources directly, without having to overcome most of the barriers. In other words, the insider is in a much better position to act—and act effectively—against an organization.

The mobility of storage, communication media, and technology enabled by distributed, grid, and cloud computing—all of which can be used to mask the source of any information "leak"—all serve to increase the complexity of the insider threat problem. Unfortunately, because protection from internal threats requires a different security approach than protection against external threats, the controls and tools used for the protection of the IS from externally initiated attacks (e.g. firewalls and intrusion detection systems) are not effective in preventing insider threats [17]. Research into the insider-threat identification problem has, accordingly, looked to elements of behavioral psychology to help not only inform the development of any technological tools that might be deployed to deter insiders from "going rogue" and intentionally compromising the organization, but more importantly, to attempt to create a behavioral/ psychological model of those individuals who are most likely to do so with a view to preemptive intervention [8, 9]. Several approaches along these lines have involved categorizing and modeling behaviors and psycholinguistic cues [2, 3, 33]. Schultz [32] defined a set of different behavioral and linguistic cues that can be used to predict insider attacks, including deliberate markers, meaningful errors, preparatory behavior, correlated use patterns, verbal behavior, and personality traits. Magklaras and Furnell [27] focused on capability and opportunity factors in proposing their model of end user sophistication, which could be used as a factor of analysis in a broader insider-threat prediction tool. Brown, Watkins et al. [3] shifted this focus a bit to look more specifically at linguistic cues. Their approach effectively "translated" observed linguistic cues into behavioral categories identified as corresponding to behaviors significantly associated deceptive insiders.

2.4. Language-action Cues in Deception

Deception in general refers to the active transmission of messages and information to create a false impression leading to a false conclusion [4]. As Buller and Burgoon [4] state, deception involves the intentional distortion of the informational content of a message by the speaker in order to convey something other than the truth. Because of the volitional and intentional nature of this action, deception as used

herein excludes the conveyance of unintentionally or accidentally incorrect or incomplete information. Cues available to detect deception differ across environment. Moreover, cues available in face-to-face communication (F2F) (such as body language, facial expression, tone of voice etc.) are not always present in a more limited CMC environment [11, 36]. Online communication is almost exclusively text-based, leaving people with only written words on which to base their trust decision. This suggests a need to look at whether *language-action cues* can reveal communicative intent. *Language-action cues* refers to the linguistic styles, phrases, language patterns, or actions in an actor's written expressions, which may manifest as an indirect or subtle signal to others [15, 16]. Indeed, Pennebaker, Mehl et al. [30] posited that words reveal the inner characteristics of an individual, and convey human emotions such as happiness, anger, anxiety and sadness. In particular among these cues, Pennebaker and King [29] suggested that linguistic styles and use of certain linguistic cues (e.g. self-references; negations; words associated with cognitive processes; and words associated with affective processes) can often reveal deceptive intent to help in detecting deception.

The proposition that *language-action cues* can be used to detect deception has been supported in a number of studies. For example, in studies examining deception in online dating profiles [13], users that were highly deceptive in their profiles included fewer self-references, less negative words and negations, more motion words, but fewer words overall when compared to less deceptive profiles [34, 35]. Zhou, Burgoon et al. [38], [40] found that deceivers tend to use more modal verbs and less self references. Likewise, in synchronous text-based CMC environments, Hancock, Curry et al. [12] found that deceivers used more words overall, more sense-based words, more other-oriented pronouns, and fewer self-oriented pronouns. In particular, they found that deceptive actors used more negations than truthful actors, and there was no difference in the frequency of causal words as between his/her truthful and false statements. However, their findings did report a difference in these two cues as between a non-incentivized deceptive actor and a deceptive actor who has been motivated to deceive.

