Liar, Liar, IM on Fire: Deceptive Language-Action Cues in Spontaneous Online Communication

Shuyuan Mary Ho
School of Information
Florida State University
Tallahassee, FL 32306-2100
smho@fsu.edu

Cheryl Booth
School of Information
Florida State University
Tallahassee, FL 32306
clb14h@my.fsu.edu

Shashanka S. Timmarajus
Computer Science
Florida State University
Tallahassee, FL 32306
st13f@my.fsu.edu

Jeffrey T. Hancock
Information Science
Cornell University
Ithaca, NY 14853
jeff.hancock@cornell.edu

Xiuwen Liu
Computer Science
Florida State University
Tallahassee, FL 32306
liux@cs.fsu.edu

Mike Burmester
Computer Science
Florida State University
Tallahassee, FL 32306
burmeste@cs.fsu.edu

***Abstract*—With an increasing number of online users, the potential danger of online deception grows accordingly – as does the importance of better understanding human behavior online to mitigate these risks. One critical element to address such online threat is to identify intentional deception in spontaneous online communication. For this study, we designed an interactive online game that creates player scenarios to encourage deception. Data was collected and analyzed in October 2014 to identify certain deceptive cues. Players' interactive dialogue was analyzed using linear regression analysis. The results reveal that certain language features are highly significant predictors of deception in synchronous, spontaneous online communication.**

Keywords—interpersonal deception theory; computer-mediated communication; language-action features; regression analysis

I. INTRODUCTION

Communication technology has continually fostered human communications by providing ever newer and more convenient online interfaces. However, while the adoption of these technologies allows for greater and more convenient global communications, it also emphasizes human vulnerabilities, not the least of these being the vulnerability to deception. Human communication is based on certain presumptive norms, including in particular a presumption of truthfulness. Yet, notwithstanding this presumption, deception has, in fact, been found to occur rather frequently, and can be expected in at least one quarter of all conversations [1].

This paper presents initial findings in an ongoing study of language-action cues in detecting deception in a computer-mediated communication (CMC) environment. Here, "Language-action cues" refer to linguistic styles, phrases, patterns, or actions in an actor's written expression and manifested as an indirect or subtle signal to other actors. Specifically, these are cues we found to be present in spontaneous online deception as carried out in synchronous, co-present, distributed and recordable media. Since interpersonal communication involves a dynamic exchange of messages between two or more people, the study focuses on social interactions where participants try to mutually influence each other. We attempt to answer this question: *What language-action cues can be attributed to intentional deception in spontaneous computer mediated communication across a pluralistic background of users*?

II. ONLINE DECEPTION

Certain norms or assumptions are implicit in an exchange of messages; in particular, communication is based on an essential presumption of truthfulness (a "truth bias") [1]. This truth bias functionally minimizes the incentive to be suspicious regarding potential cues associated with deception.

A. The Problem of CMC Deception

Much study has been done to date concerning different aspects of deception, in both face-to-face (F2F) and CMC. These studies have a common purpose; attempting to identify cues (behavioral, contextual, verbal or textual) associated with deception. From these studies, we can derive several essential elements of online deception. First, deception is a volitional and intentional act, not accidental as in the result of a mistake. Second, there are two types of deception: spontaneous ("on the fly") and planned. Third, a lie may be de minimus ("little white lie") or serious (having significant consequences), or fall somewhere in between. Finally, the mode of communication used by a deceptive actor may be synchronous [2] or asynchronous, and there is a relationship between this and the type of media chosen for a deceptive act.

The following section discusses some of the theoretical frameworks in the area of online deception and deception in general, as applied to the specific research question.

B. Theoretical Approaches

Buller and Burgoon proposed [1] interpersonal deception theory (IDT), which focuses on the interaction between participants during the act of communication and emphasizes the interpersonal nature of communication. IDT posits that deception is an iterative and strategic process on the part of all parties: one party's behaviors influence or affect the responsive behaviors of the other, and communication strategies may need

to be adjusted "on the fly" for the deceiver to avoid detection. The language choices of a deceiver's message would reflect such strategic attempts to manipulate information and shape behavior.

