Real or Spiel? A Decision Tree Approach for Automated Detection of Deceptive Language-Action Cues

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

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

Muye Liu
Florida State University
ml11an@my.fsu.edu

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

Mike Burmester
Florida State University
burmeste@fsu.edu

Abstract

As the use of computer-mediated communications has increased, the potential risk of online deception has grown—as has the importance of better understanding human behavior online to mitigate these risks. Previous research has demonstrated that linguistic features provide crucial cues to detect deception, and that reasonable accuracy in detection of deception can be achieved by applying certain classification methodologies to these cues. This paper expands on this line of inquiry, and presents findings from a study conducted in the Spring of 2015. Our findings suggest a viable process for and the feasibility of using a decision-tree classification approach to develop an automated process to detect deception in computer-mediated communications.

1. Introduction

Advances in communication technology have increased—and continue to enhance—the speed, geographical scope, and convenience of human communication. However, these technological advances also expose and emphasize the threat to certain human vulnerabilities. Not the least of these is the vulnerability to deception. Deception is commonly defined as "a message knowingly transmitted by a sender to foster a false belief or conclusion by the receiver" [1, p. 205].

Deceptive communication can be understood as involving persuasive strategies and activities [24], aiming "at influencing the beliefs, attitudes and behaviors of others by means of deliberate message distortions" [19, p. 99]. In the specific context of computer-mediated communication (CMC), deceptive communication presents a variety of risks, including attempts at social engineering, spam, phishing, identity theft, and even fraud. Thus, one's ability to understand how parties determine the truthfulness of an individual with whom they are interacting—particularly online—is important, but also challenging. These challenges include assessing identity and trusting information exchanged. Another challenge is being able to examine the actions and reactions of parties as they interact [32].

Significant research in deceptive communication and deception detection in both face-to-face (F2F) and CMC environments has been done by many communication theorists as well as computer scientists. However, how to "translate" deceptive language action cues into actionable decision-tree analysis, such that it may be possible to develop an automated process for detecting deception has been lacking in the literature. This paper attempts to address the following research question: *Can a decision tree approach be used to automate the process of identifying intentional deception in spontaneous computer mediated communication across a pluralistic background of users?* For clarity, "Language-action cues" refers to linguistic styles, phrases, patterns, or actions in an actor's written expression and manifested as an indirect or subtle signal to other actors.

In the following sections, we will first describe the logistics and nature of deception, and the extent to which decision-tree analysis has been applied to date in the area of deception detection. Then, we will discuss our research design and methodology, along with an analysis of the findings. Finally, we will discuss implications and limitations, and conclude with insights of potential future work.

2. Deception

2.1. The True Nature of Deception

Fundamentally, and irrespective of environment (CMC vs. F2F), deception refers to purposefully concealing the truth, either by omission or

1530-1605/16 $31.00 © 2016 IEEE
DOI 10.1109/HICSS.2016.462

commission [26]. In practice, however, deception is much more complex than this characterization suggests. In order to understand the true nature of deception, it is necessary first to understand the fundamental nature of human communications (whether deceptive or not) and the norms implicit in communication. In particular, it is important to understand the concept of the "truth bias," or the fundamental presumption of truthfulness [1, 12]. That is, when engaging in communication, absent any indications to the contrary, a speaker is considered "truthful unless proven deceptive." This truth bias operates "… to reduce a person's search for the [cues] that might reveal [a] lie" [18, p. 380], and functionally reduces or even removes any incentive to be suspicious and look more closely at cues (i.e., verbal, behavioral) potentially associated with deception. In short, this truth bias makes us less inclined to look for, and thus less apt to pick up on, cues that might reveal deception [18]. This perhaps accounts (if only in part) for Buller and Burgoon [1, p. 205] finding that deception is in fact somewhat pervasive, with one quarter or more of conversations including some element or amount of deception.

Studies on deception share the common focus of identifying particular cues (behavioral, contextual, verbal or textual) that can be associated with deception. In so doing, these studies reveal several essential elements of deception. First, deception should be understood as a volitional and intentional act. An inadvertent "mistake of fact" type of error, for example, does not constitute deception. Second, deception can be typed as being either spontaneous ("on the fly") or planned [27]. Third, the consequences of a deception may range from de minimus (as in a "little white lie") to significant [13]. Fourth, the type and significance of the deception are linked. For instance, a spontaneous lie is likely to be less significant than a planned lie. Fifth, a deception may be self-serving (benefitting the deceiver), other-oriented (benefitting the recipient other third parties), or both. Finally, both the mode of communication (synchronous [32] or asynchronous [28]) and the specific medium chosen may provide insight into to the type and severity of the deception [27]. That is, would someone telling a serious, planned lie be more likely to use synchronous media (such as a telephone call or instant message) or asynchronous media (such as a letter or an e-mail)?