In addition to specific words chosen by a deceptive actor, the overall descriptiveness of his/her communication and even the overall number of words used may also reveal deceptive intent. For example, studies suggest that a potential deceiver may try to gain trust and credibility to enhance the success of deception by being increasingly wordy and using peripheral expressions in their messages [39]. Indeed, although excessive descriptiveness may undermine credibility by revealing inaccuracies or inconsistencies, we may consider that sufficient description in a deceitful statement is necessary to convince another communicating actor [28]. Zhou, Burgoon et al. [38] also found that deceivers use proportionately more imagery words (sensory, spatial and temporal expressions) than truth-tellers (i.e. deceivers are said to have a higher imagery ratio). Essentially, they suggest that, because deceivers are unable to rely on experience or memory to deceive someone, they tend to use more sensory expressions (e.g., sounds, smells, physical sensations, and visual details), spatial (e.g., locations of people or objects), and temporal (e.g., time when the event happened) words. In contrast to F2F interactions, where deceivers have been found to be more laconic [5], Zhou and Zhang [40] found that in online contexts, deceivers tend to be more active, wordy, taking shorter pauses between messages and in discussion—than truth tellers [40].

Taylor, Dando et al. [33] also found several specific cues to be particularly revealing in their insider threat study using computer-mediated technologies in physical interactive context. The results of their study showed that insiders were more self-focused (using more personal pronouns, for example), and that their overall use of language associated with cognitive processing (for example, negative emotion and feelings) was different to that of their coworkers.

2.5. Promiscuous Behavior of the Motivated Deceptive Actor

As information-rich as the previous research described above is, it does not provide much insight into specifically how group members interact and react when a deceptive actor is present—and in particular when that actor has been induced or incented by an external party to act in a deceptive manner, even to the extent of undermining the interests of the group for his/ her own benefit. As mentioned above, this particular problem can be examined from both the level of individual interactions and group-level interactions. Accordingly, the hypotheses presented in this section are grouped into questions concerning individual interactions (H1(a) and (b)) and questions concerning group-level interactions (H2).

As a preliminary matter in addressing our individual-level hypotheses, it must be understood that a common characteristic of deception is the deceptive actor's attempt to conceal his/her deceptive intent in order to avoid detection. We would expect

the deceptive actor to try to sound and act much the same as if s/he were telling the truth. The more successful the deceptive actor is in this regard, the more difficult it becomes to identify or detect deceptive cues—including linguistic or language-action cues. Of course, the deceptive actor may betray him/herself through *leakage*, which Ekman and Friesen [6] described as unintentional/ unconscious or subconscious behaviors by and/or cues given off by a speaker (i.e. a deceptive actor) that belie and even undermine an intended deception. With this possibility in mind, we suggest there are observable differences in language-action cues between deceptive actors and non-deceptive actors, and that a deceptive actor will communicate and interact differently than a non-deceptive actor. Accordingly, we frame and test our hypothesis in the affirmative, as follows:

H1(a): The language-action cues of a deceptive actor's communications with his/her group will differ from those of a non-deceptive actor.

A similar analysis applies in studying the related question of whether there are observable changes in the language-action cues used by a deceptive actor after s/he has accepted an incentive for deception. Our expectation is that a deceptive actor would attempt to conceal deceptive intent by behaving the same before and after. However, we must still address leakage, and consider that a deceptive actor may or may not change the way in which s/he communicates with her/his group after receiving an incentive to deceive the group. Thus, we again frame and test our hypothesis in the affirmative, as follows:

H1(b): The language-action cues used by a deceptive actor in communicating with his/her group will be different before an incentive for deception has been introduced to him/ and after it has been accepted by him/her.

2.6. The Dynamics In Groups With A Motivated Deceptive Actor

In addition to studying the language-action cues used by an individual actor/ speaker, we also examine the language-cues present in groups—and specifically, what these may reveal concerning the presence or absence of a deceptive actor within the group. The works of Lewicki and Bunker [26] and Kramer, Brewer et al. [25] in group dynamics and group trust suggest that deceptive actors will tend to use more words designed to unite and motivate team members than other group members will.

