Computer-Mediated Deception: Collective Language-Action Cues as Stigmergic Signals for Computational Intelligence

Shuyuan Mary Ho
Florida State University
smho@fsu.edu

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

Abstract

Collective intelligence is easily observable in group-based or interpersonal pairwise interaction, and is enabled by environment-mediated stigmergic signals. Based on innate ability, human sensors not only sense and coordinate, but also tend to solve problems through these signals. This paper argues the efficacy of computational intelligence for adopting the collective language-action cues of human intelligence as stigmergic signals to differentiate deception. A study was conducted in synchronous computer-mediated communication environment with a dataset collected from 2014 to 2015. An online game was developed to examine the accuracy of certain language-action cues (signs), deceptive actors (agents) during pairwise interaction (environment). The result of a logistic regression analysis demonstrates the computational efficacy of collective language-action cues in differentiating and sensing deception in spontaneous communication. This study contributes to the computational modeling in adapting human intelligence as a base to attribute computer-mediated deception.

1. Introduction

Pairwise interaction can facilitate our cognitive understanding of each other, and language provides the means for effective communication. Differences in context, differences among the parties involved, differences in time, place, and even communication medium—each influence not only what is communicated but also how communication is perceived. For example, a humorously sarcastic remark about the economy made to friends while socializing on Friday evening may very well be entirely inappropriate were it to be made at work during a budget meeting on Monday morning. Likewise, taking the words from one conversation and applying them in another can imbue the words with an entirely different meaning to the original—and presumably intended—meaning. In the context of computer-mediated communication, language-action cues can become important indicators in identifying computer-mediated deception [1]. Cues found in language help enable collective understanding, and thus provide communication context for collective intelligence. In this regard, communicative signals shared by communicators are also critically important in assisting deceptive intent in computer-mediated communication (CMC).

As CMC users are currently being exposed to an increasing number and variety of risks associated with computer-mediated deception (e.g., unauthorized access to one's personally identifiable information, identity theft, as well as spear phishing attacks), it is becoming more and more important—and challenging—for users to be able to protect themselves from these deceptive tricks. Our ability to understand a message and thereby correctly interpret a sender's communicated intent and meaning becomes ever important. In face-to-face (F2F) communication, the receiver's assessment of context is informed not only by words, but also by other physical cues such as body language and facial expressions. In "cue lean" text-based CMC, the message receiver only has reference to the message sender's words themselves. Consequently, the intelligence generated in human interaction becomes one of the few means available to assess the intent of the individual(s) with whom one is communicating, as well as the truthfulness or reliability of the information being exchanged.

Human sensors not only have the intuitive ability to coordinate, but also the innate capability to sense and respond to anomalies. In this paper, we first describe the concept of collective intelligence; specifically, the human sensor's ability to detect deception in an interaction network is emphasized. We then review deceptive communication styles, which forms the core components of stigmergy;

URI: http://hdl.handle.net/10125/50098
ISBN: 978-0-9981331-1-9

deceivers (*agents*), cues (*signs*) across different contexts (*environment*). Next, we discuss a study conducted to investigate the stigmergy in the pairwise communication context. We compare and contrast the findings of this examination into communication cues. The paper concludes with some reflections, implications, and limitations, along with potential directions for future research.

2. Collective Intelligence

Collective intelligence refers to complex behavior created by the simple interaction between individuals that follow basic rules, and is generally defined as "the ability of a group to solve more problems than its individual members" [2]. A simple example can be the ants' or bees' ability to map out their environment. Individually, these insects experience limited capacity in processing information; however, collectively, they can decide the fields to which they exploit and the danger about to occur. The collective cognition is demonstrated and communicated through a "stigmergic signal [2]," or an innate "gene [3]," in these insects' society. In human society, people interact more autonomously due to different communication modes and medium. Nonetheless, people can work and coordinate together to solve complex problems in a surprisingly intelligent way. Malone, Laubacher *et al.* [3] illustrated the open source software development community as a prototypical example of collective intelligence.

The genes of collective intelligence can create, decide and build a genome of any shape and kind by asking simple questions of Who, What, Where, Why and How [3]. Likewise, the genes can also sense differences and anomalies in coordination activities. Individuals pick up communication cues, find information, and make their own decisions that then influence the group's intelligence and the results of collective problem solving. Although, a group's decisions can be influenced by the individuals' decisions and sharing of information, Woolley, Chabris *et al.* [4] examined the collective intelligence factor, and identified that this "c factor" does not depend on the average individual intelligence, but depends more on the average social sensitivity of group members, which reflects the composition of the group (e.g., equality in distribution of conversational turn-taking), and the way the group members interact (e.g., communication mode and medium) when they are assembled.

