Group Decision and Negotiation **13**: 81–106, 2004

Automating Linguistics-Based Cues for Detecting Deception in Text-based Asynchronous Computer-Mediated Communication

LINA ZHOU
Department of Information Systems, University of Maryland, Baltimore County, MD, USA
(E-mail: zhoul@umbc.edu)

JUDEE K. BURGOON, JAY F. NUNAMAKER, JR. AND DOUG TWITCHELL
Center for the Management of Information, University of Arizona, Tucson, AZ, USA
(E-mail: jburgoon@cmi.arizona.edu; jnunamaker@cmi.arizona.edu; dtwitchell@cmi.arizona.edu)

Abstract

The detection of deception is a promising but challenging task. A systematic discussion of automated Linguistics Based Cues (LBC) to deception has rarely been touched before. The experiment studied the effectiveness of automated LBC in the context of text-based asynchronous computer mediated communication (TA-CMC). Twenty-seven cues either extracted from the prior research or created for this study were clustered into nine linguistics constructs: quantity, diversity, complexity, specificity, expressivity, informality, affect, uncertainty, and non-immediacy. A test of the selected LBC in a simulated TA-CMC experiment showed that: (1) a systematic analysis of linguistic information could be useful in the detection of deception; (2) some existing LBC were effective as expected, while some others turned out in the opposite direction to the prediction of the prior research; and (3) some newly discovered linguistic constructs and their component LBC were helpful in differentiating deception from truth.

Key words: deception, deception detection, linguistics based cue, computer-mediated communication, natural language processing

1. Introduction

Deception generally entails messages and information knowingly transmitted to create a false conclusion (Buller and Burgoon 1994). It is a fact of life that daily communication is rife with various forms of deception, ranging from white lies, omissions, and evasions to bald-faced lies and misrepresentations. Driven by the globalization of economies and advancement of computer technology, computer-mediated communication (CMC) (Wolz et al. 1997) continues to diffuse into our everyday life, bringing with it new venues for deception. CMC can be classified into text, audio, audio/video, and multi-media based formats. Text-based CMC is conducted via transmitting textual information without audio and video signals. Such transmissions may differ in their timeliness of response. Synchronous CMC, such as Instant Messaging, has minimal time delays between message transmissions.

Asynchronous CMC, such as email, allows people to respond to incoming messages at their convenience. If CMC is both text-based and asynchronous, we call it Text-based Asynchronous CMC (TA-CMC). Given little consumption of network bandwidth and much flexibility in time, TA-CMC has gained wider adoption than other types of CMC. For example, there were over 890 million email accounts in the world by the end of 2000, up 67 percent from 1999 (Internet Society; WorldLingo). It is this ubiquitous form of CMC and specifically, the messages contained within the text body of such transmissions, which are the focus of the current investigation.

On the one hand, the ever-increasing volume of information transferred through the Internet simultaneously increases the chances of receiving deceptive messages while making it inefficient and impractical to manually filter and screen such messages. On the other hand, people tend to be truth-biased on assessing messages they receive so that the accuracy of human detection of deception remains little better than chance (Frank and Feeley 2002). Tools that augment human deception detection and thereby increase detection accuracy would therefore prove quite valuable, whether in the realm of low-stakes daily discourse or in high-stakes realms such as law enforcement, employment screening, and national security. By analyzing the messages produced by deceivers and comparing them to those produced by truth-tellers, we hope to verify a number of reliable indicators of deceit that can subsequently be built into software to automate detection. Natural language processing (NLP) is a research area that is intended to use computers to analyze and generate languages that humans use naturally. Some relatively mature NLP techniques enable software to automatically identify linguistics-based cues in texts.

Deception detection in TA-CMC has received little attention in research and field studies so far. The research on the automation of deception detection and on correlates of deception is largely separated. The contexts in which automatic deception detection has been investigated include credit card fraud (Wheeler and Aitken 2000), telecommunication fraud (Fawcett and Provost 1997), and network intrusion (Mukherjee et al. 1994), for example. These studies share the characteristic of structured original data with predefined attributes. Credit card fraud detection is a good example. A pre-determined group of attributes of each credit card transaction stored in a database is employed in discovering fraud patterns. Therefore, some conventional statistics and machine learning techniques, such as outlier detection (Aggarwal and Yu 2001) and case-based reasoning (Wheeler and Aitken 2000), can be directly applied to analyzing the data. However, the data produced in TA-CMC are free texts, which are much less manageable than the structured data due to the lack of standard composition style and common elements of messages. The natural language composition of textual messages adds more complexity and ambiguity to the task of analyzing such data. In order to extend the above-mentioned statistical and machine learning techniques to TA-CMC, we have to first transform messages into some kind of structured format. The structure should capture indicators of deception present in the messages.

As for reliable indicators of deceit, there have been numerous studies examining physiological responses, utilizing behavioral coding with well-trained experts, or applying content-based criteria to written transcripts. In virtually all of these cases, the cues or criteria have been developed for well-trained experts and would be quite difficult for laypersons

to apply. When computerization was applied, coding of a set of cues was performed manually, and the function of the computer was limited to performing statistical analysis on the coded scores (Akehurst et al. 1995; Höfer et al. 1996; Köhnken et al. 1995; Ruby and Brigham 1998; Sporer 1997; Vrij et al. 2000). Our study goes one step closer towards the ultimate goal of automating deception detection by replacing humans with computers in analyzing and rating messages based on promising cues. The indicators, or cues, that discriminate between truthful and deceptive messages can then be used to build profiles and algorithms for the automated detection of deception.

Extant literature in communication, criminology, forensic science, police study, psychology, and psychophysiology offers numerous prospective cues that might be applicable to TA-CMC (Burgoon et al. 1996; Driscoll 1994; Kraut 1978; Porter and Yuille 1996; Sapir 1987; Vrij 2000). Nevertheless, three major issues need to be addressed before adapting the existing cues to deception in TA-CMC. First, most experimental data in the prior research were collected via interview, interrogation, observation, or analysis of written statements of a specific past event. As is pointed out in Crystal (1969), language expression changes along with situational factors, such as speech community, register, genre, text and discourse type. Messages generated in TA-CMC exhibit different types of features than those from face-to-face communication. For example, e-mail is expressed through the medium of writing, though it displays several of the core properties of speech, such as expecting a response, transient, and time-governed. Nevertheless, the Internet language lacks the true ability to signal meaning through kinesic (body posture) and proxemic (distance) features, and this, along with the unavailability of prosodic features, places it at a considerable remove from spoken language (Crystal 2001). Therefore, it is inappropriate to treat e-mail either as spoken language as displayed in interviews or interrogations, or as written language as in written statements. The above comparison between the language in TA-CMC and traditional spoken or written language implies that cues derived from other studies need to be validated before applying them to TA-CMC. There is compelling evidence from prior deception research that a variety of language features, either spoken or written, can be valid indicators of deceit (Buller, Burgoon, Buslig and Roiger 1996; Burgoon, Buller, Afifi and Feldman 1996; Zuckerman et al. 1981). Yet, what kinds of cues from messages in TA-CMC could be used for alerting the recipient that deceit is occurring remains an open question that requires systematic empirical verification.

Secondly, current research demands intensive human involvement in decoding messages. Understanding and learning cues is time-consuming yet still does not necessarily produce consistency among human behavioral coders. Consequently, reliability must be checked before and after coding, and multiple coders are typically needed to ensure high reliability because of the subjectivity of human assessment. An automated approach using text-based cues would necessarily need to center on ones that can be objectively quantified.

Thirdly, among the existing text-based cues (Höfer et al. 1996; Porter and Yuille 1996; Steller and Köhnken 1989), some are strongly context sensitive and must be interpreted on the basis of a specific event, such as *unexpected complications during the incident* (Steller and Köhnken 1989), while others can be operationalized with general linguistic knowledge, such as *self reference* (Höfer et al. 1996). The latter are called linguistics-based cues (LBC).

Compared with other text-based cues, LBC do not suffer from the ground-truth problem—knowing with certainty whether what is being reported is truthful or false. This makes them less dependent upon expert knowledge and experience. Automating deception detection using text-based cues would also need to focus on those LBC that are relatively context-insensitive.

With these parameters in mind, we next review various coding systems from which we nominated cues with high discriminatory potential, present our classification scheme for LBC, then turn to a study in which we tested the effectiveness of 27 indicators in distinguishing messages encoded by truth-tellers from messages encoded by deceivers.

2. Previous work related to text analysis

Textual messages lack facial expressions, gestures, and conventions of body posture and distance, so the text itself is the only source for us to infer personal opinions and attitudes, and verify message credibility. Among the systems for analyzing textual information that have been proposed and accepted in research and/or practice are Criteria-Based Content Analysis (CBCA), Reality Monitoring (RM), Scientific Content Analysis (SCAN), Verbal Immediacy (VI) and Interpersonal Deception Theory (IDT) strategies and tactics. Even though none of them was developed specifically for TA-CMC, they provide the theoretical and evidentiary foundation for the cues included in the current investigation.