Some of the early work of deception and deception detection comes from the fields of psychology and psychiatry. Ekman and Friesen [9] examined nonverbal behavior communication—that is, body language as indicators of deception. In particular, they explored how certain nonverbal cues—unconsciously or subconsciously manifested by a party to a communication—operate to provide clues to deception. This phenomenon is referred to as nonverbal *leakage* in deceptive situations. Ekman and O'Sullivan [10] also wrote about one particular problem that nonverbal leakage can present: attribution error, or, as Ekman refers to it, the *Othello error*. During the interaction between two parties, one party ("A") may pick up cues from the other party ("B") that make "A" suspicious that "B" is being deceptive about the subject under discussion, when in fact "B" is telling the truth but is giving off these cues as a result of other factors (such as anxiety or fear, or even shock at being suspected/ falsely accused). "A" misattributes, misunderstands or misinterprets these cues as confirming his/her suspicion of deception, acts accordingly, often with potentially tragic (or at least, unfortunate) results.

As to actually detecting deception, Ekman and Friesen [9] posited that deception can be revealed based on three dimensions: (1) whether the act of deception is salient with an explicit focus on the conscious concerns, (2) the role-play between the deceiver and the deceived, and (3) an interactive process, or collaboration, to discover or to maintain deception. Granhag and Strömwall [11] further suggested three types of internal processes can be translated into three types of non-verbal deceptive behaviors: emotional, cognitive, and attempted control. The emotional approach to discover deceptive behavior describes the emotional state of the deceiver, e.g., a deceiver may display a sense of guilt. The cognitive approach identifies the complexity in deception; for example, a deceiver in a synchronous communication will speak slowly or exhibit speech disturbances. Attempted control behaviors describe the struggle of a deceiver to sound normal by trying to appear more honest and genuine; however, in the communication process, he may eventually reveal more cues from which deceptive behavior may be discerned. Based on the complexity of deception, Granhag and Strömwall [11] further evaluated verbal behaviors using a credibility assessment technique, statement validity analysis (SVA), to focus on the verbal features that correlate with deception. The findings reveal that quantity of details may be the most determinative factor (i.e. truthful statements have more details than false ones). Consistency in detail provided is another important factor (an average deceiver being less consistent in the details provided).

DePaulo, Kashy et al. [7] also focused on identifying cues in a F2F environment, although their work involved examination of the psychological ramifications of lying and how liars minimize these. Based on the framework of social distance theory

[17], DePaulo, Kashy et al. [7] posit that, in order to avoid the social discomfort attendant to deception, liars will separate or distance themselves from the person they are deceiving. In the context of choosing a channel of communication for deception, social distance theory suggests that deceivers would be more likely to choose a channel (or media type) that gives fewer cues to their interacting partner. Social distance theory further suggests that, again in order to avoid social discomfort, we can expect that most lies are self-serving—i.e. they are told for the benefit of the deceiver and not the receiver or any third party, for example to save face or preserve a particular image [22]. Indeed, particularly as regards serious lies, the deception was self-serving in 9 out of 10 lies told [23, p. 163]. Social distance theory accounts for much of what has been discussed in the foregoing sections as well, but with its own 'spin' on it. For example, DePaulo, Lindsay et al. [8] treats the Othello error, pointing out that a speaker may be insecure and have as many of what they refer to as "self-regulatory demands" as a liar when the speaker is concerned about failure (i.e. being taken for a liar).

Finally, social distance theory would suggest that relatively more lies (particularly everyday lies) are told to those having a more remote (i.e. less close) relationship with the deceiver and/or to those with whom the deceiver interacts only occasionally—presumably because the deceiver feels relatively safer in deceiving strangers or remote others vis-à-vis chances of being detected and/or the consequences of being detected.

In contrast to social distance theory is Buller and Burgoon [1] Interpersonal Deception Theory (IDT). Rather than examining leakage cues specifically, or how deceivers work to increase the psychological distance between themselves and the person they are deceiving—IDT focuses on the interpersonal nature of deceptive (and truthful) communications and behaviors—rather than the communication or message itself—and views deception as an iterative, strategic process that may be best characterized as something of a chess match between deceiver and deceived [4, 5]. This portrait of deception is given additional texture and color by Miller et al. [19] describing deception as "…a general persuasive strategy that aims at influencing the beliefs, attitudes and behaviors of others by means of deliberate message distortions" (p. 99). In short, as Miller and Stiff [20] and Stiff [24] characterize deceptive communication as an act that involves the intentional use of persuasive strategies and activities to manipulate the receiver.

Buller, Burgoon et al. [2] tested IDT by evaluating a receiver's perceptions and suspicions to understand how a deceiver strategizes and shapes his/her behavior. Their results indicated that, to avoid detection by a receiver in the course of communication, a deceiver often finds s/he must adjust his or her deceptive strategy 'on the fly'—possibly numerous times during the interaction—and therefore the deceptive strategy tends to be fluid and variable rather than solid and stable. That is, IDT assumes that, although both strategic and non-strategic behaviors manifest during interactive deception, deception is fundamentally a strategic practice engaged in to satisfy multiple (and sometimes competing) objectives of the deceiver, including impression management, relational communication, emotion management and conversation management [3]. IDT further posits that the influence of a deceiver's behavior on a receiver affects the receiver's behavior, which in turn affects the deceiver's message strategy. Thus, the language choice of a deceiver's message would reflect strategic attempts to manipulate information through non-immediate language. It is worth noting that not only deception can influence the dynamics of a communication, suspicion also has a similar influence. Burgoon, Buller et al. [4], [5] examined this phenomena from the perspective of the receiver's perceptions of the truthfulness in a deceiver's message and any suspicions the receiver's response may demonstrate.