Based on the observations of the insect swarm-behavior, Dipple, Raymond *et al.* [5] proposed a macro-level view of the communication mechanism that triggers responses in human society. Depending on environment-mediated signals, this model defines an abstract form of stigma semantics in stigmergy with the core components: the agents, the environment, and the sign(s). The agent's ability to coordinate, to sense, or to detect anomalies depends on their interpretation of the meaning as mediated by the manifestation of stigmergic signals and signs when interacting. These core components also correspond to the fundamentals of human sensors' ability to interpret and sense deceptive communication.

3. Deceptive Communication

A common focus of the numerous studies on deceptive communication is to identify particular cues (behavioral, contextual, verbal or textual) that can be associated with deception. Collectively, these studies reveal several essential aspects of deception. First, deception is "…a message knowingly transmitted by a sender to foster a false belief or conclusion by the receiver" [6]. Thus, deception is a volitional and intentional act. Simple "mistakes of fact" would not constitute deception. Neither would it be considered deception when a message sender objectively communicates false information when s/he believed the information to be true [7]. In addition, research suggests that both the mode of communication (synchronous [8] or asynchronous [9]) and the specific medium chosen may provide insight into to the type (planned or on-the-fly) and severity (serious or inconsequential) of the deception [10]. Finally, our ability to detect deception, in any environment, depends on many different factors, including the availability of certain types of cues. These cues can function as an alert to the receiver to be more critical of the information being provided. Unfortunately, in a CMC environment, the availability of cues is reduced (being limited to the text in message-based exchanges) when compared to F2F communication, thereby making detection of deception particularly challenging in CMC.

3.1. Deceptive Agents

Communication is, of necessity, interactive. It involves a sender and one or more receiver(s) who are engaged in a (more-or-less) interactive exchange. Within this exchange, there is an opportunity for the message sender to influence the receiver(s) actions or beliefs. Deceptive communication fundamentally means purposefully misrepresenting or concealing the truth, either by omission or commission [11].

Miller, Deturck *et al.* [12] described deceptive communication as "…a general persuasive strategy

that aims at influencing the beliefs, attitudes and behaviors of others by means of deliberate message distortions" (p. 99). Miller and Stiff [13] and Stiff [14] characterized deceptive communication as an act involving the intentional use of persuasive strategies and activities to manipulate the receiver. Buller and Burgoon's [6] interpersonal deception theory (IDT) further examines and explains how a deceiver (i.e., sender of a deceptive communication) strategically shapes his/her communication behaviors by studying the perceptions and suspicions of the receiver(s). As a major theoretical lens, IDT views deceptive communication as a strategic, interactive process on the part of all parties, through which the deceiver attempts to accomplish multiple objectives—including impression management, emotion management and conversational management. IDT suggests that, much like the move-and-counter-move dynamics of a chess match, the influence of the deceiver's behavior on the receiver affects the receiver's behavior, which, in turn, affects the deceiver's strategy and behavior.

There are two critical truths about deception one must appreciate. First, deception is common, occurring in approximately one–quarter of all communications [6], and second, if we charge humans to detect deception, they tend to be bad at it [1, 15].

3.2. Language-action Cues and Signs

3.2.1. F2F. F2F communication has an "advantage" over CMC in terms of deception detection, in that both verbal and nonverbal (i.e., physical) cues are available to the message receiver. Indeed, physical nonverbal cues are more-or-less exclusive to F2F, and include everything from body language and facial expressions to the tone and pitch of voice and pace of speech. Ekman and Friesen [11] specifically studied such nonverbal/ physical communication behaviors 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*. Granhag and Strömwall [16] likewise examined deception in a F2F context, using a credibility assessment technique, statement validity analysis (SVA), to evaluate both verbal and nonverbal behaviors during an in-person F2F interview event. Their findings indicate that speech rate, pauses, gaze aversion, and smiles/laughs were all salient and statistically significant nonverbal cues to deception. While physical-based cues (e.g., body language, facial expressions, and even vocal pitch and tone and pace of speech) are virtually non-existent in text-based CMC environment, other verbal and nonverbal cues are indeed shared between CMC and F2F communication.