2.1. Criteria-based content analysis (CBCA)

CBCA was developed as one of the major elements of Statement Validity Assessment (SVA), a technique developed to determine the credibility of child witnesses' testimonies in trials for sexual offenses and recently applied to assessing testimonies given by adults (Raskin and Esplin 1991). It is based on the Undeutsch hypothesis that a statement derived from memory of an actual experience differs in content and quality from a statement based on invention or fantasy (Steller and Köhnken 1989; Undeutsch, 1989). The findings of recent research reveal that people are able to detect deception above the level that would be expected by chance by utilizing SVA and CBCA (Vrij 2000). CBCA focuses on the presence of specific semantic content characteristics. There are 19 criteria in the original CBCA (Steller and Köhnken 1989) which are grouped into four major categories: general characteristics, specific contents, motivation-related contents, and offense-specific elements. Trained evaluators examine the statement and judge the presence or absence of each criterion.

CBCA has shown some limitations in the application to detecting deception. Many factors may influence the presence of CBCA criteria, such as age of the child witness, cognitive interview, and stressful events (Vrij 2000). As part of SVA is targeted at children, some CBCA criteria do not work for adults (Landry and Brigham 1992). The purpose of CBCA was to detect truths rather than deception, as demonstrated in the evaluation of the

criteria on the testimonies of suspects or adult witnesses who talk about issues other than sexual abuse (Porter and Yuille 1996; Ruby and Brigham 1997; Steller and Köhnken 1989). Moreover, some criteria in CBCA need strong background knowledge about the concerning event in addition to familiarity with CBCA criteria in the validity checking.

2.2. Reality monitoring (RM)

RM was originally designed for studying memory characteristics. It implies that a truthful memory will differ in quality from remembering an event that has been made up. The former is likely to contain perceptual information (visual details, sounds, smells, tastes, and physical sensations), contextual information, and affective information (details about how someone felt during the event), while the latter is likely to contain cognitive operations (such as thoughts and reasoning) (Johnson and Raye 1981). Considering that deception is likely based on imagined rather than self-experienced events, RM has been applied in the context of deception detection. Among the eleven deception studies on the RM criteria (Sporer 1997) surveyed by Vrij (2000), eight showed that spatial and temporal information occurs more frequently in truthful than in deceptive statements, and seven found similar patterns for perceptual information. However, the criteria on cognitive operations were only supported by one study (Hernandez-Fernaud and Alonso-Quecuty 1997). In a crime simulation study (Porter and Yuille 1996), none of the three criteria selected from RM, frequency of verbal hedges, number of self-references, and number of words, was found to significantly differentiate between experiment conditions ranging from completely false to truthful confession. We are reluctant to draw any firm conclusions from such comparisons, as they were conducted in the interrogative context.

RM was found to be more useful for analyzing adults' statements than studying children's because children do not differentiate between ongoing fact and fantasy as clearly as adults do (Lindsay and Johnson 1987). RM might be particularly useful for analyzing statements about events that happened recently rather than a long time ago. People have a tendency to fill in gaps, particularly with imagined events, in order to make their stories sound interesting and coherent. Consequently, differences between perceived and imagined events become smaller when people are asked to put their memories into words (Johnson 1988; Vrij 2000).

2.3. Scientific content analysis (SCAN)

Given the transcript or written statement of a subject, SCAN is able to discriminate between adult criminal investigation statements of doubtful validity and those that are probably accurate (Driscoll 1994). Among the indicators listed in SCAN, some are suggestive of deceit when they are present (Sapir 1987), such as lack of memory and missing links; some are indicative of deception when they are absent, such as connections, spontaneous corrections, first person singular and other pronouns, past tense verbs, denial of allegations, unnecessary links, and changes in language; and others are contingent upon where they

occur, such as emotion and time. Due to such complexity in assessing a statement, it is recommended to pay extreme caution when multiple issues may be involved (Driscoll 1994).

A field study (Smith 2001) found that officers who used the SCAN technique, and those untrained officers who drew upon their experience as detectives to assess the statements, were all able to correctly identify at least 80% of the truthful statements and 65% of the deceptive statements. The officers who had not received SCAN training and used their general intuition to assess the statements were only able to correctly assess 45% of deceptive statements. However, an analysis of the use of SCAN criteria used by different assessors revealed low levels of consistency (Smith 2001). The written statement must be made without assistance from any other individual in order for SCAN to be effective.

2.4. Interpersonal deception theory (IDT) strategies and tactics

IDT (Buller and Burgoon, 1996) was developed to explain and predict deception and its detection in interpersonal contexts. As part of that theory development, Buller and Burgoon (1994), Burgoon, Buller, Guerrero, Afifi, and Feldman (1996; see also Jacobs, Brashers, and Dawson 1996, and McCornack 1992) proposed a series of general strategies and specific tactics that deceivers may employ to manage the information in their messages and to evade detection. Tests of IDT (e.g., Buller, Burgoon, Buslig, and Roiger 1994, 1996; Burgoon et al. 1996), along with prior research and a recent meta-analysis (DePaulo, Lindsay, Malone, Muhlenbach, Charlton, and Cooper 2003), have served to clarify what strategies and specific verbal indicators may be valid. They can be summarized as follows:

(a) *quality (truthfulness) manipulations* – deceivers may opt to deviate from the truth completely or partially. Half-truths and equivocations may be deceptive through the inclusion of adjectives and adverbs that qualify the meaning in statements. Other strategies below may further result in receivers drawing wrong inferences about the true state of affairs.

(b) *quantity (completeness) manipulations* – deceivers may be more reticent and less forthcoming than truth-tellers. They may exhibit reticence by using fewer words and sentences or less talk time than truth-tellers. Their messages may be incomplete syntactically, by giving perceptually less information than would normally be expected as a response, or semantically, by failing to present actual detailed content such as factual statements. Deceivers' language describing imagined events may also fail to reflect the rich diversity of actual events, as noted in CBCA and RM. Thus, two extensions of the concept of reduced completeness may include reduced content specificity and reduced lexical (vocabulary) and content diversity.

(c) *clarity (vagueness and uncertainty) manipulations* – deceivers' messages may be less clear by virtue of using contradictory or impenetrable sentence structures (syntactic ambiguity) or by using evasive and ambiguous language that introduces uncertainty (semantic ambiguity). Modifiers, modal verbs (e.g., should, could), and generalizing or "allness" terms (e.g., "everybody") may increase uncertainty.

(d) *relevance manipulations* – deceivers may give responses that are semantically indirect (e.g., forms of polite speech) or irrelevant (such as irrelevant details). They may also be syntactically indirect (e.g., following a question with a question).

(e) *depersonalism (disassociation) manipulations* – deceivers may use language to distance themselves from their messages and the contents of those messages. Nonimmediate language (described more fully below) such as lack of pronouns, especially first person pronouns, and use of passive voice reduce a sender's ownership of a statement and/or remove the author from the action being described. Other linguistic features such as use of more second person pronouns may imply dependence on others and lack of personal responsibility.

(f) *image- and relationship-protecting behavior* – "verbal and nonverbal behaviors used to make oneself appear sincere and trustworthy and to sustain the self-presentation one has created" (Buller and Burgoon 1994, p. 204). Verbal tactics may include avoidance of discrediting information (e.g., admitted lack of memory, expressions of doubt) and avoidance of negative affect in one's language (partially intended to cover any accidental betrayal of true feelings of guilt, fear of detection, etc.).

These strategies and tactics together point to a number of plausible text-based indicators of deception that may, despite deceivers' efforts to the contrary, reveal their deceptive intent. For example, the withdrawal and distancing associated with quantity and depersonalism manipulations may result in an overall pattern of uninvolvement that itself may give deceivers away. Other indicators that are nonstrategic (i.e., unintended) – such as cues related to nervousness, arousal, tension, negative affect, and incompetent speech performance – include mostly nonverbal cues. Two exceptions, unpleasantness and inexpressiveness, may also manifest themselves through use of adverbs and adjectives that express negative feeling states and attitudes and through less expressive or intense language.

2.5. Verbal immediacy (VI)

VI was originally proposed as a means of inferring people's attitude or affect (Mehrabian and Wiener 1966). The general construct of *immediacy-nonimmediacy* refers to verbal and nonverbal behaviors that create a psychological sense of closeness or distance. Verbal *nonimmediacy* thus encompasses any indication through lexical choices, syntax and phraseology of separation, non-identity, attenuation of directness, or change in the intensity of interaction between the communicator and his referents. The basic principle of assessing VI is via a literal interpretation of the words rather than their connotative meanings (Wiener and Mehrabian 1968). For example, while "you and I selected" may be equivalent to "we selected" in meaning, the former is considered more nonimmediate than the latter.