Social distance theory provides another framework for thinking about and evaluating different cues as predictors of deception. DePaulo et al. [3] posited that in order to avoid the social discomfort associated with lying, deceivers will separate or distance themselves from the person to whom they are lying. A third framework, media richness theory [4] looks broadly at media choice as an indicator of deception. Media richness theory posits that because lying is subject to the interpretation of the individual being lied to, deceivers will tend to choose media that allows for multiple cues, immediate feedback and an opportunity for personalization, presumably in an attempt to obfuscate their lie by, for example, sending conflicting cues.

Finally, the feature-based model proposed by Hancock et al. [5], views deception cues through the lens of the specific features of the particular media used in deception. This model assumes fundamentally that lies are spontaneous, which suggests that deceptive actors will tend to lie using media that are synchronous, distributed but non-recordable. These theories inform our understanding of online deception and provide a solid theoretical and evidentiary means of analysis for the detection of deception.

C. Deceptive Language-Action Features

Our ability to detect deception, in any environment, depends on numerous factors, particularly the availability of certain types of cues alerting the receiver to be more critical of the information being provided. In a CMC environment, the availability of such cues is reduced (being limited to the text in message-based exchanges), relative to F2F communication. Nonetheless, certain communication cues, verbal and non-verbal, can still be observed and catalogued [6].

Examples of verbal deception cues in CMC, which look at words and syntax, include use of first-person references, emotional words, negations, inhibition words, prepositions, and conjunctions. Verbal cues also include syntactic cues such as overall "wordiness" of the communication and relevancy or meaningfulness of information provided, quantity of and consistency in detail, frequency of use of sensory or spatiotemporal words, and expansiveness and complexity of vocabulary used, and overall linguistic style (casual or formal). In addition, deception cues include those associated with verbal and non-verbal immediacy (i.e. ways in which an actor can associate or distance him/herself from the content of his/her message).

Our research is different from previous work in this area in several ways, but most particularly in that 1) the game-based approach we employ offers both an opportunity and *a motivation* for the actors to deceive, and 2) it specifically examines linguistic cues to planned deception, specifically in synchronous, co-present, distributed and recordable media.

III. A Sociotechnical Research Design

An interactive game interface was developed based on to the chat feature of Google+ Hangout. We presented players with interactive scenarios requiring them to write either deceptive or truthful statements. This online game ("*Truth or Dare*") simulates a real-time interactive deception scenario through synchronous communication channels. Each game involves two participants, who are placed in assigned pairings by the research team, and are then randomly assigned an *outer role* as either an *initiator* or a *detector* in each gaming session. The initiator in each scenario is also randomly assigned an *inner* role – either *saint* (truthful) or *sinner* (deceptive). The initiator establishes the *ground truth* before the beginning of each scenario by truthfully answering a question on a particular topic. This provides a baseline against which to assess the truthfulness or deceptiveness of his/her subsequent responses to the questions posed during the scenario by the detector, who must decide whether he/she feels the speaker is being truthful or deceptive in the scenario based on these exchanges.

IV. Data Collection and Analysis

Our dataset was collected during Fall 2014. The accuracy of the detector's predictions can be checked and verified against the ground truth as recorded in the MySQL database. The original sample size of those recruited during the Fall of 2014 was 24 participants, for a total of 12 game sessions, however some of these game sessions were incomplete and so had to be excluded from analysis. The final dataset captured 7 game sessions played by 14 participants with approximately 1,400 lines of script. The duration of each session varied from 30 minutes to 50 minutes.

A. Regression Analysis

The linguistic cues were extracted using Linguistic Inquiry and Word Count (LIWC), and were tested under linear regression analysis to define the relationship between the variables. Table 1 illustrates the two best linear regression models where the role of the initiator (i.e. saint or sinner) was the independent variable, and the linguistic cues were the dependent variables. The primary correlated predictors of deception are discussed below.