3.2.2. CMC. Text-based CMC is "cue lean," in that it lacks the physical cues to deception available in F2F communication. However, certain communication cues can nonetheless still be observed and catalogued within a CMC environment [17-19]. These communication features and language-action cues, such as first-person references, emotion words, inhibition words, prepositions, and conjunctions, have all been shown to be indicators that can differentiate deceivers from truth tellers [17]. Use of more or fewer sensory or spatiotemporal words, and changes in the diversity and complexity of language have also been shown to be indicative [20]. And, as in F2F, level of detail (less or more) is also suggestive of deception in CMC—although in CMC, relevance of detail appears to be more significant than "detail" *per se*. That is, deceivers in CMC tend to be wordier than truth-tellers, but the additional words (i.e., details) provided are not necessarily relevant or meaningful [19].

Another language-action cue that is important in CMC (also in F2F communication) is *immediacy* (i.e., ways in which a speaker can associate, or distance him/herself from the content of his/her message) [8]. *Immediacy* (whether verbal or non-verbal) is particularly important in detecting deception. In the physical environment, nonverbal immediacy cues include eye contact, body language, facial expression, etc. While these specific cues were first studied in a F2F environment, certain cues—such as delay in response—are also present in CMC, and operate similarly in both environments to create a psychological distance between deceiver and his/her communication partner [10].

It is worth noting that, in contrast to F2F interactions, CMC deceivers statistically tend to take shorter *pauses* between messages (i.e., time between two consecutive messages sent by them) than truth-tellers [19]. Deceivers also have been found to have shorter *response latency* (i.e., time between receiving a message and responding to it) than truth-tellers, which is consistent with results of studies investigating response latency in F2F communication [19]. Research has also suggested that deceivers tend to use more restricted vocabulary and syntax, and to be more casual in their linguistic style [12].

Many of these cues have been examined for the purpose of developing an automated process to detect deceptive intent, and many of them—including quantity and consistency of detail—are measureable

in a dynamic exchange of text messages by focusing on specific features such as the use of adverbs, adjectives, and inclusive words. It is thus possible to benchmark verbal indicators (such as word count and details of information disclosed) and capture certain nonverbal behaviors (latency and usage of expression words) in CMC environment, which can then be statistically computed [19].

3.3. Environment

Identification of deception is a complex problem that often requires ground truth verification [11, 15]; nevertheless, deception can be detected in interpersonal communication [1, 21-23] as well as group communication [24, 25]. Research into CMC deception cues has examined a variety of media types and modes of communication, while also exploring the role of media choice and mode of communication. The mode and medium of the communication provide dependencies that shape people's communication behavior. Here, communication mode refers to whether the parties are interacting in real time (synchronous) or are communicating via messages exchanged back-and-forth over time (asynchronous). Moreover, the mode of any communication—whether it takes place in a "virtual" CMC environment, or in a "real" F2F environment—can also influence how one interacts and communicates. One notable early study exploring this problem at a high level was done by Hancock, Thom-Santelli *et al.* [26]. Participants were asked to journal their interactions and lies for seven (7) consecutive days. The results were the foundation of their "feature-based" model for studying deceptive CMC, which attempts to derive cues to deceptive communication by examining the specific features of the medium chosen—particularly looking at the communication mode (i.e., synchronous, asynchronous, or either/both) associated with or supported by the medium. Other salient factors include whether the medium records the communication or not, and whether the communication is distributed. A fundamental assumption of this model is that deception is spontaneous, and therefore is more likely to occur when media is "synchronous and distributed, but non-recordable" [10].

In this section, we briefly discuss studies that focus on both asynchronous and synchronous media types and communication modes in interpersonal as well as in group context.

3.3.1 Asynchronous Communication. One of two main theoretical "schools of thought" coming out of, or supported by, this research—*social distance theory*—would seem most applicable when discussing asynchronous communication. According to social distance theory [27], the prevailing social disapproval of deception, and the accompanying psychological discomfort experienced by deceivers makes deceivers attempt to distance or separate themselves from their deception and the individual(s) they are attempting to deceive. Therefore, according to *social distance theory*, deceivers will tend to choose media offering fewer cues to the receiver of the communication and, thus, will be more likely to use an asynchronous mode of communication.