Recently, Ho, Fu et al. [15] found that groups containing a deceptive actor tend to use more words of inclusion (e.g., "and," "with," and "include"); more words of exclusivity ("but," "without," "exclude") to create a sense of uniqueness; more words of certainty ("always," "never") to bolster group confidence; and more words of achievement (e.g., "win," "earn," and "hero") to create a sense of goal-focus and common purpose, than groups without a deceptive actor.

The studies of Lewicki and Bunker [26], and Kramer, Brewer et al. [25] further inform our understanding of the effect of the introduction of a deceptive actor into a group. Based on their work, we speculate that group interaction will be altered as a result of the introduction of a deceptive actor. For example, the group's language-action cues may indicate that group members are less willing to cooperate with the putative deceptive actor. An increased usage of cues such as cognition, certainty, inclusivity, exclusivity, and fewer affective cues may also be observed. In short, we assume that there will be a change in group communication after a deceptive actor has been introduced into the group. Accordingly, we posit that:

H2: *The language-action cues of a group's interactions will differ significantly before and after an incentive for deception has been introduced to and accepted by an actor.*

3. Research Design

In order to investigate our research question and test our specific hypotheses, we used the Google+ Hangout platform to design and develop an online multi-player/ team gaming environment, which simulates various group-interaction scenarios (Figure 1). Communication between and among participants occurs in a cue-lean, synchronous, text-based communication (CMC) environment. Specifically, the game database captures the text record of interactions and exchanges between and among players on a team, and by and between a designated focal actor in each team and an external overseer figure. In this way, we are able to examine the similarities and differences in behaviors and responses in groups with and without a deceptive actor, and before and after an incentive to deceive is accepted by a member of a team. Likewise, we can examine similarities and differences observed in the behaviors and responses of a deceptive actor both before s/he has been offered an incentive to deceive his/her team and after s/he has accepted it, and the behaviors and responses of such a motivated deceptive actor and a non-deceptive actor.

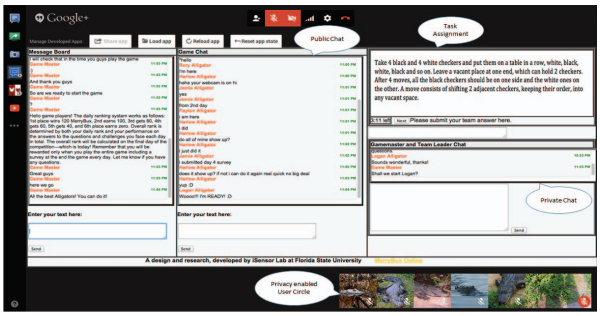

Figure 1. Illustration of an online game interface designed and developed for data collection

3.1. Game Overview

To study our research question and test our hypotheses, we developed and deployed an online game called "Collabo" (Figure 1). Participants (players) were randomly assigned to one of several virtual teams. Each virtual team was also designated by random assignment as being either a control group ("no bait") team or a treatment group ("bait") team. A pseudonym was given to each player for purposes of game interactions, in order to protect his or her information privacy.

Game play consisted essentially in each of these teams being given a series of assigned tasks that must be completed within a given timeframe. Specifically, each team was given a set of seven logic problems, and was given six minutes to solve each puzzle. Game sessions, each lasting approximately 42 minutes, were played daily over five consecutive days, with a new set of logic problems being provided for each session (i.e. each day). Although each team was presented with the same set of seven questions on the same day, in order to avoid potential confounds, the specific order in which the questions were presented was randomized by the game system across all teams. This was a competitive exercise, with each team competing against the others to be the team correctly completing the most puzzles within the specified time across the entire 5 days of play.

The game also involved an external party (i.e. a member of the research team), who served as an overseer figure (referred to as the "Game Master"). The Game Master purports to have a stake or interest in each team's outcomes in the game, and, accordingly each Team Leader was responsible for reporting to him/her on the progress of their team (i.e. they submit their team's solutions to the day's problems on the team's behalf). The Game Master had visibility to all team chats/ communications, but interacts only with the Team Leaders. A private chat function, available only to Team Leaders and the Game Master, was used for this purpose.