Subsequent tests of IDT have generally relied on either non-strategic leakage cues (e.g. visual and tactile cues pointing to deceptive intent), or uncovering non-immediacy cues that may also be useful in detecting deception [3-5]. However, despite identifying these potential ways in which deception may be detected, humans continue to perform poorly in detecting lies [10].

2.2. The Problem of CMC-based Deception

The problems of CMC-based deception incurred when deceivers must contend in an online environment and the resultant cues that can be observed differ accordingly. Indeed, as media richness theory posits [6, 25], while not exclusively belonging to CMC, the type of communication method (F2F or CMC) chosen by a deceiver can itself provide cues or clues as to the type and significance of the deception.

In essence, because deception is subject to the interpretation of the individual being lied to, deceptive actors will tend to choose media that provide multiple cues, immediate feedback and an opportunity for personalization—all presumably in an attempt to obfuscate their lies by sending conflicting cues [6, 25]. Thus, whereas social distance theory would predict that deceivers would tend to choose "less rich"

media such as e-mail, media richness theory predicts that deceivers would be expected to choose richer media/means of communication—particularly face-to-face exchanges.

Media richness theorists suggest that the nature of the message (equivocal or unequivocal) drives the choice of medium for transmission for that message [6, 25]. The *richness* of the medium is determined by evaluating four factors: feedback (immediate or delayed); number of cues available to the receiver (including social cues); language variety (i.e. the type and variety of symbols used to convey the particular message); and personal focus (i.e. infusing the message with personal feeling/ emotions) [6]. The richer the media type, the easier it is selected in the communication of equivocal messages (i.e. where there might be ambiguity requiring clarification) than in the communication of unequivocal messages. In the context of identifying cues to deception in a message, media richness theory seemingly suggests the opposite of social distance theory. That is, because deception is subject to the interpretation of the individual being lied to (i.e. lies are equivocal in nature), liars will tend to choose to lie using rich media that provides multiple cues, immediate feedback and an opportunity for personalization—all presumably in an attempt to obfuscate their lie by, for example, sending conflicting cues. So, whereas according to social distance theory, liars would tend to choose less rich (i.e. cue lean) media such as e-mail, according to media richness theory, liars would be expected to choose richer media/means of communication—particularly face-to-face exchanges.

Hancock, Thom-Santelli et al. [15] proposed a feature-based model, which looks at the specific features of the media chosen as a means of deriving cues to deception. Several questions can be asked. For instance, does the media allow for real-time communication? Is the exchange recorded/recordable? Is the media distributed? Are the communicating parties collocated in the same location e.g., copresent, or are they in different locales? This model assumes fundamentally that deception is spontaneous, suggesting that deception is more likely to occur when media is "synchronous and distributed, but non-recordable" [27, p. 209]. Again, while this isn't exclusive to CMC, the feature-based model is clearly applicable to the analysis of CMC issues.

In addition to the impact of media features on deception, Hancock et al. [13] demonstrated the significance of linguistic features such as first-person singular, emotional toned words, inhibition words, prepositions and conjunctions, as indicators that can differentiate truth tellers from deceivers. Zhou et al. [30], on the other hand, evaluated deceptive cues in a desert survival context and found that deceivers tended to be wordy using peripheral expressions in their messages. In contrast to F2F interactions where deceivers have been found to be more concise, Zhou and Zhang [31] found that in asynchronous online contexts, deceivers tend to be more active in language usage and take shorter pauses between messages. That said, however, they also made an important distinction between synchronous and asynchronous communication: the exercise in their particular experiment—which involved asynchronous communication—involved, by its nature and design, an element of persuasiveness on the part of the deceptive actor in the dyad, which, of necessity, required the deceptive actor to be strategic and the deception to be planned and thus unlikely to have been facilitated by a synchronous means of communication.

These theories, derived from previous research in various fields including psychology, criminal justice, linguistics and criminology also inform the framework of online deception and provide a solid theoretical and evidentiary means of analysis for the detection of deception. In particular, exploring the concept of immediacy in communication, discussed in more detail below and which is common to all four of the theories described above, is instructive in this regard.

2.3. Deceptive Language-action Cues

As the previous sections have made clear, our ability to detect deception, in any environment, depends on numerous factors, including the availability of certain types of cues. These cues function as an alert to the receiver to be more critical of the information being provided. In a CMC environment, the availability of cues is reduced (being limited to the text in message-based exchanges), relative to F2F communication. Nonetheless, certain communication cues can still be observed and catalogued within a CMC environment [13, 16, 31].