A notable study examining language-action cues captured from e-mails was conducted by Zhou, Twitchell *et al.* [28]. This study evaluated deceptive cues in a team-based "desert survival scenario" game, and found that deceivers tended to be wordier as compared to truth-tellers—particularly in terms of using more verbs, modifiers and noun phrases in peripheral expressions to provide useless or irrelevant information. Moreover, Zhou and Zhang [19] found that, in the context of asynchronous online communication, deceivers tend to be more active in language usage, and take shorter pauses between messages and were more non-immediate than truth-tellers (using more group references and modal verbs). Another study examining CMC deception in an asynchronous mode looked at the truthfulness of online dating profiles. Toma and Hancock [29] found that language-action cues involving emotion were "more powerful in predicting deception" than cognitive cues—with the notable exception that word count (a cognitive cue) was again highly significant. Likewise, Ott, Choi *et al.* [30] examined asynchronous CMC through the lens of online hotel reviews, investigating which linguistic features were most indicative of a truthful review. The results suggest that truthful reviews included more "sensorial and concrete language" (especially concerning spatial configurations—i.e. overall room space and space usage) than false or deceptive ones, while deceptive/ fake reviews included more superlatives. Language-action cues in transcripts of 911 (emergency) telephone calls has also been studied [31], to determine which cues are most indicative of bogus or fake calls. The results from this study show that deceptive callers were found to show more "inhibition" (meaning, for example, delaying or telling the dispatcher to "hold on a minute"). Deceptive callers were also found to use more words associated with immediacy (1st person pronoun) and non-immediacy (3rd person pronoun).

3.3.2. Synchronous Communication. A second major theoretical framework—*media richness theory*—seems most applicable to synchronous communication. According to media richness theory, much, if not most, deceptive communication is equivocal in nature (i.e., intentionally ambiguous), and thus is open to interpretation by the receiver (i.e., the individual the communicator is attempting to deceive). Therefore, deceptive actors will tend to choose media types that provide them multiple cues, an opportunity for personalization, and immediate feedback (i.e., synchronous and spontaneous)—allowing them to ensure the equivocal nature of their message and adjust their deceptive communication strategy 'on the fly,' and thereby obfuscate their deceptive intent [32, 33]. Four factors are used in determining the *richness* of the medium: feedback (spontaneous, 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 [32]). The richer the medium, the better able it is to convey equivocal messages.

Studies investigating CMC deception in a synchronous mode of communication have tended to focus on instant-messaging/ chat. For example, a study by Hancock, Curry *et al.* [34] found that word count was a significant predictor of deception in semi-synchronous communication. Participants in the dyad (consisting of one truthful and one deceptive partner) were given time (5 minutes) before play was to begin in which to review the fixed set of questions to be used, and plan their responses. Thus, the deceptive player was, in essence, given an opportunity to plan and prepare a strategy for implementing the intended deception(s). Further, Toma and Hancock [29] suggested that first person pronouns were used more by truth-tellers than deceivers, and that deceivers use fewer self-references in CMC synchronous chat, but more third-person references—consistent with social distance theory. To note, the differences of research methodology employed in the above-mentioned studies illustrates that *timing* is an important indicator for identifying computer-mediated deception. The collective representation of language-action cues varies depending on the *time* allocated to deceivers.

3.3.3. Group Communication. Although many studies have focused on interpersonal interactions in CMC deception, a growing number of studies in this area have also examined CMC deception from a group-dynamics perspective. Taylor, Dando *et al.* [35], for example, examined language-action cues (specifically, personal pronouns, negative emotions, feelings, cognitive processes, discrepancy and tentative) in the context of deception through (asynchronous) e-mail exchanges, within and between teams in a common physical location. Taylor, Dando *et al.*'s [35] findings were consistent with findings from Toma and Hancock [29], [36], which indicate that the deceptive "insiders" used more personal pronouns than the others in their group. Additionally, Taylor, Dando *et al.* [35] found that the designated deceivers used more words associated with cognitive processes (particularly, discrepancy and tentative). This finding is contrary to the findings of Hancock, Toma *et al.* [18] which suggested that affect (emotion)-related words were more significant in detecting deception in online dating profiles. While both studies involved asynchronous communication, Taylor, Dando *et al.*'s [35] study examined group communication and interaction, and thus provides a different context than Hancock, Toma *et al.*'s [18] interpersonal communication studies on deception in online dating profiles.