As noted above, play was competitive among teams. Overall team performance over the course of five days was recognized and rewarded through awards made using a micropayment system called MerryBux, with the team having the highest overall score (i.e. the team who correctly solved the most problems in the allotted time) earning the greatest amount of MerryBux. The "value" of MerryBux won during play ostensibly corresponded to the amount of the Amazon gift card they each receive at the end of the game. The Team Leader determines the exact distribution of his/her team's winnings amongst the team's members.

While the "prizes" (in MerryBux) awarded for finishing in first, second, third place etc., were known to all players, an additional financial award ("bait") was introduced secretly by the Game Master to Team Leaders in the treatment groups, by way of the private chat function mentioned previously. This incentive, which was introduced after Day 2 of play, was 200 additional MerryBux, and was offered on the following two conditions: (1) the Team Leader must keep the deal a secret from his/her team; and (2) if his/her team wins (i.e. finishes in first place), the Team Leader must split the incentive equally among team members (although, the Team Leader has no such obligation if his/her team does not win). In this way, these particular Team Leaders were presented with an ethical dilemma: whether to a) collaborate with their teammates to achieve the best team outcome and, if they win, distribute the additional MerryBux along with whatever amount of MerryBux the team won based on its performance, or b) undermine the team's collaborative efforts, either actively (as in failing to submit answers within the required time, or changing the team's answer before submission, to make sure it is incorrect) or passively (as in simply withholding any contribution to team collaboration in solving the puzzles, or allowing a response s/he knows to be incorrect to be submitted), accept less in team winnings, but keep the entire amount of the incentive for him/ herself[1].

4. Data Collection and Analysis

This experiment, with data collected during 2014, consisted of six virtual teams (three control groups, and three treatment groups) (Table 1). Each team consisted of three or four members, and a randomly

[1] The Game Master articulates that this extra incentive is approved by the research protocol as reviewed and approved by the Institutional Review Board (IRB) with protocol #15316 at Florida State University.

assigned focal actor. Participants were randomly recruited from within the student population of Florida State University. A total of 27 students participated in this study; 17 were males and 10 were females. Players' names were replaced with pseudo-names to protect privacy. Ages ranged from 18 to 65 years old.

Table 1: Control/Treatment Group Designations

Experiment Design	Control / No Bait	Treatment / Bait
Negative Feedback	Team A	Team C
Positive Feedback	Team B	Team D
Neutral Feedback	Team E	Team F

Conversation logs were archived, cleaned, processed and analyzed using Linguistic Inquiry and Word Count (LIWC) to identify relevant language-action cues (selected categories noted below). Basic linguistic and psychological categories were extracted using LIWC (Table 2). The dataset for this study consists of approximately 8,000 lines of text, and each line of chat included an average of 38.98(55.15) words.

Table 2. LIWC Categories and Examples

LIWC CATEGORIES	CODING SCHEMA	Examples
Affective Process	affect	happy, cried, abandon
Positive Emotion	posemo	love, nice, sweet
Negative Emotion	negemo	hurt, ugly, nasty
Anxiety	anx	worried, fearful, nervous
Anger	anger	hate, kill, annoyed
Sadness	sad	crying, grief, sad
Cognitive Process	cogmech	cause, know, ought
Insight	insight	think, know, consider
Causation	cause	because, effect, hence
Discrepancy	discrep	should, would, could
Certainty	certain	always, never
Inclusive	incl	and, with, include
Exclusive	excl	but, without, exclude
Achievement	achieve	earn, hero, win
Auxiliary Verbs	auxverb	am, will, have, should, would, could
Negations	negate	no, not, never

The archived data provide evidence that all the deceptive Team Leaders did distribute Team Players' daily allowance; however, none of the deceptive Team Leaders shared any amount of the additional MerryBux they received. Moreover, the data indicate that, in each case, the deceptive Team Leaders accepted the "bait". Further, in each case, the data show that these deceptive Team Leaders were (merely) passively deceptive (i.e. no attempts at sabotage or the like were evident).