VI can be classified into three major categories: spatio-temporal, denotative specificity, and agent-action-object categories, each of which is further broken down into many subcategories (Wiener and Mehrabian 1968). VI has been applied to conversation analysis and coded on a scale with positive scores signifying approach and negative scores signifying avoidance (Borchgrevink unpublished; Donohue 1991). Avoidance is indicated by some nonimmediacy sub-categories, such as spatial and temporal terms, passive voice, presence of modifiers, and other expressions such as volitional words, politeness, and automatic phrasing. Detailed criteria for scoring nonimmediacy result in positive and negative scores assigned for the presence of each attribute. These are summed so that the higher the nega-

tive score for any utterance, the greater the probability that it is part of a communication about a negative experience or intended to distance the communicator from the listener and/or the message itself (Mehrabian and Wiener 1966). Since deception is frequently associated with negative affect and/or attempts to disassociate oneself from one's communication, VI measures are plausible indicators of deceit.

Many other individual studies and meta-analyses (DePaulo et al. 1985, 2003; Zuckerman et al. 1981) that covered certain kind of cues from texts could be mentioned here, but they largely have their origins in one of the above criteria or theories.

In summary, the review of literature clarifies that many aspects of text, such as content and style, have been employed as cues to deception. It should be emphasized that many of the above-mentioned criteria were developed for interrogation or interview contexts. The subjects in the experiments and the witnesses or suspects in the field studies were asked to describe or answer questions about a specific past event or experience, making such cues as temporal and spatial information, perceptual information, and quantity of details applicable. In TA-CMC, people are also likely to discuss some ongoing events or future decisions. With this change of context, we need to validate what cues may still be appropriate for TA-CMC and what factors may alter the previously discovered patterns. Furthermore, the emerging capacities of natural language processing, coupled with the principles of VI, open opportunities to discover new cues to deception in TA-CMC.

2.6. Linguistics based cues (LBC) and natural language processing (NLP)

LBC are involved with linguistic information in text unit(s), including words, terms, phrases, sentences, or an entire messages. A term is defined as a meaningful unit that consists of one or more content words and has distinct attributes (Zhou et al. 2002), whereas a phrase is composed of multiple words and/or terms. Many LBC can be extracted from the aforementioned criteria and constructs: contextual embedding in CBCA (Steller and Köhnken 1989); affective information in RM (Johnson and Raye 1981); first person singular pronouns and denial of allegations in SCAN (Sapir 1987); and spatio-temporal information and passive voice in VI (Wiener and Mehrabian 1968). As is evident from Table 1, previous approaches show some overlap. We have therefore synthesized these to produce a more parsimonious list of LBCs that are amenable to automation. Because past research has re-

Table 1. A sample list of LBC, their sources and depth of analyses

Cues	Sources	Depths of analyses
Passive voice	VI	Mo, Sy
Self reference	RM, SCAN	Mo
Negative statements	VI, SCAN	Mo, Sy, Ls
Generalizing terms	VI	Mo, Ls
Uncertainties	VI	Mo, Ls
Temporal information	CBCA, RM, VI	Mo, Sy, Ls
Spatial information	CBCA, RM, VI	Mo, Sy, Ls
Affect	RM, VI	Mo, Ls

lied on very time-consuming manual behavioral coding by human judges and because many cues require subjective interpretation that may vary substantially from one judge to the next, we turned to NLP techniques to assist with automating cue identification.

NLP enables people to communicate with machines using natural communication language by automatically analyzing and understanding human language with computers. Inspired by the process of human language understanding (breaking down larger textual units into smaller ones and integrating the understanding of small units into that of the whole text), NLP analyzes texts by going through sub-sentential, sentential, and discourse processing. Based on depth of analysis, the sub-sentential processing can be further classified into phonological analysis, morphological analysis, syntactic parsing, semantic analysis, and so on. Since the phonological analysis is usually performed on speech rather than written text, it is beyond our consideration in this study. Morphological analysis attempts to determine the part-of-speech of each word in a sentence, while syntactic parsing looks for the structure of a sentence following certain syntactic grammar. Full syntactic parsing into a hierarchical tree structure is not always necessary and may produce many ambiguous results; therefore, shallow parsing, extracting only the syntax one needs from a sentence, has gained popularity in practice. A shallow parser may identify some phrasal constituents, such as noun phrases, without indicating their internal structures and their functions in the sentence (Karlsson and Karttunen 1997). Semantic and discourse analyses dig deeper into the meaning and context and are very complex and difficult to automate. Therefore, we temporarily ignored LBC that require these two types of analyses except for those involving limited lexical semantic processing dealing with meaning of word(s). As a result, we focus on LBC that are involved with Morphological (Mo), Syntactic (Sy) and Lexical Semantic (Ls) analyses in this study (in short, MoSyLs). All the cues listed in Table 1 belong to these types. The third column in Table 1 also records the NLP analyses, noted in shorthand by the first two characters, that can be performed to identify a specific cue. For example, *temporal information* drawn from CBCA, RM, SCAN, and VI requires morphological, syntactic and lexical semantic analyses to automatically identify.

Most of prior studies combine LBC with other types of cues in detecting deception in face-to-face settings. What remains theoretically challenging is how applying pure LBC to deception would work in TA-CMC. We began by identifying the most promising MoSyLs cues from existing criteria and constructs, then merged them into a candidate cue list for testing in a TA-CMC simulation study. Encouraged by the research on stylistic analysis as a predictor of newspaper credibility (Burgoon et al. 1981), we added three other stylistic indices: complexity, pausality, and emotiveness. Complexity can be measured as the ratio of syllables to words or characters to words. Pausality, or amount of punctuation, may also be an indication of degree of sentence complexity. Emotiveness is the ratio of adjectives plus adverbs to nouns plus verbs, which was selected as an indication of expressivity of language. To measure actual emotional and feeling states, we included the amount of positively or negatively valenced terminology included in the messages and differentiated between positive and negative affect to determine if the total amount of affect or the valence of the affect made a difference. Finally, in TA-CMC, typos are both unavoidable and easily correctable if wanted. Thus, typos in a message may reflect informality of language in the communication, which might be another useful aspect to view deceptive messages.

2.7. Hypotheses

In building our hypotheses to test automated LBC for detecting deception in messages created in TA-CMC, our overriding premise was that LBC improves the performance of deception detection in general. Thus, we expected linguistic indicators to successfully discriminate between deceivers and truth-tellers.

Based on the preceding literature review, we might normally expect deceivers to minimize the amount of information that is presented and that could later be verified and determined to be deceptive. We might also expect some cognitive difficulty associated with deceit that could limit the amount and quality of discourse being presented and result, for example, in repetitive phrasing and less diverse language. Due to the possible arousal of guilty feelings, deceivers might be expected to take a low-key, submissive approach, to disassociate themselves from their messages through a higher degree of nonimmediacy, and to inadvertently reveal negative affect. We might expect more passive voice, modal verbs, objectification, other indicators of uncertainty, generalizing terms, fewer self-references, and more group references as means of increasing uncertainty and vagueness and as further disassociation. Due to over-control and less conviction about what is being said, the expressiveness of the language of deceptive senders might also be expected to be lower than truthful senders and to include less positive affect or less affect altogether. In order to create a sense of familiarity, which should activate positive biases, deceivers might show higher informality of language than truth-tellers.

Our conjectural language is due to the fact that the nature of TA-CMC and the task being used may alter many of these predictions. With regard to TA-CMC, several factors related to a reduction in interactivity argue against some of the above patterns (Burgoon, Bonito and Stoner 2003). First, participants interact at a distance, and proximity or lack of it is a big factor in how people relate to one another. At a distance, participants feel less connected to one another and therefore deceivers may experience less negative emotions about deceiving. Deceivers may even go to the other extreme by showing a positive state of mind on their falsified opinions in order to achieve their communication goal with remote partners. Second, the text medium gives deceivers fewer modalities to control and therefore more opportunities to attend carefully to the one modality they must monitor. Third, asynchronicity enables greater control and forethought, greater time for deceivers to plan, rehearse and edit what they say. This can reduce the cognitive difficulty of the task as well as the anxiety associated with answering "on the fly."

With regard to the task itself, deceptive senders were given the goal of convincing receivers to make decisions contrary to what they knew to be correct. The fact that the task was a persuasive one, one requiring deceivers to generate arguments and "evidence" to support their claims if they were to succeed, introduced a major change from previous experiments and raised the distinct possibility that deceivers would generate more, not less, discourse as part of advancing their arguments in behalf of their position. Research by Burgoon, Blair and Moyer (2003) had found that deceivers in their experiment were more motivated than truth-tellers to succeed in appearing truthful and that text communication was not particularly taxing. Thus we thought deceivers under TA-CMC might actually

produce longer (higher quantity) messages than truth-tellers. Moreover, a decision-making task with a strong persuasive component in it demands increasing expressiveness of deceivers' language in order to enhance the persuasiveness of their opinions, and use of positively valenced adjectives and adverbs (e.g., "great") and informal language might be especially useful both in building rapport and in reducing the appearance of trying to manipulate the partner. We reasoned that complexity, diversity, and specificity might still be limited, however, due to some continued cognitive taxation and lack of reliance on real memory. People commonly deceive by concocting lies or being equivocal and evasive. In the former case, the messages lack the support of rich and real memory, so they tend not to include specific details and lack language to refer to said same. In the latter case, deceivers may deliberately leave out specific details. In either case, the complexity, diversity, and specificity of language of senders in the deception condition should be lower than those in the truth condition. We also expected that senders would continue to introduce uncertainty in their language and disassociate themselves from their messages through nonimmediacy.