These communication features and cues, such as first-person references, emotional words, inhibition words, prepositions, and conjunctions, have been shown to be indicators that can differentiate deceivers from truth tellers [13]. Level of detail (less or more), use of more or fewer sensory or spatiotemporal words, and changes in the diversity and complexity of language are among some of these features [21].

Another important feature is verbal immediacy, referring to ways in which an actor can associate (or distance) him/herself from the content of his/her message [32]. Such cues (whether verbal or non-verbal) are particularly important in detecting

deception. In the physical environment, non-verbal immediacy cues include eye contact, body posture, facial gestures, etc. While these specific cues were derived without reference to CMC, there are certain immediacy cues in CMC environments, including the delay in response, which, according to social distance theory [27], creates a psychological distance between liar and lie.

In addition, Zhou and Zhang [31] found that deceivers tended to be more wordy in their messages, but provided less relevant or meaningful information. [31], Zhou and Zhang [32] also found that deceivers tend to use more restricted vocabulary and syntax, use fewer self-references and to be more casual in their linguistic style. Finally, in contrast to F2F interactions, deceivers tend to be more active in language usage and take shorter pauses between messages [31].

From an online CMC environment perspective, then, quantity and consistency of detail may be measureable via language-action cues in a dynamic exchange of text messages by focusing on features such as the use of adverbs, adjectives, and inclusive words. Based on this, it is possible to benchmark verbal indicators (such as word count and details of information disclosed) and capture certain non-verbal behaviors (latency and usage of expression words) which can then be statistically computed using the findings from Zhou and Zhang [31] work.

2.4. Spontaneous vs. Planned Deception

Deceptive communications can be planned (in a CMC context, social media profiles, online reviews, and blogs are examples), or constructed on-the-fly and spontaneous (in a CMC context, interactive chat descriptions or tweets are examples). As Whitty, Buchanan et al. [27] discussed, there are observable differences in deception cues between these two types of lies. Moreover, they further pointed out that this distinction of deception type (planned vs. spontaneous) may have implications for the preferred communication medium of the deceiver [27]. For example, a planned deception often requires time to construct, while a spontaneous lie obviously would not—and some media types are more amenable to this than others (compare, for example, e-mail to instant messages). In addition, Whitty, Buchanan et al. [27] further posit that planned lies tend to be more serious than spontaneous lies—so this distinction between types of deception also has implications for the magnitude of the deception itself.

2.5. Decision Tree Approach in Deception Detection

Zhou, Burgoon et al. [29] provided detailed discussions on comparing the effectiveness of CMC-based deception detection across four primary classification methods: discriminant analysis, logistic regression, neural networks and decision trees. Their work tested the accuracy of each of these approaches (i.e. how well each did in terms of correctly identifying deception); however, this study did not address how well these methods lend themselves to application as the basis for development of an automated deception detection system (indeed, the authors specifically state that such is not their intention). As Zhou, Burgoon et al. [29] suggested, decision tree analysis provides reasonable accuracy in the context of "important" cues (i.e. cues known to have statistical significance in revealing deceptive intent). Our study as depicted in this paper builds on this thread by operationalizing one of the classification approaches—decision tree analysis—from the perspective of potential for use in developing an automated system.

To sum, our research is different from previous work in several ways. In particular, our approach collects spontaneous conversational data based on randomized interpersonal scenarios. These scenarios generate automated insertion of text for the study of text-based cues across various topics, demographics, age groups, and genders. In addition, the game-based approach we employ is innovative in that it offers both an opportunity and *a motivation* for the actors to deceive (or, at a minimum, express their intent to do so). Moreover, our study specifically examines linguistic cues to planned deception, specifically in synchronous, co-present, distributed and recordable media. Finally, our study adopts decision-tree analysis to further examine the language-action cues as decision points as base that differentiates deceiver from truth-teller.

3. A Sociotechnical Research Design

The research approach developed to answer the research question posed above focused on developing specific metrics for language cues and word choice as information behaviors, and analyzed communication patterns that distinguish between different communication typologies. An interactive game interface was developed connecting to the chat feature of Google+ Hangout, presenting players with interactive scenarios requiring them to write either deceptive or truthful statements. The framework, illustrated in Figure 1, provides a conceptual basis for understanding, analyzing and designing ways to explore the dynamics of intentional deception. The identification of text-based cues from these scenarios

provides a means of understanding and measuring the decision parameters needed to detect online deception. It also enables us to observe how people lie successfully (or unsuccessfully) in different circumstances.