Ho, Hancock *et al.* [25], [24] examined behaviors of a deceptive "insider" in spontaneous synchronous chat-based group-dynamics context, and suggested that deceptive "insiders" in computer-mediated synchronous interactions will tend to use more words associated with cognitive processes. Ho, Hancock *et al.*'s [25] findings are consistent with those of Taylor, Dando *et al.* [35], and suggest that deceptive "insiders" tend to use words more associated with cognitive processes in their communication with their peers in either asynchronous or synchronous communication.

4. Method

Our study incorporates core components of the environment-mediated stigmergic signals by analyzing instances in pairwise interaction: deceivers (*agents*) and *truth-teller* (*agents*), collective language-action cues (*signs*) across different scenarios (*environment*) as context. The data collection and data cleaning process of this study are described in this section.

4.1. Data Collection

The data were collected in 2014 and 2015[1]. Each game session consists of two players; a speaker and a detector. Data were collected across a total of 80

[1] The Florida State University's Institutional Review Board has approved human subject data collection (Protocols #2014.13490 and #2015.15885).

game sessions. 40 participants (22 males and 18 females) were randomly assigned into pairs, with each pair playing a total of 4 game sessions. Players were between 18 and 68 years of age. Players' names were replaced with pseudo-names to protect their privacy.

Each game session lasted approximately 30 minutes, and consisted of about 4 role-play exchanges. At the end of each such exchange, the players' roles were automatically changed.

4.2. Data Cleaning Process

The collected data 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. Any individual message containing fewer than 50 words total was excluded from the dataset because these messages do not contain meaningful sentences but gibberish words (e.g., yes, um, ok, etc.). The final data set used in analysis consisted of a total of 2,196 lines of chat and 7,271 words.

5. Research Design and Data Analysis

This study investigates the collective language-action cues as stigmergic signals that is most indicative of deceptive intent in interpersonal deception, specifically in spontaneous synchronous chat-based CMC (*environment*). The research approach focuses on developing specific metrics for collective language-action cues as represented by information behavior (*signs*), and analyzes communication patterns that distinguish between deceptive vs truthful actors (*agents*). This online game 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.

5.1. Environment

An interactive online game, called "*Real or Spiel*[2]," was designed and developed to present players with real-time interactive simulated scenarios requiring them to exchange either deceptive or truthful statements specifically using instantaneous, synchronous communication channels [21]. Figure 1 is a screenshot taken from a live game, and includes role assignment and ground truth question.

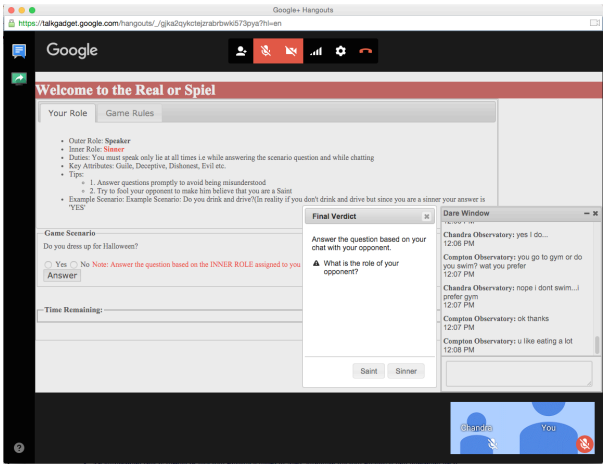

Figure 1. Game interface [1]

5.2. Agents

Each game scenario involves two participants (i.e., players), who are placed in randomly assigned pairings by the research team, and 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)—by the RAND() random operator. The speaker establishes the *ground truth* before 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 visited Puerto Rico?" If, for example, the speaker has visited the location, s/he would establish the ground truth by answering "yes." This provides a baseline against the assessment of 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 visited the above-mentioned location, and the speaker would try to convince the detector that s/he has not visited such a location in the past. 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.

5.3. Language-action Cues and Signs

The linguistic cues were extracted according to the categories established in the Linguistic Inquiry and

[2] Developed at Florida State University.

Word Count (LIWC) tool [20, 37] after the data had been cleaned. The specific LIWC categories we included in our analysis are those set forth in Figure 1. In order to avoid or minimize possible multicollinearity problems, cues directly corresponding to the "main" LIWC headings of "cogmech" and "affect" were ultimately excluded, in favor of including cues from their respective subcategories (i.e., posemo, negemo, certain, incl, excl, etc.).