From the raw transcripts, we derived two sets of clean data in order to answer our hypotheses: 1) group-level data, from the communications between the Team Leader and the team members during play, based on communications between the Team Leader and the team players while solving puzzles; and 2) individual-level data, from the private conversations and negotiations between each Team Leader and the Game Master.

We then subdivided these data sets by timeframe: data collected during days 1-2 (pre-bait) and data collected during days 3-5 (post-bait). Word count (i.e. in each LIWC category) was used as the basic unit of analysis. The data for the individual-based analysis (H1(a) and H1(b)) was converted into word counts. The data for the group-based analysis (H2) was normalized to percentages. Statistical testing—specifically, paired sample *t*-tests—was then done on the data sets using IBM SPSSv22.

5. Hypotheses Testing and Results

Our findings with respect to each of our hypotheses are discussed in the following sections.

5.1. H1: Not Supported

Hypothesis 1(a) posits that the language-action cues of a deceptive actor's communications with his/her group will differ from those of a non-deceptive actor. The results of a paired sample *t*-test run on the collected data show that that no significant differences were observed. Accordingly, H1(a) is not supported.

Hypothesis 1(b) asserts that language-action cues used by a deceptive actor in communicating with his/her group will be different before an incentive for deception has been introduced to him/ and after it has been accepted by him/her. The results of a paired sample *t*-test show that no significant differences were observed. Accordingly, H1(b) is not supported.

These findings with respect to H1(a) and H1(b) are consistent with existing literature, and are thus neither disappointing nor surprising. Essentially, with these hypotheses, we were examining how effectively a deceptive actor manages to "act natural" during a deception. As to H1(a), our findings indicate that the language-action cues present in the communications of a deceptive actor are more or less the same as are present in the communications of non-deceptive actors and that, in fact, the deceptive actor was fairly effective in concealing his/her deceptive intent. And, as to H1(b), our findings show that the deceptive actors were successful in maintaining a consistent communication style and language-action cue usage with their respective groups before and after accepting the "bait," with the result that their group

members were unaware of any change in their interactions with the deceptive actor, and the deceptive actor, again, effectively concealed his/her deceptive intent.

These results differ in important ways from prior work by Taylor, Dando et al. [33] on insider threat. In that study, the focal actor was asked to "act" as a deceptive insider, which may have enhanced the focal actor's sense of empowerment and led to more forceful actions in carrying out the deception. In contrast, in our experiment the focal actors were not aware they were being lured to act against their team. The focal actors in our experiments faced an ethical dilemma and responded, rather than being sanctioned to act deceptively. Another important difference between the present study and the study of Taylor, Dando et al. [33] is the communication setting. The settings in Taylor, Dando et al. [33] included both physical interaction and online communication, whereas the present experiment was conducted entirely online. The difference between the present findings and prior work may be a result of these changes in setting.

5.2. H2: Supported

Hypothesis 2 submits that the language-action cues of a group's interaction will differ significantly before and after a deceptive actor has accepted an incentive for deception. A comparison of the differences between a treatment groups' interactive behavior before and after the introduction of the financial incentives revealed statistical differences in language-action cues within the categories of affective process and positive emotions (Table 3).

Table 3: Paired Samples Test (H2)

		Paired Differences					
		Mean	Std. Deviation	Std. Error Mean	T	df	Sig. (2-tailed)
Pair 1	Affect-Affect_2	.40417	1.74455	.50361	.803	11	.439
Pair 2	Posemo-Posemo_2	-.14417	2.12648	.61386	-.235	11	.819
Pair 3	Negate-Negate_2	-1.13083	.97212	.28063	-4.030	11	.002

It is worth noting that in the earlier Ho, Fu et al. [15] study, data from Teams C and D were compared due to the fact that the Team E focal actor did not seem to be a real insider threat case (s/he does not appear to face the same ethical dilemma as the other two deceptive actors). The affective and positive emotions cues of the treatment groups (Teams C and D) seem to show statistical significance after their focal actor had accepted the bait. However, in our present study, we included data from all treatment groups (Teams C, D, and E) even when the focal actor in Team E did not appear to really betray his/her team—but did accept the bait. In these cases, the compromised insider tended "to mask the source of any information leak" even when they were not "actively" engaged in sabotaging their team. Here we identified a different cue; the groups containing compromised deceptive actors tend to use more negate words after the actor had accepted the bait.