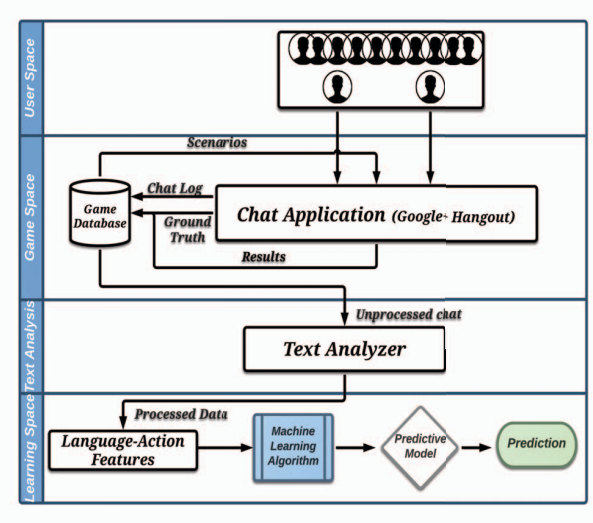

Figure 1. Design framework

The game ("*Real or Spiel*") is an online game that simulates a real-time interactive deception scenario through synchronous communication channels. It consists of four distinct pieces: 1) the user space, which holds and manages registration of users and user profiles, and maintains a list of active users available for online game sessions; 2) the game space, which has two main components, the database and the chat application, and is responsible for selecting scenarios managing game sessions, collecting surveys and logging the players' chats, 3) the text analysis space, where the chats collected in the game space are analyzed; and 4) the learning space of the language-action cues, which is intended to support future development of a "live" machine learning system that would be able to detect deception in CMC environments.

Each game involves two participants (i.e. players), who are placed in assigned pairings by the research team, and are then randomly assigned an *outer role* as either an initiating *speaker* or a *detector* in each gaming session. The speaker in each scenario is also randomly assigned an *inner* role—either *saint* (truthful) or *sinner* (deceptive). The speaker establishes the *ground truth* at the beginning of each scenario by truthfully answering questions on a particular topic from common knowledge domains such as finance, skills or personal experiences. For example, the ground truth question might be something like "Have you ever been given a speeding ticket?" If, for example, the speaker has been given one, s/he would establish the ground truth by answering "yes." This provides a baseline against which to assess the truthfulness or deceptiveness of his/her subsequent responses to the questions posed by the detector during the scenario. In the above example, the detector would ask questions designed to learn whether or not the speaker had received a ticket, and the speaker would try to convince the detector that s/he has not been given one. At the end of each scenario, the detector tries to determine whether the speaker was being deceptive or truthful based on question-and-answer exchanges.

Processing of the data collected during the gaming sessions occurs in the text analysis layer, which consists of two segments or layers. The first layer of the text analyzer cleans and categorizes the raw, unprocessed data using a statistical text analysis tool to derive relevant statistical inferences that link language-action cues (such as use of negative/ positive words, self-references, word count and level of embellishment or description), to possible deceptive cues. From this, a rich catalog of language-action cues is created that may be used to facilitate machine learning. The second layer normalizes the processed data from the previous layer, to evaluate the consistency of the given conversation, as well as the use in the conversation of various linguistic cues used to measure deception (such as spontaneity, emotion, self-references, etc.) [13].

4. Data Collection and Analysis

The data was collected during Spring 2015. Each game session consists of two players; a speaker and a detector. The data set used for analysis included a total of 10 games sessions. There were 20 participants; 12 males and 8 female players with ages ranging from 18 to 65 years old. Players' names were replaced with pseudo-names in order to protect their privacy.

Table 1. Game Session / Phases

Phase/Player	Player 1	Player 2
Phase 1	Speaker & Saint	Detector
Phase 2	Detector	Speaker & Saint
Phase 3	Speaker & Sinner	Detector
Phase 4	Detector	Speaker & Sinner

Each game session lasted approximately 30 minutes game, and is broken into four distinct phases as shown in Table 1. Game sessions were launched with the objective that each player was able to change or rotate outer roles such that s/he is a speaker twice and a detector twice. Each phase lasts approximately

7.5 minutes, after which, the roles of the players were automatically swapped. Thus, over the 10 game sessions from which we collected data, there were 40 total phases.

4.1 Data Cleaning and Preparation

The data collected were cleaned and spell checked. The spell checker corrected most of the spelling errors in the chat text, and common instant-messaging abbreviations ("LOL", "U", "2", "4", etc.) were converted to their corresponding full written forms. In addition, we excluded from our analysis data from any phrase containing fewer than 50 words total. As a result, 20 phases of dataset were removed and excluded from the analysis. There were approximately 300 lines of script processed in our analysis. The mean for the total word count across the data set was 109 words per session for truth-tellers, and 79 words per session for deceivers.

Once the data had been cleaned, the linguistic cues from the data were extracted according to the categories established in the Linguistic Inquiry and Word Count (LIWC) tool [21, 23], and the text corpus was converted into numerical representation. Table 2 depicts specific LIWC categories examined. These data were processed, analyzed and calculated first on a percentage basis and then on a word-count basis.

Table 2. LIWC Categories and Coding Schema

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

4.2. Decision Tree

Decision trees are a powerful and popular mechanism for describing data. The rules derived from the decision trees can be easily understood and even implemented directly as a database query for retrieval purposes. Moreover, since the size of the data set in our case is relatively small, the decision rules are not complex. Accordingly, once the data had been cleaned and prepared as described above, we chose to apply a decision-tree analysis approach using Matlab R2015a to derive rules or questions from our data.