Table 1. Language-action cues extracted by LIWC

LIWC Categories	CODING SCHEMA	Examples
Affective Process	affect	happy, cried, abandon
Positive Emotion	posemo	love, nice, sweet
Negative Emotion	negemo	hurt, ugly, nasty
Cognitive Process	cogmech	cause, know, ought
Certainty	certain	always, never
Inclusive	incl	and, with, include
Exclusive	excl	but, without, exclude
Discrepancy	discrep	should, would, could
Insight	insight	think, know, consider
Causation	cause	because, effect, since
Negations	negate	no, not, never
Pronouns	pronoun	
1st person singular	self-reference	I, me, myself
2nd person	other reference	you
Quantifiers	quant	few, many, much
Social	social	family, friends, humans
Tentative	tentat	maybe, perhaps, guess
Word Count	WC	n/a

We analyzed the final dataset using logistic regression, with the dichotomous outcome variable "Deceiver" (0=truthful/ 1=deceiver). Our independent/ predictor variables were the LIWC categories illustrated in Table 1, plus 'time-lag.'

Table 2. Fitness of the model

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	35.847	16	.003
	Block	35.847	16	.003
	Model	35.847	16	.003

The model incorporating these variables was statistically significant (p=0.003), as indicated in Table 2. The specific variables and corresponding statistical significance are set out in Table 3. In this model, we found that two language-action cues were statistically significant: Word Count (p=0.008) and Insight (p=0.02). However, the rest of the language-action cues, individually, were not statistically significant in predicting or identifying potential deception.

Nonetheless, logistic regression analysis indicates the model itself is significant with a Chi-square of χ^2=35.8, p<0.01 (Table 2). While multicollinearity may provide one possible explanation for this apparently contradictory result (i.e., non-significant predictors, but strong overall model), we had already eliminated any variables that might have created such a problem. Instead, in this case, we attribute this phenomenon to the nature of communication: the context itself. It is the context—the combination of words—that is most indicative of deception, rather than the words alone. Thus, even if individual language-action cues themselves are not significant, a particular *combination* of language-action cues may well be significant in the ability to indicate deception.

Table 3. Categorical variables

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	WC	-.016	.006	7.125	1	.008	.984
	I	-.089	.099	.803	1	.370	.915
	you	-.097	.154	.391	1	.532	.908
	negate	.074	.084	.768	1	.381	1.076
	quant	.261	.170	2.350	1	.125	1.298
	social	-.060	.124	.232	1	.630	.942
	posemo	.157	.106	2.182	1	.140	1.170
	negemo	.438	.244	3.220	1	.073	1.550
	insight	.424	.183	5.373	1	.020	1.528
	cause	.090	.222	.166	1	.684	1.095
	discrep	.167	.253	.437	1	.509	1.182
	tentat	-.086	.113	.572	1	.450	.918
	certain	.082	.174	.219	1	.639	1.085
	incl	-.144	.151	.905	1	.341	.866
	excl	-.110	.112	.956	1	.328	.896
	Time_lag	-.018	.024	.611	1	.434	.982
	Constant	.712	1.943	.134	1	.714	2.037

a. Variable(s) entered on step 1: WC, I, you, negate, quant, social, posemo, negemo, insight, cause, discrep, tentat, certain, incl, excl, Time_lag.

5.4. Results and Discussion

The initial logistic regression on this model was run with a (default) cut value of 0.5 (depicted in Table 4), and yielded an overall accuracy of 75% in correctly classifying "0s" (truth-tellers' statements) and "1s" (deceivers' statements).

However, because the focus of our study is on identifying deceivers (i.e., classification as a "1"), the accuracy of the model specifically in categorizing "1s" is equally important as overall accuracy. The classification table (Table 4) shows the model to be 75% accurate in correctly classifying both "1s" as deception and as "0s" as truthful statements (Table 4), as well as having 75% overall accuracy.

On initial review of the classification at 0.5 cut value would seem to be a fairly good model (Table 4). However, as the objective of our study is to identify the model that optimizes the combination of overall accuracy *and* accuracy as to detecting deceivers (i.e., "1s"), we ran two additional models, using different cut-values.

Table 4. Classification at 0.5 cutoff

Classification Table[a]

			Predicted		
			Deceiver		Percentage
	Observed		0	1	Correct
Step 1	Deceiver	0	30	10	75.0
		1	10	30	75.0
	Overall Percentage				75.0

a. The cut value is .500

The second iteration of the logistic regression analysis (depicted in Table 5) on this model was run with a cut value of 0.4. The model illustrated in Table 5 yielded an accuracy rate for classification of "1s" of 85% with a slightly lower overall accuracy of 74% classifying truth-teller compared to the 75% accuracy of Table 4.