In addition, the increased usage of words in the affective process and positive emotion categories suggests that players displayed more amicable behavior in their chat (Table 3). Although the language-action cues of the treatment groups are statistically significant, they may have been influenced by the group discussions when solving the puzzles. This shows that players in a group can indeed sense subtle changes in the behavior of the focal actor on their team (i.e. after an incentive was accepted).

6. Limitations

In analyzing our results—and particularly our results against H1(a) and (b)—we acknowledge that our sample size may have been too small to conclusively "rule out" the significance of individual-level interaction cues in detecting deception.

Another limitation is that all deceptive Team Leaders were passively, as opposed to actively, deceptive. Therefore, our results perforce do not distinguish or differentiate between these two categories of deceptive actors. This limitation is, however, simply a by-product of how this particular instance of the game was played, and the specific game master's approach to incenting the deceptive Team Leaders. In other instances of game-play [14], the deceptive Team Leaders were in fact actively deceptive. Accordingly, this particular limitation can be easily overcome going forward to make it more likely that both passive and active deceptive Team Leader types can be observed at the same time. One simple way in which we propose to modify future instances of data collection using this game is to provide a higher level of incentive to those Team Leaders who are seen to be actively undermining their teams' performance than those who simply fail to act.

7. Conclusions and Future Work

It is clearly a challenge to detect changes in the behavioral intent of a collaborator when s/he decides

to deceive a group of which s/he is a part. This is the crux of the insider-threat problem. The results of our study provide evidence that subtle but identifiable patterns of words used by a virtual team interacting with a deceptive group member may be an early indicator of insider threat. We submit that this is a positive step in the ongoing attempt to identify individuals who may be the source of an insider threat. Our findings support the earlier work of Ho, Fu et al. [15]; that language action cues fluctuate significantly after a deceptive actor has accepted an incentive for deception. This validates the feasibility of developing an automated computational model for detecting deception in synchronous online communications.

In this regard, we will focus future studies on identifying additional language-action cues suggestive of deceptive intent to contribute to the creation of a more finely tuned LIWC tool. Our research is also looking at ways in which these LIWC categories can be used to automatically detect deceptive intent using different machine learning approaches to identify significant cues to deception-based language-action patterns. As this line of inquiry proceeds, it may be possible to employ these in some iteration to the identification of potential insider-threat risks.

8. Acknowledgements

The authors wish to thank the National Science Foundation EAGER grants #1347113 and #1347120, 09/01/13—08/31/15, the Florida Center for Cybersecurity Collaborative Seed Grant 03/01/15—02/28/16, and the Florida State University Council for Research and Creativity Planning Grant #034138, 12/01/13—12/12/14.

9. References

[1] Austin, J.L., ***How to do things with words***, 2nd ed. 1962. Cambridge, MA: Harvard University Press.

[2] Brown, C.R., F.L. Greitzer, and A. Watkins. *Toward the development of a psycholinguistic-based measure of insider threat risk focusing on core word categories used in social media*. in *2013 Americas Conference on Information Systems*. 2013. Chicago, Illinois: AIS, 1-8.

[3] Brown, C.R., A. Watkins, and F.L. Greitzer. *Predicting insider threat risks through linguistic analysis of electronic communication*. in *2013 46th Hawaii International Conference on System Sciences*. 2013. Wailea, Hawaii: IEEE, 1849-1858. doi:10.1109/HICSS.2013.453.

[4] Buller, D.B. and J.K. Burgoon. *Interpersonal deception theory.* Communication Theory, 1996. **6**(3): 203-242.