We first developed a decision tree from the percentage-based data. Figure 2 depicts the decision tree derived from this analysis. Table 3 illustrates the pseudo code developed from the percentage-based decision-tree analysis, to implement the rules derived from it.

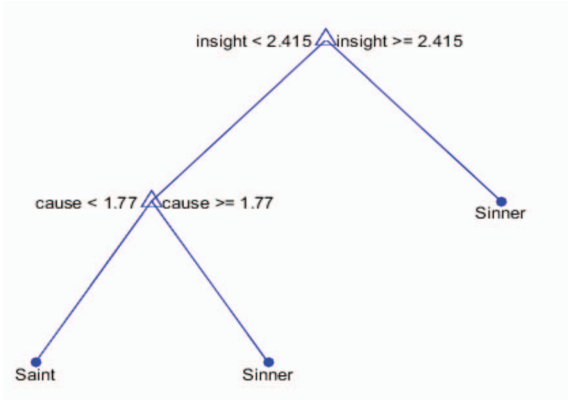

Figure 2. Decision Tree Based on Percentage

Table 3. Pseudo Code - Percentage Based

Pseudo Code <Percentage Version>

```
if insight < 2.415
   if cause < 1.77
      classify actor as Saint
   else
      classify actor as Sinner
   end
else
   classify actor as Sinner
end
```

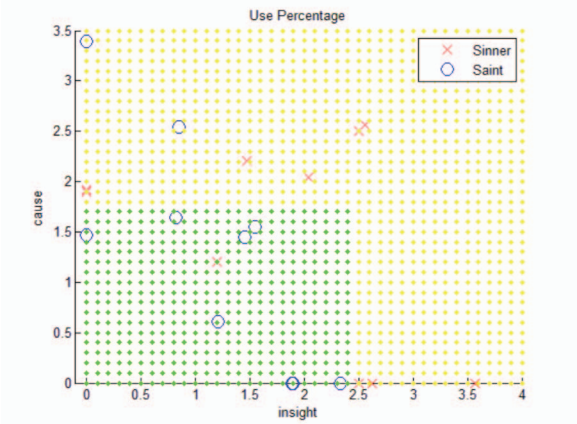

Figure 3. Percentage Plot Data Visualization

Figure 3 depicts the plot of the percentage-based data analysis. The decision regions are shown in color. The predictor *insight* appears on the horizontal (*x*) axis, and the predictor *cause* appears on the vertical (*y*) axis. Each data point consists of a pair of values (*insight, cause*). As can be seen, the data are

not well separated in this case. Two (2) Saint data points fall within the Sinner region and one (1) Sinner data point falls within the Saint region.

In order to address this issue, we further analyzed our data on a word-count basis. Figure 4 depicts the decision-tree derived from that analysis.

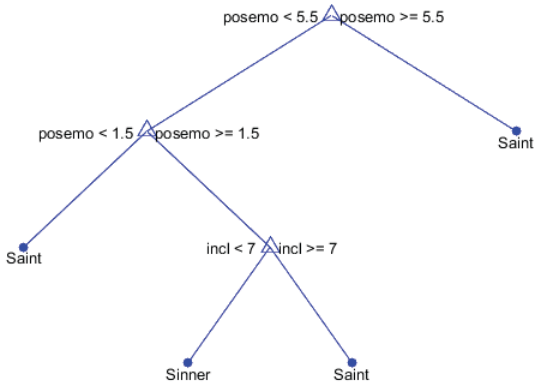

Figure 4. Decision Tree Based on Word Count

Table 4 details the pseudo code developed from the word count-based decision-tree analysis, to implement the rules derived from it.

Table 4. Pseudo Code - Word-Count Basis

Pseudo Code <Word Count Version>

```
if posemo < 5.5
   if posemo < 1.5
      classify actor as Saint
   else
      if incl < 7
         classify actor as Sinner
      else
         classify actor as Saint
      end
   end
else
   classify actor as Saint
end
```

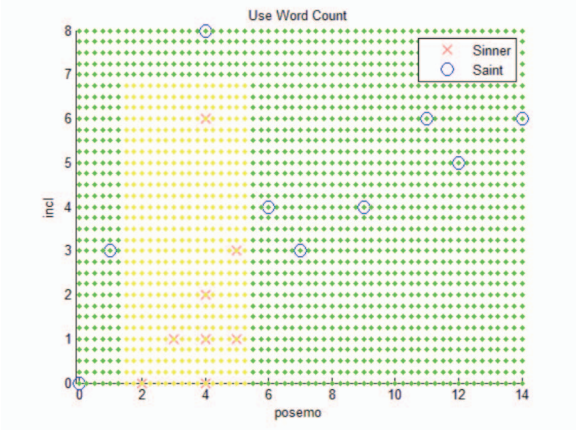

Figure 5: Word Count Plot Data Visualization

Figure 5 depicts the plot of the word-count based data analysis. The predictor *posemo* appears on the horizontal (*x*) axis, and the predictor *inclusivity* appears on the vertical (y) axis. Each data point consists of a pair of values (*posemo, incl*). The data in this case separates well. All Saint data points fall within the Saint region and all Sinner data points fall within the Sinner region.