Hypothesis 1. Deceptive senders display higher (a) quantity, (b) expressivity, (c) positive affect, (d) informality, (e) uncertainty, and (f) nonimmediacy, and less (g) complexity, (h) diversity, and (i) specificity of language in their messages than truthful senders.

Inasmuch as language used by a sender has impact on that of the receiver, the issue of deception in interpersonal contexts can be approached from a dyadic and dialogic rather than monadic and monologic perspective, as suggested in IDT (Buller and Burgoon 1996). In our experiment, we labeled the initiator of a communication as the sender and the other party as the receiver. Senders were assigned to the truthful or deceptive condition, but receivers in both conditions were presumably truthful. Thus we could examine deceptive versus truthful discourse in two ways: by comparing deceptive senders to truthful senders (i.e., independent group comparisons) and by comparing deceptive senders to their truthful receivers (i.e., within-group comparisons). The second hypothesis thus extended the comparison of deceptive and truthful senders to that of deceptive senders and naïve (truthful) receivers:

Hypothesis 2. Deceptive senders display higher (a) quantity, (b) expressivity, (c) positive affect, (d) informality, (e) uncertainty, and (f) nonimmediacy, and less (g) complexity, (h) diversity, and (i) specificity of language in their messages than their respective receivers.

3. Method

The research experiment was a 2×2 repeated measures design varying experimental condition (deception, truth) and dyad role (sender, receiver). Participants were assigned one of the two roles in one of the two conditions and performed a task for three consecutive days under the same condition.

3.1. Participants

Participants (N = 60) were freshmen, sophomore, junior, and senior students (57% female, 42% male) recruited from a Management Information Systems course at a large southwestern university who received extra credit for experimental participation. Ranking grade level from low to high as 1 to 4, the average grade of completed subjects was 3.12. Failure to comply with the full requirements over the course of the entire experiment resulted in attrition, with 30 dyads successfully completing the entirety of the experiment. Among the 30 dyads, 14 were collected from the truth condition, and 16 from the deception condition. The messages from each subject were aggregated across three days to derive stable estimates.

3.2. Procedures

Participants completed the experiment by logging onto a designated web server from either labs on campus or from home. They were randomly assigned to two-person groups (dyad) and the dyads were randomly assigned to treatments depending on the order they logged in. Within dyads, participants were randomly assigned the role of "sender" or "receiver."

The task consisted of a modified version of the Desert Survival Problem (Lafferty and Eady 1974). The modified version presented participants with a scenario in which their jeep had crashed in the Kuwaiti desert and their primary goal was to achieve, through discussion, a consensus ranking of 12 items they should salvage in terms of their usefulness to survival. The task in the experiment was carefully selected to meet several criteria. First, it elicited high involvement by participants. Second, the experiment occurred in a natural setting, where subjects typically communicate with each other, increasing ecological validity. Third, it created opportunities for deception in exchanging electronic messages. Fourth, the single task was clearly defined in the instructions and supplemented with additional background knowledge. Fifth, the experiment was supported by an integrated system, which is embedded with flow control of the entire procedure, helping subjects interpret the task consistently and perform the task easily. Last but not least, it went beyond the traditional paradigm of structured interviews to the kind of decision-making task relevant to group work.

A list of *n* (10–12) salvageable items was available, depending on the scenario. Participants in each dyad exchanged their ideas by sending messages to each other via an email messaging system. Each sender first ranked the items based on his or her own truthful or deceptive opinion, composed an email message presenting his or her ranks and reasoning, and sent the email to the receiver within a half-day time slot. Each receiver read the message from his or her sender, re-ranked the items if necessary, and wrote a response to the sender within the given time slot. The senders started to receive messages from their partner from the second half day. The sender and receiver in each dyad communicated back and forth once for each of the three consecutive days before reaching a final decision. Deceptive senders were given special instructions on deceiving their partners when they first logged in, while truthful senders were instructed to offer their true opinions to their

partners. None of the receivers was informed of the senders' condition during the experiment. They were only told to collaborate with their partner to complete the decision-making task. Additionally, the system did not reveal the identities of the subjects to their partners, protecting anonymity.

On the second and the third days of the task, a random scenario was given to each dyad where one of items was removed from consideration. These items were removed to elicit discussion between the partners and give the task a sense of realism and urgency. The scenarios included such events as the dyad's water being spilled or the plastic sheeting being blown away in a storm.

The study was performed entirely using a web-based messaging system. Volunteers were given a web-site address and instructed on when to begin. The subjects completed each day's task by logging into the system from any web-enabled computer. Although performing the study using a web-based messaging system outside of the laboratory reduced the amount of experimental control that could be exercised, it allowed the subjects to perform the tasks at their convenience without the pressures and unnatural feel of the laboratory.

3.3. Independent variables

3.3.1. Dyad

Participants were randomly assigned to the (arbitrarily labeled) sender or receiver role. Senders were the participants who initiated the online communication. Receivers were the other member of each dyad and were the first to receive a message and reply to it. Due to the close relationship between a received messages and the corresponding response, sender and receiver behavior were not independent of one another, resulting in the need to treat dyad membership as a repeated, or within- , factor in the statistical design.

3.3.2. Deception condition

Senders were randomly assigned to the deception or truth condition. In the deception condition, senders were explicitly instructed to deceive the receiver about how they ranked the items; in the truth condition, senders offered their true opinions to the receiver.

3.4. Dependent variables and measures

Based on prior studies, the linguistic features of messages in TA-CMC, and the possibility of automation, we selected 27 LBC as dependent variables. Considering the correlations between some dependent variables, we grouped the LBC into eight linguistic constructs: quantity, complexity, uncertainty, nonimmediacy, diversity, affect, specificity, expressiveness, and informality. All the linguistic constructs and their component dependent variables and measures are summarized in Table 2.

A shallow parse was sufficient for identifying the LBC selected in this experiment. We adopted an NLP tool called iSkim (Zhou et al. 2002), which combines the accuracy of the EngCG-2 morphological tagger (Samuelsson and Voutilainen 1997; Voutilainen 2000) with

Table 2. Summaries of linguistic constructs and their component dependent variables and measures

Quantity

1. Word [a]: a written character or combination of characters representing a spoken word.

2. Verb [a]: a word that characteristically is the grammatical center of a predicate and expresses an act, occurrence, or mode of being.

3. Noun phrase [a]: a phrase formed by a noun, its modifiers and determiners.

4. Sentence [a]: a word, clause, or phrase or a group of clauses or phrases forming a syntactic unit which expresses an assertion, a question, a command, a wish, an exclamation, or the performance of an action, which usually begins with a capital letter and concludes with appropriate end punctuation.

Complexity

5. Average number of clauses: $\frac{\text{total \# of clauses}}{\text{total \# of sentences}}$

6. Average sentence length: $\frac{\text{total \# of words}}{\text{total \# of sentences}}$

7. Average word length: $\frac{\text{total \# of characters}}{\text{total \# of words}}$

8. Average length of noun phrase: $\frac{\text{total \# of words in noun phrases}}{\text{total \# of noun phrases}}$

9. Pausality: $\frac{\text{total \# of punctuation marks}}{\text{total \# of sentences}}$

Uncertainty

10. Modifiers [b]: describes a word or makes the meaning of the word more specific. There are two parts of speech that are modifiers - adjectives and adverbs.

11. Modal verb [a]: an auxiliary verb that is characteristically used with a verb of predication and expresses a modal modification.

12. Uncertainty: a word that indicates lack of sureness about someone or something [a].

13. Other reference: third person pronoun.

Nonimmediacy[c]

14. Passive voice: a form of the verb used when the subject is being acted upon rather than doing something.

15. Objectification [a]: an expression given to (as an abstract notion, feeling, or ideal) in a form that can be experienced by others and externalizes one's attitude.

16. Generalizing terms: refers to a person (or object) as a class of persons or objects that includes the person (or object).

17. Self reference: first person singular pronoun.

18. Group reference: first person plural pronoun.

Expressivity

19. Emotiveness: $\frac{\text{total \# of adjectives + total \# of adverbs}}{\text{total \# of nouns + total \# of verbs}}$

Diversity

20. Lexical diversity: $\frac{\text{total \# of different words or terms}}{\text{total \# of words or terms}}$, which is the percentage of unique words or terms in all words or terms.

Table 2. Continued

21. Content word diversity: $\frac{\text{total \# of different content words or terms}}{\text{total \# of content words or terms}}$, where content words or terms primarily express lexical meaning.