[5] DePaulo, B.M., J.J. Lindsay, B.E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper. *Cues to deception.* Pyschological Bulletin, 2003. **129**: 74-112.

[6] Ekman, P. and W.B. Friesen. *Nonverbal leakage and clues to deception.* Psychiatry, 1969. **32**: 88-106.

[7] Ekman, P. and M. O'Sullivan. *Who can catch a liar?* American Psychologist, 1991. **46**(9): 913-920.

[8] Greitzer, F.L., L.J. Kangas, C.F. Noonan, C.R. Brown, and T. Ferryman. *Psychosocial modeling of insider threat risk based on behavioral and word use analysis.* e-Service Journal, 2013. **9**(1): 106-138. doi:10.2979/eservicej.9.1.106.

[9] Greitzer, F.L., L.J. Kangas, C.F. Noonan, A.C. Dalton, and R.E. Hohimer. *Identifying at-risk employees: Modeling psychosocial precursors of potential insider threats*. in *2012 45th Hawaii International Conference on System Sciences*. 2012. Maui, Hawaii: IEEE, 2392-2401. doi:10.1109/HICSS.2012.309.

[10] Habermas, J., ***The theory of communicative action***. Reason and the Rationalization of Society. 1981. Boston, MA: Beacon Press.

[11] Hancock, J., J. Birnholtz, N. Bazarova, J. Guillory, J. Perlin, and B. Amos. *Butler lies: Awareness, deception and design*. in *CHI'09*. 2009. Boston, MA: ACM.

[12] Hancock, J., L.E. Curry, S. Goorha, and M. Woodworth. *On lying and being lied to: A linguistic analysis of deception in computer-medicated communication.* Discourse Process, 2008. **45**(1): 1-23. doi:10.1080/01638530701739181.

[13] Hancock, J., C. Toma, and N. Ellison. *The truth about lying in online dating profile*. in *CHI'07*. 2007. San Jose, CA: ACM, 449-452. doi:10.1145/1240624.1240697.

[14] Ho, S.M., ***Cyber insider threat: Trustworthiness in virtual organizations***. 2014: Lambert Academic Publishing. 1-436.

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

[16] Ho, S.M., J.T. Hancock, C. Booth, X. Liu, S.S. Timmarajus, and M. Burmester. *Liar, Liar, IM on Fire: Deceptive language-action cues in spontaneous online communication*. in *IEEE International Conference on Intelligence and*

Security Informatics. 2015. Baltimore, MD: IEEE, 157-159. doi:10.1007/978-1-4799-9889-0/15.

[17] Ho, S.M. and J. Hollister. *Cyber insider threats in virtual organizations*. In Encyclopedia of Information Science and Technology, edited by Khosrow-Pour, M.E. 2015. Information Resource Management Association: Hershey, PA. 1517-1525.

[18] Hosmer, L.T. *Trust: The connecting link between organizational theory and philosophical ethics.* Academy of Management Review, 1995. **20**(2): 379-403.

[19] Jarvenpaa, S.L. and D.E. Leidner. *Do you read me? The development and maintenance of trust in global virtual teams*, in *INSEAD Working Paper Series*, edited by Austin, U.o.T.a., 1997: Fontainebleau, France. 1-44.

[20] Jarvenpaa, S.L. and D.E. Leidner. *Communication and trust in global virtual teams.* Journal of Computer-Mediated Communication, 1998. **3**(4). doi:10.1111/j.1083-6101.1998.tb00080.x.

[21] Jarvenpaa, S.L. and D.E. Leidner. *Communication and trust in global virtual teams.* Organization Science, 1999. **10**(6): 791-815.

[22] Jones, S. and S. Marsh. *Human-computer-homan interaction: Trust in CSCW*. ACM SIGCHI Bulletin, 1997. **29**(3): 36-40. doi:10.1145/264853.264872.

[23] Kanawattanachai, P. and Y. Yoo. *Dynamic nature of trust in virtual teams.* Journal of Strategic Information Systems, 2002. **11**(3): 187-213. doi:10.1016/S0963-8687(02)00019-7.