Table 1. Regression models.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.917[a]	.841	.655	.302	.841	4.528	7	6	.042
2	.941[b]	.885	.701	.281	.885	4.806	8	5	.050

a. Predictors: (Constant), assent, excl, I, negate, cogmech, posemo, affect
b. Predictors: (Constant), affect, excl, negate, quant, assent, cogmech, I, posemo

B. Manifested Language-Action Cues

Self References: Usage of the first person personal pronoun "I" by deceptive actors was statistically significant. Deceptive actors used more "I" references than truthful actors. In our study, because participants were engaged in anonymous interactions, there would be no inherent need to feel a desire to protect self-image, or protect the "other."

Negations (negate): Words such as "no," "not" and "never" were used more frequently by truthful actors than by deceptive actors.

Cognitive Process (Cogmech): Our results suggest that words reflecting cognitive process (i.e., active thinking), such as "cause", "know" and "ought" were used more freqently by deceptive actors than by truthful actors. While it appears that few studies have taken a measure of this LIWC category as a whole, Hancock et al. [7] found that deceivers (specifically, motivated deceivers) often use fewer causal words. Similarly, Newman et al. [8] examined the subcategory "exclusivity" (i.e. using words such as "but", "except" and "without") – and found that deceptive actors use fewer of these types of words than truthful actors. We again speculate that the apparent contradiction of our results is due to the types of deception participants engaged in.

Affective Processes (Affect): Usage of words expressing positive emotion (posemo) (e.g. "love", "nice", and "sweet") and words expressing negative emotion (negemo) (e.g. "hurt", "ugly" and "nasty") were measured separately. However, as neither posemo nor negemo word usage was found to be statistically significant individually, both of these were combined into the larger LIWC heading of "affect", which did yield a statisfically significant combined result. These results indicate that deceptive actors tended to use more "affect" words (i.e. words showing posemo and/or negemo) than truthful actors.

V. Implications and Limitations

The findings derived from this admittedly small sample seem on balance to support findings in similar studies; that deceivers tend to use more words overall in their communications than truth tellers [7]: the mean of the deceivers' word count was 72 words, while 55 words were used by truth tellers across all sessions. Our results indicate that descriptive and embellishing phrases were more often used in deceptive communications than in truthful ones and, overall, deceptive communications were longer (used more words) than truthful communications. Additionally, our results seem to generally indicate that, as predicted by media richness theory and the feature-based model, deceptive actors tend to use language that shortens the social distance between themselves and the person with whom they are communicating. 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.

VI. Conclusions and Future Work

The aforementioned results suggest that the deceptive linguistic cues used for transcript and online profile description can be significantly different from the deceptive linguistic cues used in spontaneous communication. Deception is not only a strategic process [1], that involves persuasive activities, but also is context-sensitive. Deception strategies differ based on modes of communication; asychronous vs. synchronous. By identifying key text-based cues in a given dataset, it may be possible to use these correlated communication patterns to develop predictive models of human interactive behavior online and thereby predict the truthful or deceptive nature of statements in spontaneous computer-mediated communication. The methodology employed permits further generalization of the perceptions in spontaneous online conversations to automate deception intelligence in learning machines.

Future research will include using 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 mentioned above. The ultimate objective is eventually to design and implement a "live" machine learning system that is able to detect the potential for deception in CMC environments.

Acknowledgment

The authors wish to thank the National Science Foundation EAGER grants #1347113 and #1347120, 09/01/13–08/31/15, and the Florida State University Council for Research and Creativity Planning Grant #034138, 12/01/13–12/12/14.

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] 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.

[3] 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.

[4] 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.

[5] Hancock, J.T., J. Thom-Santelli, and T. Ritchie. *Deception and design: The impact of communication technologies on lying behavior*. 2004. 130-136.

[6] Hancock, J., J. Birnholtz, N. Bazarova, J. Guillory, J. Perlin, and B. Amos. *Butler lies: Awareness, deception and design*. 2009. Boston: ACM, 517-526.

[7] 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**: 1-23.

[8] 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.