22. Redundancy: $\frac{\text{total \# of function words}}{\text{total \# of sentenses}}$, where function words express primarily grammatical relationships.

Informality
23. Typographical error ratio: $\frac{\text{total \# of misspelled words}}{\text{total \# of words}}$

Specificity[c]
24. Spatio-temporal information: information about locations or the spatial arrangement of people and/or objects, or information about when the event happened or explicitly describes a sequence of events.
25. Perceptual information: indicates sensorial experiences such as sounds, smells, physical sensations and visual details
Affect[c]
26. Positive affect[a]: conscious subjective aspect of a positive emotion apart from bodily changes.
27. Negative affect[a]: conscious subjective aspect of a negative emotion apart from bodily changes.

a: Source of definition: *www.webster.com*
b: Source of definition: *http://englishplus.com/grammar/glossary.htm*
c: Individual measures in the construct are calculated per message unit, i.e. frequency counts divided by the total number of words, to adjust for differential message lengths.

the information produced by EngLite syntax (http://www.conexoroy.com/lite.htm) and named entity extraction. Some types of named entities, such as location and time, were directly related to the selected LBC. The software provided critical information for measuring the LBC in Table 2. Based on iSkim's output, another tool, CueCal, was developed to derive the value of each individual cue. For example, the cue *lexical diversity* was measured using the following steps: iSkim first reduced all the words that have inflectional changes into their base forms (stems), and then CueCal identified terms in addition to words, counted the total number of words or terms as well as unique words or terms, and finally divided the latter by the former to derive the value of lexical diversity.

4. Results

Hypotheses were tested with 2 × 2 repeated-measure analyses of variance. Multivariate analyses were initially conducted on sets of related variables, followed by simple effect tests on the 27 individual dependent variables. Dyad was set to sender in testing the simple effect of deception in Hypothesis 1 and condition was set to deception in testing the simple effects of dyad in Hypotheses 2. Table 3 lists the means and standard deviations of all dependent variables.

Table 3. Means (standard deviations) for LBC (dependent measures)

Cues	Condition*	Sender	Receiver
Word	T	272.4[124.3]	273.5[142.7]
	D	391.3[123.5]	329.7[137.6]
Verb	T	56.8[28.5]	58.3[38.3]]
	D	92.9[29.7]	71.5[33.8]
Noun phrase	T	97.5[49.8]	96.2[54.8]
	D	132.7[39.6]	110.8[48.3]
Sentence	T	18.8[10.3]	21.2[12.5]
	D	25.8[8.9]	19.6[9.0]
Modifier	T	29.6[14.7]	32.1[17.2]
	D	48.3[19.5]	34.8[16.5]
Modal verb	T	0.057[0.026]	0.046[0.028]
	D	0.073[0.02]	0.05[0.015]
Uncertainty	T	0.013[0.013]	0.014[0.012]
	D	0.012[0.012]	0.011[0.009]
Other reference	T	0.007[0.009]	0.004[0.005]
	D	0.003[0.005]	0.005[0.007]
Passive voice	T	0.015[0.01]	0.013[0.01]
	D	0.018[0.013]	0.015[0.013]
Objectification	T	0.009[0.012]	0.008[0.01]
	D	0.008[0.009]	0.008[0.01]
Generalizing term	T	0.028[0.016]	0.028[0.017]
	D	0.021[0.015]	0.017[0.01]
Self reference	T	0.035[0.029]	0.035[0.027]
	D	0.022[0.016]	0.033[0.022]
Group reference	T	0.016[0.013]	0.019[0.015]
	D	0.03[0.023]	0.02[0.015]
Emotiveness	T	0.272[0.085]	0.304[0.108]
	D	0.289[0.06]	0.249[0.089]
Avg. number of clauses	T	0.95[1.0]	0.55[0.32]
	D	0.55[0.25]	0.62[0.44]
Avg. sentence length	T	19.6[15.0]	16.1[8.8]
	D	15.2[4.6]	17.3[5.5]
Avg. word length	T	3.9[0.22]	3.9[0.22]
	D	4.0[0.24]	3.9[0.15]
Avg. NP length	T	1.7[0.27]	2.3[2.4]
	D	1.7[0.17]	2.0[1.4]
Pausality	T	3.1[2.2]	2.9[2.4]
	D	1.9[0.53]	2.8[1.9]

Table 3. Means (standard deviations) for LBC (dependent measures)

Cues	Condition*	Sender	Receiver
Typographical error ratio	T	0.005[0.008]	0.006[0.008]
	D	0.01[0.007]	0.01[0.008]
Lexical diversity	T	0.719[0.073]	0.719[0.111]
	D	0.637[0.074]	0.679[0.107]
Content diversity	T	0.732[0.089]	0.737[0.118]
	D	0.641[0.088]	0.70[0.093]
Redundancy	T	7.507[6.525]	5.972[3.791]
	D	5.501[1.871]	6.336[2.176]
Spatio-temporal information	T	0.043[0.012]	0.047[0.022]
	D	0.047[0.018]	0.048[0.016]
Perceptual information	T	0.015[0.011]	0.016[0.011]
	D	0.018[0.011]	0.02[0.012]
Positive affect	T	0.004[0.005]	0.007[0.007]
	D	0.009[0.007]	0.006[0.006]
Negative affect	T	0.003[0.005]	0.003[0.005]
	D	0.004[0.004]	0.002[0.002]

*T: truth condition; D: deception condition

Hypothesis 1 received support on numerous measures. The multivariate analysis on *quantity* measures showed that messages from deceptive senders were significantly different from those from truthful senders on quantity, Wilk's $\lambda = 0.607$, $F(4, 25) = 4.043$, $p = 0.012$, $\eta^2 = 39.3\%$. Compared with truthful senders, deceptive senders used more words, $F(1,28) = 6.877$, $p = 0.014$, verbs, $F(1,28) = 11.446$, $p = 0.002$, noun phrases, $F(1,28) = 4.644$, $p = 0.040$, and sentences, $F(1,28) = 4.054$, $p = 0.054$ (equivalent one-tailed p-value = 0.028). A univariate analysis on *informality*, $F(1, 28) = 3.89$, $p = .058$, $\eta^2 = 12\%$, was significant as a directional test (i.e., a t-test at $p < 0.05$, one-tailed). Deceivers used more informality in the form of more typographical errors. The multivariate analysis was likewise significant on *diversity* measures, Wilk's $\lambda = 0.717$; $F(3, 26) = 3.58$, $p = 0.027$, $\eta^2 = 29\%$, and *uncertainty* measures, Wilk's $\lambda = 0.658$; $F(4, 25) = 3.242$, $p = 0.028$, $\eta^2 = 34.2\%$. As predicted, deceivers displayed less lexical diversity, $F(1, 28) = 9.322$, $p = 0.005$, and content diversity, $F(1, 28) = 8.116$, $p = 0.008$, and more modifiers, $F(1,28) = 8.55$, $p = 0.007$, and modal verbs, $F(1, 28) = 3.88$, $p = 0.059$ (equivalent one-tailed p-value = 0.029), than truthful senders. The multivariate effect for *affect* failed to achieve conventional levels of significance, $F(2, 27) = 2.85$, $p = 0.07$, $\eta^2 = 17\%$, but the follow-up univariate analyses showed that deceptive senders produced more positive affect, $F(1, 28) = 5.27$, $p = 0.029$, than truthful senders. The multivariate tests on *complexity*, $F(5, 24) = 1.30$, $p = 0.297$, *nonimmediacy*, $F(5, 24) = 1.746$, $p = 0.162$, *specificity*, $F(2, 27) = 0.60$, $p = 0.58$, and *expressivity*, $F(1, 28) = 0.43$, $p = 0.517$, respectively, also failed to yield significant results. However, the univariate analyses revealed that compared with truthful senders, deceptive senders created signifi-

cantly less pausality, $F(1, 28) = 4.63$, $p = 0.04$; and used more group references, $F(1, 28) = 4.15$, $p = 0.051$ (equivalent one-tailed p-value = 0.025), than truthful senders. Thus, as hypothesized, deceptive senders created longer, more informal, more uncertain and nonimmediate, less complex, and less diverse messages than truth-tellers.