[24] Klimoski, R.J. and B.L. Karol. *The impact of trust on creative problem solving groups.* Journal of Applied Psychology, 1976. **61**(5): 630-633. doi:10.1037/0021-9010.61.5.630.

[25] Kramer, R.M., M.B. Brewer, and B.A. Hanna. *Collective trust and collective action: The decision to trust as a social decision*, in *Trust in Organizations: Frontiers of Theory and Research*, Kramer, R.M. and T.R. Tyler. 1996. Sage Publications, Inc.: Thousand Oaks, CA. 357-389.

[26] Lewicki, R.J. and B.B. Bunker. *Developing and maintaining trust in working relationships*, in *Trust in Organizations: Frontiers of Theory and Research*, Kramer, R.M. and T.R. Tyler. 1996. Sage Publications: Thousand Oaks, CA. 114-139.

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

[28] Pennebaker, J.W., C.K. Chung, M. Ireland, A. Gonzales, and R.J. Booth. *The development and psychometric properties of LIWC2007*, 2007.

[29] Pennebaker, J.W. and L.A. King. *Linguistic styles: Language use as an individual difference.* Journal of Personality and Social Pyschology, 1999. **77**(6): 1296-1312. doi:10.1037/0022-3514.77.6.1296.

[30] Pennebaker, J.W., M.R. Mehl, and K.G. Niederhoffer. *Psychological aspects of natural language use: Our words, our selves.* Annual Review of Psychology, 2003. **54**: 547-577. doi:10.1146/annurev.psych.54.101601.145041.

[31] Rotter, J.B. *A new scale for the measurement of interpersonal trust.* Journal of Personality, 1967. **35**(4): 651-665. doi:10.1111/j.1467-6494.1967.tb01454.x.

[32] Schultz, E.E. *A framework for understanding and predicting insider attacks.* Computers & Security, 2002. **21**(6): 526-531.

[33] Taylor, P.J., C.J. Dando, T.C. Ormerod, L.J. Ball, M.C. Jenkins, A. Sandham, and T. Menacere. *Detecting insider threats through language change.* Law and Human Behavior,, 2013. **37**(4): 267-275. doi:10.1037/lhb0000032.

[34] Toma, C. and J. Hancock. *Reading between the lines: Linguistic cues to deception in online dating profiles*. in *International Conference on Computer Supported Cooperative Work (CSCW 2010)*. 2010. Savannah, Georgia: ACM. doi:978-1-60558-795-0/10/02.

[35] Toma, C. and J. Hancock. *What lies beneath: The linguistic traces of deception in online dating profiles.* Journal of Communication, 2012. **62**: 78-97. doi:10.1111/j.1460-2466.2011.01619.x.

[36] Wright, R.T. and K. Marett. *The influence of experiential and dispositional factors in phishing: An empirical investigation of the deceived.* Journal of Management Information Systems, 2010. **27**(1): 273-303.

[37] Zand, D.E. *Trust and managerial problem solving.* Administrative Science Quarterly, 1972. **17**(2): 229-239. doi:10.2307/2393957.

[38] Zhou, L., J.K. Burgoon, D.P. Twitchell, T. Qin, and J.F. Nunamaker Jr. *A comparison of classification methods for predicting deception in computer-mediated communication.* Journal of Management Information Systems, 2004. **20**(4): 139-165.

[39] Zhou, L., D.P. Twitchell, T. Qin, J.K. Burgoon, and J.F. Nunamaker Jr. *An exploratory study into deception detection in text-based computer-mediated communication*. in *Proceedings of the 36th Hawaii International Conference on Systems Sciences*. 2003. Hawaii: IEEE. doi:0-7695-1874-5/03.

[40] Zhou, L. and D. Zhang. *Can online behavior unveil a deceiver? .* in *Proceedings of the 37th Hawaii International Conference on System Sciences*. 2004. Hilton Waikoloa Village Big Island, Hawaii: IEEE Press.