4.3 Language-action Cues

Cognitive Process (Cogmech): Our results suggest that words reflecting cognitive process (i.e., active thinking) are used more freqently by deceptive actors than by truthful actors. Few studies have examined this LIWC category as a whole. However, different studies have looked at specific subcategories. For example, the results of a study by Hancock, Curry et al. [14] suggest that deceptive actors (specifically, motivated deceptive actors) use fewer causal words—than truthful actors. Similarly, results of a study by Newman, Pennebaker et al. [21] suggest that deceptive actors use fewer words of exclusivity than did truthful actors. While these results would appear to be in contrast to our own findings, we speculate that the reason for this apparent contradiction is the difference in the specific nature, kind and quality of the deceptive act(s) in which the participants in the respective studies were asked to engage.

Affective Processes (Affect): In our initial analysis, we measured usage of words expressing positive emotion (*posemo*) and words expressing negative emotion (*negemo*) separately. Neither posemo nor negemo word usage was found to be statistically significant individually. However, we did obtain a statistically significant result when we combined words from both categories into the larger LIWC heading of "affect" for analysis. Specifically, results from this combined analysis suggests that deceptive actors will tend to use more "affect" words (i.e. words showing *posemo* and/or *negemo*) than truthful actors.

5. Implications and limitations

Although the small sample size is a notable limitation to our study, our results nonetheless appear generally to support findings in similar studies [14, 28, 31]. Our results indicate that descriptive and embellishing phrases such as "interesting" and "valuable" were more often used in deceptive communications than in truthful ones. Additionally, our results seem to generally indicate that, contrary to the predictions of social distance theory, deceptive

actors tend to use language that shortens the social distance between themselves and the person with whom they are communicating—thereby supporting media richness theory and the feature-based model. In particular, our results differed from some other studies in that ours suggest that deceptive actors tend to use more "I" references than truthful actors. This particular discrepancy certainly bears more targeted investigation—and would benefit from having a larger data sample. Finally, although the time lag between players' responses was not included in the regression analysis (as it was not a specific factor of interest in the context of the study), it is worth mentioning that the time lag between a detector's questions and a sinner's replies tended to be longer than the time lag between the detector's question and a saint's response. This phenomenon also merits further exploration, and would likewise benefit from a larger dataset.

6. Conclusions and Future Work

The results discussed above indicate that the deceptive language-action cues used in spontaneous communication are (or can be) materially different to the cues used in transcript and online profile description. They support the assertion that deception is not only a strategic process [1] involving persuasive activities and interactions [24], but it is also is context-sensitive. They further demonstrate that the strategies employed by deceptive actors will differ based on different modes (asynchronous, synchronous) of communication. Finally, our results provide further evidence and support to the idea that the identification of key text-based cues, correlated to deception, can be effectively used to develop models of behavior that can be used to predict or detect deceptive (or truthful) behavioral intent in a CMC environment. The methodology employed—in particular, decision-tree analysis and pseudo-code development—allows for generalization of these significant cues to a broader population, while filtering out less significant cues, and hence can be used to inform the development of automated deception intelligence learning machines.

This paper presents the foundations for developing a machine learning system that can identify deception in a spontaneous CMC environment (Figure 1). Future research will include employing interactive social media games to simulate additional deception scenarios, and mapping out additional known and unknown deceptive language action cues. Future study will also include analysis of the response time lag discussed above. The ultimate objective is eventually to design and implement a "live" machine learning system that is able to detect deception in CMC environments.

7. 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.

8. References

[1] Buller, D.B. and J.K. Burgoon. *Interpersonal deception theory.* Communication Theory, 1996. **6**(3): 203-242. doi:10.1111/j.1468-2885.1996.tb00127.x.

[2] Buller, D.B., J.K. Burgoon, A. Buslig, and J. Roiger. *Testing interpersonal deception theory: The language of interpersonal deception.* Communication Theory, 1996. **6**(3): 268-289. doi:10.1111/j.1468-2885.1996.tb00129.x.

[3] Burgoon, J.K. and D.B. Buller. *Interpersonal deception: III. Effects of deceit on perceived communication and nonverbal behavior dynamics.* Journal of Nonverbal Behavior, 1994. **18**(2): 155-184.

[4] Burgoon, J.K., D.B. Buller, L. Dillman, and J.B. Walther. *Interpersonal deception. IV. Effects of suspicion on perceived communication and nonverbal behavior dynamics.* Human Communication Research, 1995. **22**(2): 163-196.

[5] Burgoon, J.K., D.B. Buller, A.S. Ebesu, C.H. White, and P.A. Rockwell. *Testing interpersonal deception theory: Effects of suspicion on communication behaviors and perceptions.* Communication Theory, 1996. **6**(3): 243-267. doi:10.1111/j.1468-2885.1996.tb00128.x.

[6] Daft, R.L., R.H. Lengel, and L.K. Trevino. *Message equivocality, media selection, and manager performance: Implications for information systems.* MIS Quarterly, 1987. **11**(3): 355-366.