As predicted in Hypotheses 2, multivariate analyses within dyads produced differences between deceptive senders and their truthful partners on *affect*, Wilk's $\lambda = 0.49$; $F(2, 14) = 7.38$, $p = 0.006$, partial $\eta^2 = 51\%$; *uncertainty*, Wilk's $\lambda = 0.225$; $F(4, 12) = 10.362$, $p = 0.001$, partial $\eta^2 = 77.5\%$; and *expressivity*, Wilk's $\lambda = 0.72$; $F(1, 15) = 75.88$, $p = 0.028$, partial $\eta^2 = 28\%$; and produced a near-significant effect on *diversity*, Wilk's $\lambda = 0.61$; $F(3, 13) = 2.75$, $p = 0.085$, partial $\eta^2 = 39\%$. The follow-up univariate analyses showed that compared with truthful partners, deceptive senders showed greater uncertainty in the form of modifiers, $F(1,15) = 8.93$, $p = 0.009$; and modal verbs, $F(1, 15) = 44.04$, $p < 0.001$. In addition, their language displayed more negative affect, $F(1, 15) = 7.35$, $p = 0.016$, and emotiveness, $F(1, 15) = 5.88$, $p = 0.028$, but had lower lexical diversity, $F(1, 15) = 6.62$, $p = 0.021$, and content diversity, $F(1, 15) = 8.46$, $p = 0.011$. Although analyses failed to produce significant multivariate differences on *quantity*, $F(4, 12) = 2.023$, $p = 0.155$; *nonimmediacy*, $F(5, 11) = 1.543$, $p = 0.255$; *specificity*, $F(2, 14) = 0.194$, $p = 0.825$; *complexity*, $F(5, 11) = 1.66$, $p = 0.225$; or *informality,* $F(1, 15) = 0.001$, $p = 0.981$, univariate analyses showed that relative to their non-deceptive partners, deceptive senders produced more language in the form of more words, $F(1, 15) = 3.278$, $p = 0.09$ (equivalent one-tailed p-value = 0.045); sentences, $F(1, 15) = 3.567$, $p = 0.078$ (equivalent one-tailed p-value = .039), and verbs, $F(1,15) = 5.92$, $p = 0.028$; and they were lower on pausality (complexity) than receivers, $F(1, 15) = 3.78$, $p = 0.071$ (equivalent one-tailed p-value = 0.035). Their language was also more nonimmediate, as shown by fewer self references, $F(1, 15) = 3.585$, $p = 0.078$, and more group references, $F(1, 15) = 3.675$, $p = 0.074$ (equivalent one-tailed p-values = 0.039 and 0.037). In sum, Hypothesis 2 received substantial support. Deceivers exhibited greater expressivity, uncertainty, quantity, and nonimmediacy, and lower complexity and diversity. Contrary to expectations, but consistent with prior research, deceivers showed more negative rather than positive affect, and specificity failed to emerge as a discriminator.

5. Discussion

5.1. Major findings

This investigation sought to determine the viability of using LBC to distinguish truthful from deceptive messages. Taken together, our two hypotheses received considerable support for all classes of linguistic features studied except specificity. Consistent with our hypothesis but contrary to much prior research, deceivers displayed higher quantity – of words, verbs, noun phrases, and sentences. Their messages were also more expressive than their partners and they appeared more informal, as they had more typographical errors than truth-tellers. Consistent with other research and our hypotheses, deceptive subjects in this study displayed less diversity at both the lexical and content level than did truth-tellers.

They also used nonimmediate and uncertain language in the form of less self-reference, more group references, more modal verbs, and more modifiers. Moreover, their messages were less complex, as evident by less punctuation (pausality). One anomaly was that, although affective references were higher by deceivers, the between-groups comparison showed more positive affect, whereas the within-dyads comparison showed more negative affect, leading to the very tentative conclusion that deceivers in general used more affective language. Finally, specificity in terms of spatio-temporal or perceptual references was not found to vary between truth-tellers and deceivers, although that might be attributable to our reliance on an as-yet very small dictionary of spatio-temporal and perceptual terms.

How do these results compare with prior investigations? The greater quantity of language runs contrary to IDT's prediction of deceivers typically opting to say less, and the greater expressivity is counter to past face-to-face findings showing deceivers to be nondemonstrative and inexpressive. The uncertainty and nonimmediacy are consistent with general strategies of obfuscation and equivocation. The trend that deceivers showed more affective information than truth-tellers runs contrary to what has been found in RM investigations, and the lack of evidence in support of spatio-temporal information was contrary to the prediction in CBCA, RM, and VI. However, virtually all of the differences between this experiment's findings and those of prior investigation can be laid at the feet of the unique characteristics of asynchronous, distributed, text-based communication and the specific task. Unlike interviews, in which respondents must construct answers spontaneously in real time, with little opportunity for prior planning, rehearsal, or editing, deceivers in this investigation had ample opportunity to create and revise their messages so as to make them as persuasive as possible. Additionally, unlike interviews requiring narratives about specific events, the task was an advocacy one that required participants to offer their opinions and to give reasons for their recommendations, a task likely to elicit more rather than less discourse and with little concrete basis for partners to suspect duplicity.

Some LBC that were advanced in this investigation have not been considered previously. These include complexity, expressivity, informality, and content diversity, all of which were found effective in distinguishing truthful from deceptive messages. Additionally, the enrichment of the quantity construct with verb and noun phrase quantities in addition to word quantity greatly increased the distinction between experimental conditions. The importance of breaking down affect into positive and negative categories was illustrated by the fact that only positive affect could significantly differentiate between truthful and deceptive senders, and only negative affect could differentiate between senders and receivers in the deception condition.

While the majority of dependent constructs and LBC in Table 2 were found effective to detect deception in this study, some specific indicators such as passive voice, objectification, generalizing terms, other references, and redundancy were not effective discriminators in this study. It is possible that these indicators will not prove to be reliable cues to deception, possibly because some are easy for deceivers to readily manipulate in a TA-CMC setting and thus to approximate the language of truth-tellers. However, it may be that under different contexts and tasks, they will emerge as relevant and should therefore not be discounted at this early stage of investigating LBC. We can infer from the above results that diversity,

uncertainty, quantity, and affect constructs were relatively robust and were applicable to distinguishing deceivers from both truthful senders and receivers, whereas other constructs may only be effective in at most one of the above comparisons. In view of the level of interaction between senders and receivers and the type of deception task, we can clearly see that detecting deception is an extremely complex task, with many dynamic and contextual factors.

5.2. Implications for automating deception detection

Despite its complexity, automating deception detection with accuracy beyond the level of chance is still a reachable goal as further research yields more effective cues and technology advances allow for the use of more complex cues. The significant results from the computer-generated measures used in this study demonstrate that a computational approach is a valid one in tapping the various variables being examined. Given a list of computerized cues and their preferable conditions, deception detection could become available to laypersons.

The identification of cues to deception is the first step in automating deception detection. It is especially important to identify those cues that can easily be implemented using current technology. Cues, such as identifying logical inconsistencies, that require specific domain knowledge or deep semantic understanding may be powerful, but are currently computationally infeasible across domains. Focusing on these complex cues may unnecessarily delay the automation of deception detection. All of the cues presented in this study are easily implementable using current commercial and open-source technology. As natural language processing continues to advance, however, more potential cues such as logical consistency, contextual embedding, and avoidance behaviors may become available for testing for possible use in improving automated deception detection.

A system designed for automated deception detection could be based on machine learning techniques that derive weights for the various cues presented in this study. (The automation of identifying possible cues to deception enables developing a fully automated deception detection system by taking advantage of machine learning algorithms.) Specifically, we could first use machine learning to discover the weights of cues from previously classified messages in a given context. Those cues or indicators could then be used to create a set of profiles for deceptive messages in that context. Finally, the values of indicators in a message could be fed as features into a system that learns to combine evidence to generate high-confidence warnings of deception. This machine learning approach to deception detection has the ability to adjust to different strategies of deception that appear in different contexts. The persuasive nature of the Desert Survival task used in this study may have resulted in the deceivers creating more words in their messages than truth-tellers. In other contexts, such as the criminal interrogations studied in much of the previous research, a deceiver may be trying to conceal facts and produce fewer words in a message than a truth-teller. Any generalizable automated deception method or system would need to adapt to different contexts. (Thus, adaptability will be a desirable feature of automated systems that can successfully detect deception in different contexts.)

5.3. Limitations

At least four plausible explanations exist to explain why our findings failed to support some of the LBC that had emerged in face-to-face settings. The communication goal of deceptive senders was to persuade the receivers to accept an incorrect solution to the problem. When deceivers feel the need of diverting others' attention from the right path in order to fulfill their goals, they are more likely to adopt a persuasive strategy. In order to influence their partners' decision, deceivers tend to invent substantial "evidence" to justify their misleading suggestions. Even though the deceivers generally go through cognitive difficulties and have little memory to recollect during the deceiving process, they are put into an advantaged position in the text-based asynchronous setting. These advantages include the invisibility of typical signs of cognitive difficulty such as delayed speech and hesitation between speeches and an abundance of time to fabricate messages. Without knowing who their partners are in TA-CMC, deceivers may begin building trust with their partners by intentionally demonstrating their "credibility" in performing the task. We can infer from the research results on virtual communication (Chidambaram 1996; Jarvenpaa and Leidner 1998) that physically distributed deceivers may have the motivation to build trust and a pseudo-relationship with their remote partners regardless of their diverse communication goals. Building a relationship of trust may be another practical deceiving strategy used to make up for the lack of memory, leading to longer messages. The observation of experimental messages confirmed our supposition. For example, Message 1 is from a truthful sender, and Message 2 and 3 are excerpts from two deceivers' messages.