Table 5. Classification at 0.4 cutoff

Classification Table[a]

			Predicted		
			Deceiver		Percentage
	Observed		0	1	Correct
Step 1	Deceiver	0	26	14	65.0
		1	7	33	82.5
	Overall Percentage				73.8

a. The cut value is .400

We further ran a third round of logistic regression analysis at a cut value at 0.6 (depicted in Table 6). This model yielded an accuracy rate in classifying "1s" of only 65%, although the model classifies truth-teller yielded 85% accuracy, and the overall model had an accuracy of 75% (which was not different from the results where the cut-off value was set to 0.5).

Table 6. Classification at 0.6 cutoff

Classification Table[a]

			Predicted		
			Deceiver		Percentage
	Observed		0	1	Correct
Step 1	Deceiver	0	34	6	85.0
		1	14	26	65.0
	Overall Percentage				75.0

a. The cut value is .600

To reiterate, the objective of our study is aimed at computationally identifying deceptive agents based on collective language-action cues in text-based communication. Comparing results derived from the three different cut values, we suggest the model with cut value at 0.4 as illustrated in Table 5 is the optimal combination of overall accuracy at 73.8% due to the high accuracy as to classifying deceivers at 82.5%.

6. Limitations and Future Work

One of the limitations involves with the adoption of the Google+ Hangout as the players' communication platform. For example, players frequently experienced technical problems in logging into the Google+ pseudo accounts created for the game, and launching the game interface we developed. It is our observation that these difficulties not only confused and distracted participants/ players, but also detracted from the overall amount of time they spent in the game itself, thus reducing our ability to collect more conversational data. To address this problem, our future work includes reconstructing the game on an independent/ stand-alone platform. We also plan to design and develop an automated participant assignment (i.e., pairing) system within the new platform, so that players are no longer manually paired.

We believe that research participation should be carried out with benefits of both learning and fun experience to participants. We thus plan to increase the strength of competitive aspect of the game by systematically reporting participants' guesses (i.e., correct or incorrect answers) during the game. By providing more feedback to the participants, it may help them make better decisions. At the same time, we can also observe how research participants make both deception decisions as well as detector's decisions in capturing the liar.

The final noteworthy limitation with respect to this study involves the sample size of the dataset, which we acknowledge is fairly small. We anticipate that in our future work, we will run the study to a broader audience of potential participants in order to get a larger dataset. Nonetheless, we submit that the results of the current study are still suggestive and encouraging.

7. Implications and Conclusions

Computer-mediated deception can be modeled based on the core components of stigmergic signals including the agents, the environment, and the signs of communicative intent. This study demonstrates the efficacy of modeling stigmergic signals to differentiate deceivers (*agents*) from truth-tellers (*agents*) based on collective language-action cues (*signs*) in synchronous pairwise interaction (*environment*). Our results demonstrate that, in the context of text-based synchronous CMC, it is the

overall combination of language-action cues (*signs*), rather than specific words used by deceivers to deceive, as most indicative of deception. Research exploring machine learning approach appears promising in detecting computer-mediated deception [23]. Moreover, the merit of our game design specifically emphasizes not just synchronous communication—but *spontaneity* within synchronous communication [22]. That is, our results provide computational intelligence in differentiating computer-mediated deception in synchronous spontaneous CMC, and the results inform the design of a crowd-sourced online polygraph system in the CMC context where F2F interaction is not available.

8. Acknowledgement

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. The authors acknowledge and appreciate the research efforts and contributions from Cheryl Booth, Sai Surya Shashanka Timmarajus, Kashyap Vemura, and Aravind Hariharan.

9. References

[1] Ho, S.M., J.T. Hancock, C. Booth, and X. Liu. *Computer-mediated deception: Strategies revealed by language-action cues in spontaneous communication.* Journal of Management Information Systems, 2016. **33**(2): 393-420. doi: 10.1080/07421222.2016.1205924.

[2] Heylighen, F. *Collective intelligence and its implementation on the Web: Algorithms to develop a collective mental map.* Journal of Computational & Mathematical Organization Theory, 1999. **5**(3): 253-280. doi: 10.1023/A:1009690407292.

[3] Malone, T.W., R. Laubacher, and C. Dellarocas. *The collective intelligence genome.* MIT Sloan Management Review, 2010. **51**(3): 21-31.