[7] DePaulo, B.M., D.A. Kashy, S.E. Kirkendol, M.M. Wyer, and J.A. Epstein. *Lying in everyday life.* Journal of Personality and Social Pyschology, 1996. **70**(5): 979-995. doi:0022-3514/96.

[8] 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.

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

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

[11] Granhag, P.A. and L.A. Strömwall. *Repeated interrogations: Verbal and non-verbal cues to*

deception. Applied Cognitive Psychology, 2002. **16**(3): 243-257. doi:10.1002/acp.784.

[12] Grice, P. *Further notes on logic and conversation*, in *Studies in the way of words*. 1989. Harvard University Press: Cambridge, MA. 41-57.

[13] Hancock, J.T., J. Birnholtz, N. Bazarova, J. Guillory, J. Perlin, and B. Amos. *Butler lies: Awareness, deception and design*. in *SIGCHI Conference on Human Factors in Computing Systems*. 2009. Boston, MA: ACM, 517-526. doi:10.1145/1518701.1518782.

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

[15] Hancock, J.T., J. Thom-Santelli, and T. Ritchie. *Deception and design: The impact of communication technologies on lying behavior*. in *SIGCHI Conference on Human Factors in Computing Systems*. 2004. ACM, 129-134. doi:10.1145/985692.985709.

[16] Hancock, J.T., C. Toma, and N. Ellison. *The truth about lying in online dating profile*. in *SIGCHI Conference on Human Factors in Computing Systems*. 2007. San Jose: ACM, 449-452. doi:10.1145/1240624.1240697.

[17] Hoffman, E., K. McCabe, and V.L. Smith. *Social distance and other-regarding behavior in dictator games.* The American Economic Review, 1996. **86**(3): 653-660.

[18] McCornack, S.A. and H.S. Park. *Deception detection accuracy in dating relationships: The other side of trust*, in *Communication Yearbook 9*, McLaughlin, M.L.e. 1986. Sage Publication: Beverly Hills, CA.

[19] Miller, G.R., M.A. Deturck, and P.J. Kalbfleisch. *Self-monitoring, rehearsal, and deceptive communication.* Human Communication Research, 1983. **10**(1): 97-117. doi:10.1111/j.1468-2958.1983.tb00006.x.

[20] Miller, G.R. and J.B. Stiff, ***Deceptive communication***. Vol. 14. 1993. Newbury Park, CA: Sage.

[21] Newman, M.L., J.W. Pennebaker, D.S. Berry, and J.M. Richard. *Lying words: Predicting deception from linguistic styles.* Personal Social Psychology Bulletin, 2003. **29**(5): 665-675. doi:10.1177/0146167203251529.

[22] Ott, M., Y. Choi, C. Cardie, and J.T. Hancock. *Finding deceptive online spam by any stretch of the imagination*. in *HLT'11*. 2011. Portland, Oregon: Association for Computational Linguistics, 309-319.

[23] 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.

[24] Stiff, J.B. *Theoretical approaches to the study of deceptive communication: Comments on interpersonal deception theory.* Communication Theory, 1996. **6**(3): 289-296. doi:10.1111/j.1468-2885.1996.tb00130.x.

[25] Trevino, L.K., R.H. Lengel, and R.L. Daft. *Media symbolism, media richness and media choice in organizations.* Communication Research, 1987. **14**(5): 553-574. doi:10.1177/009365087014005006.

[26] van Swol, L.M., M.T. Braun, and D. Malhotra. *Evidence for the Pinocchio Effect: Linguistic differences between lies, deception by omissions, and truths.* Discourse Processes, 2012. **49**(2): 79-106. doi:10.1080/0163853X.2011.633331.

[27] Whitty, M.T., T. Buchanan, A.N. Joinson, and A. Meredith. *Not all lies are spontaneous: An examination of deception across different modes of communication.* Journal of the American Society for Information Science and Technology, 2012. **63**(1): 208-216. doi:10.1002/asi.21648.

[28] Zhou, L., J.K. Burgoon, and D.P. Twitchell. *A longtudinal analysis of language behavior of deception in email*. in *IEEE Intelligence and Security Informatics Conferences (ISI)*. 2003. Springer-Verlag Berlin Heidelberg, 102-110. doi:10.1007/3-540-44853-5_8.

[29] 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): 139-166.

[30] 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 Annual Hawaii International Conference on Systems Sciences (HICSS)*. 2003. Hawaii: IEEE. doi:10.1109/HICSS.2003.1173793.

[31] Zhou, L. and D. Zhang. *Can online behavior unveil a deceiver?* in *Proceedings of the 37th Annual Hawaii International Conference on Systems Sciences (HICSS)*. 2004. Hilton Waikoloa Village Big Island, Hawaii: IEEE Press. doi:10.1109/HICSS.2004.1265079.

[32] Zhou, L. and D. Zhang. *Following linguistic footprints: Automatic deception detection in online communication.* Communications of the ACM 2008. **51**(9): 119-122. doi:10.1145/1378727.1389972.