Message 1: "*Water first, then coat to keep the sun and cold off, map and compass to navigate, canvas as umbrella and blanket, matches for night, transparent plastic for sand storms, book of plants?, knife, flashlight, mirror and gun I am uncertain what to do with.*"

Message 2: "*I wanted to let you know that when I was in High School I spent three days on a ""survival mission"" living in the snow covered woods with only limited supplies. Upon completion my YMCA team spent the next few weeks learning about survival in various other environments. I just thought that would offer some credibility on how I ranked my items. I took into account the time we would be there and the fact that situations like this are always filled with group conflict. I have a lot of recources on this and have refered to them. . .*"

Message 3*: "well, because i have actually taken a class about desert survival, i know what i'm talking about. The most important thing to have is water. . .*"

Deceivers' use of more group references also conforms to the goal of winning the trust of their partners. One may argue that the deceivers' longer messages were because of their less obvious ranking choices compared with the truth-tellers' more straightforward decisions. We do not think this factor is of concern for the following reasons: (1) few of the subjects have had real experience of surviving in the desert, so the correct answers are not self-evident; (2) even if several items are obviously more important than others, it is not easy for a dyad to achieve agreement on the ranking of 10 or more items; (3) when truth-tellers sense the irrationality in their partner's suggestions, they may jump into the defending position and produce long messages as well. Therefore, we believe that the difference in

message length between deceives and truth-tellers was a result of using an asynchronous and text-based form of communication as well as the nature of the experimental task.

A second explanation for why some sensory details were absent is also related to the nature of the task. Even though the task was set in a certain time and location, participants were mainly involved in discussing better ways of solving the immediate problem assigned to them. The genre of messages was more of argumentation rather than narration. Therefore, the discriminatory capability of spatio-temporal information and the specificity construct was not evident in this study. Clearly, further research is required before fully discounting the utility of these constructs in TA-CMC.

A third potential explanation might be the differences between TA-CMC and interviews or interrogations. In a structured interview, the interviewer has control over the subject and length of the interview. Given the reduced interaction resulting from asynchronicity and distance in the TA-CMC environment of this study, deceivers tended to produce more, rather than less language, and their language was richer in emotiveness and affective information. Since text was the sole channel deceivers could use to convey information and it had adequate capacity, they may have converted cues that would have otherwise been conveyed in other channels, such as body, facial, and/or voice, into text by providing more and richer language. Therefore, the pattern of quantity and expressiveness of language and affective information in deceptive messages in TA-CMC was a complete reversal of those shown in other real-time and face-to-face communication. This reversal also fits in with the view that deceivers are highly strategic (Buller and Burgoon 1994). When circumstances argue for trying to evade detection by saying less, which is what having to produce deception extemporaneously ought to encourage, they do so. Nonetheless, when there is time to create a more plausible and detailed fabrication, deceivers also do so. In addition, due to the loose structure and informal style of messages in TA-CMC, punctuations are not used with caution. Some subjects did not give a full stop to their sentences until reaching the end of their messages, while others simply used phrases rather than complete sentences in their messages. As a result, the effect of redundancy may have been nullified if there was any, and that of pausality was opposite to the prediction. The differences in punctuation and grammaticality in text-based communication raises a special challenge that researchers in TA-CMC will have to address.

Embedding decision-making within the task may serve as the last explanation for why our findings differ from past research. While making decisions on an unfamiliar task, both truth-tellers and deceivers are likely to display objectification, uncertainty, and generalizing terms. Therefore, these nonimmediacy markers are likely to be evident in the messages of all participants, regardless of their truthfulness, which would have dampened the effects of these LBC.

As mentioned, most of the LBC did receive support in the investigation. However, this study suggested that (1) notwithstanding their apparent utility with untrue reports in other contexts, some linguistic criteria from CBCA, RM, SCAN, and VI may not be valid in identifying intentionally falsified opinions in TA-CMC; and (2) some new LBC that were found to be effective for TA-CMC may not be extensible to other contexts.

Questions remain as to the external validity of these results to other tasks or contexts. We selected a decision-making task in this experiment. It is likely that deceivers may adopt

different types of strategy when they are given a different task, which may further lead to changes in linguistic behavior.

Language was not sufficiently natural. Due to the nature of the task, some subjects used an ordered list of items as the main corpus of their messages. As a result, the names of different items frequently appeared in the messages, which reduced lexical diversity and sentence length. Since these problems occurred in both deception and truth conditions, we can assume that the effect of listing in lieu of writing out narrative detail was equally applicable across conditions.

The reciprocal effect of receivers' behavior on sender communication was not considered here. In the experiment, some "intelligent" receivers, who had rich background knowledge, might have demonstrated less susceptibility to suggestions from deceiving senders. Being "caught" in the deception might have altered the sender's language.

We provided incentives for student subjects to participate in the experiment, but we did not offer special incentives for deceptive senders to succeed in deception. The questionnaire for deceivers conducted after each round of message exchange included some self-reported measures of deception, which helped monitor deceivers' behavior and intention to some degree. However, it might be better to give explicit motivation for accomplishing the task of deception.

7. Future research

Cross-validation studies are crucial with this type of applied research. We plan to test the effective cues found in this study with messages created for other types of tasks in real environments. Measurement of cues to deception can also be conducted beyond the word level, such as phrases, sentences, and messages. Thus, we may compare the effectiveness of LBC at different levels. We plan to further test the effectiveness of LBC in distinguishing truth from deception by examining messages composed by real-life deceivers.

The impact of deception strategy or speech act type on LBC, and longitudinal analyses of how LBC change as communication goals change, are also issues that merit exploration. How deceivers manipulate their linguistic behaviors over time is still largely absent in literature. Linguistic behavior is necessarily fluid, and it is likely that LBC will change as deceivers' communication goals change and they switch deception strategies.

Even though NLP techniques for semantic and pragmatic analyses are not yet mature, it is still possible to perform some kind of discourse analysis. It may enhance the effectiveness of deception detection by combining automated discourse-based cues, knowledge bases, and MoSyLs cues. Meantime, we can continue to enrich the list of LBC.

Deception research in face-to-face settings has received strong support from theories such as information management in IDT (Burgoon et al. 1996). In view of the contrary findings in this study, we believe that it is necessary to develop new constructs and theories to reveal the relationship between the underlying process deceivers go through and their exhibited behavior in the new medium of TA-CMC.

7. Concluding remarks

The results of this first investigation into LBC that are amenable to automation have validated some cues while challenging others derived from existing criteria and constructs such as CBCA, RM, SCAN, IDT and VI and are uneven in their congruence with results from face-to-face interactions. Based on this study and prior research, we conclude that many deception cues are highly context- and task-dependent, and that discerning profiles of reliable cues will necessitate clear delineation of the conditions under which deception is taking place. Nevertheless, the results indicate that nearly all the linguistic features we considered – quantity, informality, expressivity, affect, uncertainty, nonimmediacy, diversity, specificity, and complexity – are potentially relevant discriminators. CBCA, RM, and SCAN criteria have been tested in laboratory and field studies. Extending the above criteria to the context of TA-CMC, and revising them based on these empirical data, is appealing. After all, TA-CMC is a popular communication format and has distinguishable features from other communication types, thus the importance of developing a set of criteria adapted to TA-CMC.

In our opinion, the idea of detecting deception with automated LBC is feasible. Nonetheless, the same cue profiles are unlikely to apply uniformly across contexts. Future research must not only continue to validate what LBC are applicable to CMC but also what features of communication contexts and tasks will modify the patterns that are manifested. We hope the current investigation stimulates further inquiries along these lines.

Acknowledgements

This research was partially supported by funding from the U.S. Air Force Office of Scientific Research under the U.S. Department of Defense University Research Initiative (grant #F49620-01-1-0394). The views, opinions, and/or findings in this report are those of the authors and should not be construed as an official Department of Defense position, policy, or decision.

References

Aggarwal, C. C. and P. S. Yu. (2001). "Outlier Detection for High Dimensional Data," Proceedings of the ACM SIGMOD, Santa Barbara, California, 37–46.

Akehurst, L., G. Köhnken, and E. Höfer. (1995). "The Analysis and Application of Statement Validity Assessment," Proceedings of the Fifth European Conference on Psychology and Law, Budapest, Hungary.

Borchgrevink, C. P. (unpublished). "Verbal Immediacy: *Verbal immediacy coding scheme*."

Buller, D. B. and J. K. Burgoon. (1994). "Deception: Strategic and Nonstrategic Communication," *in* J. A. Daly, and J. M. Wiemann (eds.), *Strategic Interpersonal Communication*. Hillsdale, NJ: Erlbaum, 191–223.

Buller, D. B. and J. K. Burgoon. (1996). "Interpersonal Deception Theory," *Communication Theory* 6, 203–242.

Buller, D. B., J. K. Burgoon, A. Buslig, and J. Roiger. (1994). "Interpersonal Deception: VIII. Nonverbal and Verbal Correlates of Equivocation from the Bavelas et al. (1990) Research," *Journal of Language and Social Psychology* 13, 396–417.