[4] Woolley, A.W., C.F. Chabris, A. Pentland, N. Hashmi, and T.W. Malone. *Evidence for a collective intelligence factor in the performance of human groups.* Science, 2010. **330**: 686-688. doi: 10.1126/science.1193147.

[5] Dipple, A., K. Raymond, and M. Docherty. *General theory of stigmergy: Modeling stigma semantics.* Cognitive Systems Research, 2014. **31-32**: 61-92. doi: 10.1016/j.cogsys.2014.02.002.

[6] 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.

[7] Vrij, A., ***Detecting lies and deceit: The Psychology of lying and the implications for professional practice***. 2000. New York, NY: John Wiley & Sons Ltd. 276. ISBN: 0-471-85316-X.

[8] 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.

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

[10] 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.

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

[12] 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.

[13] Miller, G.R. and J.B. Stiff, ***Deceptive communication***. Vol. 14. 1993. Newbury Park, CA: Sage. 142.

[14] 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.

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

[16] Granhag, P.A. and L.A. Strömwall. *Repeated interrogations: Verbal and non-verbal cues to deception.* Applied Cognitive Psychology, 2002. **16**(2002): 243-257. doi: 10.1002/acp.784.

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

[18] Hancock, J., C. Toma, and N. Ellison. *The truth about lying in online dating profile*. in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2007)*. 2007.

San Jose: ACM, 449-452. doi: 10.1145/1240624.1240697.
[19] Zhou, L. and D. Zhang. *Can online behavior unveil a deceiver?* in *Proceedings of the 2004 Annual Hawaii International Conference on Systems Sciences (HICSS-37)*. 2004. Hilton Waikoloa Village Big Island, Hawaii: IEEE Press. doi: 10.1109/HICSS.2004.1265079.
[20] 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.
[21] Ho, S.M., J.T. Hancock, C. Booth, X. Liu, M. Liu, S.S. Timmarajus, and M. Burmester. *Real or Spiel? A decision tree approach for automated detection of deceptive language-action cues*. in *Proceedings of the 2016 Hawaii International Conference on System Sciences (HICSS-49)*. 2016. Kauai, Hawaii: IEEE, 3706-3715. doi: 10.1109/HICSS.2016.462.
[22] 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.
[23] Ho, S.M., X. Liu, C. Booth, and A. Hariharan. *Saint or Sinner? Language-action cues for modeling deception using support vector machines*. in *Proceedings of the 2016 International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS'16), LNCS 9708*. 2016. Washington DC: Springer, 325-334. doi: 10.1007/978-3-319-39931-7_31.
[24] 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.
[25] Ho, S.M., J.T. Hancock, C. Booth, M. Burmester, X. Liu, and S.S. Timmarajus. *Demystifying insider threat: Language-action cues in group dynamics*. in *Proceedings of the 2016 Hawaii International Conference on System Sciences (HICSS-49)*. 2016. Kauai, Hawaii: IEEE, 2729-2738. doi: 10.1109/HICSS.2016.343.
[26] Hancock, J.T., J. Thom-Santelli, and T. Ritchie. *Deception and design: The impact of communication technologies on lying behavior*. 2004. 130-136.
[27] 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.
[28] 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 2003 36th Annual Hawaii International Conference on Systems Sciences (HICSS-36)*. 2003. Hawaii: IEEE. doi: 10.1109/HICSS.2003.1173793.
[29] Toma, C.L. and J.T. Hancock. *Reading between the lines: Linguistic cues to deeption in online dating profiles*. in *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW'10)*. 2010. Savanna, Georgia: ACM, 5-8. doi: 978-1-60558-795-0/10/02.
[30] Ott, M., Y. Choi, C. Cardie, and J.T. Hancock. *Finding deceptive online spam by any stretch of the imagination*. in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT'11)*. 2011. Portland, Oregon: Association for Computational Linguistics, 309-319.
[31] Burns, M.B. and K.C. Moffitt. *Automated deception detection of 911 call transcripts.* Security Informatics, 2014. **3**(8): 1-9. doi: 10.1186/s13388-014-0008-2.
[32] 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.
[33] 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.
[34] 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.
[35] 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.
[36] Toma, C.L. and J.T. Hancock. *What lies beneath: The linguistic traces of deception in online dating profiles.* Journal of Communication, 2012. **62**(1): 78-97. doi: 10.1111/j.1460-2466.2011.01619.x.
[37] 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.