Buller, D. B., J. K. Burgoon, A. Buslig, and J. Roiger. (1996). "Testing Interpersonal Deception Theory: The Language of Interpersonal Deception," *Communication Theory* 6, 268–289.

Burgoon, J. K., D. E. Buller, L. K. Guerrero, W. A. Afifi, and C. M. Feldman. (1996). "Interpersonal Deception: XII. Information Management Dimensions Underlying Deceptive and Truthful Messages," *Communication Monographs* 63, 52–69.

Burgoon, J. K., M. Burgoon, and M. Wilkinson. (1981). "Writing Style as Predictor of Newspaper Readership, Satisfaction and Image," *Journalism Quarterly* 58, 225–231.

Chidambaram, L. (1996). "Relational Development in Computer-Supported Groups," *MIS Quarterly* 20, 143–165.

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

Crystal, D. (2001). *Language and the Internet*. Cambridge: Cambridge University Press.

DePaulo, B. M., J. T. Stone, and G. D. Lassiter. (1985). "Deceiving and Detecting Deceit," *in* B. R. Schlenker (ed.), *The Self and Social Life*. New York: McGraw-Hill.

DePaulo, B. M., J. J. Linsay, B. E. Malone, L. Muhlenbac, K. Charlton, and H. Cooper. (2003). "Cues to Deception," *Psychological Bulletin* 129, 74–118.

Donohue, W. A. (1991). "Verbal Intensity: *Communication, Marital Dispute, and Divorce Mediation*," Hillsdale, NJ: Lawrence Erlbaum Associates.

Driscoll, L. N. (1994). "A Validity Assessment of Written Statements from Suspects in Criminal Investigations Using the Scan Technique," *Police Studies* 17, 77–88.

Ekman, P. (1985). *Telling Lies*. New York: W. W. Norton & Company.

Fawcett, T. and F. Provost. (1997). "Adaptive Fraud Detection," *Data Mining and Knowledge Discovery Journal* 1, 291–316.

Frank, M. G. and T. H. Feeley. (in press). "To Catch a Liar: Challenges for Research in Lie Detection Training," *Journal of Applied Communication Research.*

Hernandez-Fernaud, E. and M. Alonso-Quecuty. (1997). "The Cognitive Interview and Lie Detection: A New Magnifying Glass for Sherlock Holmes?" *Applied Cognitive Psychology* 11, 55–68.

Höfer, E., L. Akehurst, and G. Metzger. (1996). "Reality Monitoring: A Chance for Further Development of CBCA?" Proceedings of the Annual Meeting of the European Association on Psychology and Law, Sienna, Italy.

Internet Society, *info.isoc.org*.

Jarvenpaa, S. L. and D. E. Leidner. (1998). "Communication and Trust in Global Virtual Terms," *Journal of Computer-Mediated Communication* 3.

Johnson, M. K. (1988). "Reality Monitoring: An Experimental Phenomenological Approach," *Journal of Experimental Psychology: General* 117, 390–394.

Johnson, M. K., M. A. Foley, A. G. Suengas, and C. L. Raye. (1988). "Phenomenal Characteristics of Memories for Perceived and Imagined Autobiographical Events," *Journal of Experimental Psychology: General* 117, 371–376.

Johnson, M. K. and C. L. Raye. (1981). "Reality Monitoring," *Psychological Review* 88, 67–85.

Karlsson, F. and L. Karttunen. (1997). "Sub-Sentential Processing," *in* R. Cole, J. Mariani, H. Uszkoreit, A. Zaenen, and V. Zue (eds.), *Survey of the State of the Art in Human Language Technology*, Pisa, Italy: Cambridge University Press.

Köhnken, G., E. Schimossek, E. Aschermann, and E. Höfer. (1995). "The Cognitive Interview and the Assessment of the Credibility of Adults' Statements," *Journal of Applied Psychology* 80, 671–684.

Kraut, R. E. (1978). "Verbal and Nonverbal Cues in the Perception of Lying," *Journal of Personality and Social Psychology* 36, 380–391.

Lafferty, J. and P. Eady. (1974). *The Desert Survival Problem*. Plymouth, Michigan: Experimental Learning Methods.

Landry, K. and J. C. Brigham. (1992). "The Effect of Training in Criteria-Based Content Analysis on the Ability to Detect Deception in Adults," *Law and Human Behavior* 16, 663–675.

Lindsay, D. S. and M. K. Johnson. (1987). "Reality Monitoring and Suggestibility: Children's Ability to Discriminate Among Memories from Different Sources," *in* S. J. Ceci, J. Toglia, and D. F. Ross (eds.), *Children's Eyewitness Memory*. New York: Springer-Verlag, 91–121.

Mehrabian, A. and M. Wiener. (1966). "Nonimmediacy Between Communicator and Object of Communication in a Verbal Message: Application to the Inference of Attitudes," *Journal of Consulting Psychology* 30, 420–425.

Mukherjee, B., L. T. Heberlein, and K. N. Levitt. (1994). "Network Intrusion Detection," *IEEE Network* 8, 26–41.

Porter, S. and J. C. Yuille. (1996). "The Language of Deceit: An Investigation of the Verbal Clues to Deception in the Interrogation Context," *Law and Human Behavior* 20, 443–458.

Raskin, D. C. and P. W. Esplin. (1991). "Statement Validity Assessment: Interview Procedures and Content Analysis of Children's Statements of Sexual Abuse," *Behavioral Assessment* 13, 265–291.

Ruby, C. L. and J. C. Brigham. (1997). "The Usefulness of the Criteria-Based Content Analysis Technique in Distinguishing Between Truthful and Fabricated Allegations," *Psychology, Public Policy, and Law* 3, 705–737.

Ruby, C. L. and J. C. Brigham. (1998). "Can Criteria-Based Content Analysis Distinguish Between True and False Statements of African-American Speakers?" *Law and Human Behavior* 22, 369–388.

Samuelsson, C. and A. Voutilainen. (1997). "Comparing a Linguistic and a Stochastic Tagger," Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics, Madrid, Spain, 246–253.

Sapir, A. (1987). "The LSI Course on Scientific Content Analysis (SCAN)," Laboratory For Scientific Interrogation, Phoenix, AZ.

Smith, N. (2001). "Reading Between the Lines: An Evaluation of the Scientific Content Analysis Technique (SCAN)," *in* C. F. Willis (ed.), *Policing and Reducing Crime Unit: Police Research Series*, London: Crown.

Sporer, S. L. (1997). "The Less Travelled Road to Truth: Verbal Cues in Deception Detection in Accounts of Fabricated and Self-Experienced Events," *Applied Cognitive Psychology* 11, 373–397.

Steller, M. and G. Köhnken. (1989). "Criteria-Based Content Analysis," *in* D. C. Raskin (ed.), *Psychological Methods in Criminal Investigation and Evidence*. New York: Springer Verlag, 217–245.

Undeutsch, U. (1989). "The Development of Statement Reality Analysis," *in* U. Undeutsch (ed.), *Psychological Methods in Criminal Investigation and Evidence*. The Netherlands: Kluwer, Dordrecht, 101–121.

Voutilainen, A. (2000). "Helsinki Taggers and Parsers for English," *in* J. M. Kirk (ed.), *Corpora Galore: Analysis and Techniques in Describing English*. Amsterdam and Atlanta: Rodopi.

Vrij, A. (2000). *Detecting Lies and Deceit: The Psychology of Lying and the Implications for Professional Practice*. Chichester, England, New York: John Wiley Inc.

Vrij, A., W. Kneller, and S. Mann. (2000). "The Effect of Informing Liars about Criteria-Based Content Analysis on their Ability to Deceive CBCA-Raters," *Legal and Criminological Psychology* 5, 57–70.

Wheeler, R. and S. Aitken. (2000). "Multiple Algorithms for Fraud Detection," *Knowledge-Based Systems* 13, 93–99.

Wiener, M. and A. Mehrabian. (1968). "*Language Within Language: Immediacy, A Channel in Verbal Communication*," New York: Appleton-Century-Crofts.

Wolz, U., H. Walker, J. Palme, P. Anderson, Z. Chen, J. Dunne, G. Karlsson, A. Laribi, S. Mannikko, and R. Spielvogel. (1997). "Computer-Mediated Communication in Collaborative Educational Settings," *ACM SIGCUE Outlook* 25, 51–68.

WorldLingo, *http://www.worldlingo.com/resources/language_statistics.html*.

Zhou, L., Q. E. Booker, and D. Zhang. (2002). "ROD – Toward Rapid Ontology Development for Underdeveloped Domains," Proceedings of the 35th Hawaii International Conference on System Sciences, Big Island, Hawaii, Jan. 7–10.

Zuckerman, M., B. M. DePaulo, and R. Rosenthal. (1981). "Verbal and Nonverbal Communication of Deception," *in* L. Berkowitz (ed.), *Advances in Experimental Social Psychology*. New York: Academic Press, 1–59.