Are you Lying Now?

A Linguistic Examination of Deceptive Utterances in Online Conversation

Honors Thesis

Presented to the College of Agriculture and Life Sciences, Department of Communication

of Cornell University

in Partial Fulfillment of the Requirements for the

Research Honors Program

by

Barrett E. Amos

May 2008

Jeffery T. Hancock
Abstract

Extensive research has been done to identify linguistic cues to deception, especially in the rapidly growing field of computer-mediated communication. However, most past research contains an important methodological flaw: the failure to break down deceptive and truthful topics into individual utterances. When assigning research participants to a deceptive role (say, asking them to lie to an unknowing receiver) previous studies have generally given the deceivers a topic to lie about and then asked them to go ahead and communicate with their partner. The deceivers were assumed to be lying whenever they were talking about the topic on which they were supposed to deceive their partners. However, in practice not all of the utterances within a deceptive conversation topic are lies. Some are truths used to support the overall lie. Past research has failed to make this distinction, drawing into question previous findings on linguistic cues. This study sought to validate four of the more well-known linguistic cues to deception by examining them at the utterance level. The results reveal that while there often is a distinction between the linguistic cues at the overall topical level (deceptive topics vs. truthful topics), those markers do not always hold true at the utterance level. Even more interestingly, there is often a large difference between truths told in support of an overall deceptive topic and truths told in support of an overall truthful topic. These findings open up new areas of research into the linguistic cues to deception.
Are You Lying Now? A Linguistic Examination of Deceptive Utterances in Online Conversation

People lie every day. And they lie a lot. In a recent diary study of everyday lies (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996), people were deceptive in a whopping 31 percent of their social interactions. That means, on average, individuals reported deceiving their communication partner(s) in one out of every three conversations. In a world where deception is so frequent, it is important to be able to seek out and identify the deception wherever it occurs in everyday interactions.

The Nature of Deception

Social interactions are the basis of human communication, and deception is a prevalent and complex phenomenon that plays a significant role in these interactions. Previous research suggests that as many as one third of all interactions contain some form of deception (DePaulo, 1996), which can be defined as a deliberate attempt to create a false belief in a communication partner (Vrij, 2000). That attempt, whether successful or not, must not have given the partner forewarning of the deception. This definition precludes untrue utterances such as joking, sarcasm, and irony as forms of deception because the sender means for the receiver to detect the non-truth of the message. Messages that are unintentionally misleading are also not described as deceptive, but are instead simply mistakes or errors. While some researchers do draw a distinction between deception and lies (Nyberg, 1993), this study will consider the terms interchangeable so long as they fit the above definition.

The importance of being able to detect and distinguish deceptive from non-deceptive communication is clearly of practical interest. Law enforcement, government agencies, the military, and even businesses could all benefit from any clues that would help them evaluate the truthfulness of a given message. Fortunately, substantial research over the past 50 years has gone into determining what
cues a deceiver unintentionally leaks. These cues can be verbal or non-verbal, but tend to focus on “psychological leakage,” the non-strategic behaviors that are assumed to betray the deceiver’s intentions and true feelings or beliefs (Vrij, Edward, Roberts, & Bull, 2000). Along similar lines, the emotional cues a deceiver displays are considered to be indicative of detection apprehension (Vrij, 2000; Ekman, 2001). For a good overview of cues specific to deception, see DePaulo et al’s recent meta-analysis (DePaulo, Lindsay, Malone, Muhlenbruck, Charlton, & Cooper, 2003).

The main problem with the research examining these cues is that few indicators remain invariant across all situations, cultures, genres and other communication contexts. However, it does seem that a small set of these indicators can have good predictive ability in specific contexts (Vrij et al, 2000). Since non-verbal cues are arguably harder to control, most deception detection focuses on the small but systematic non-verbal differences between deceivers and truth tellers (Burgoon, Blair, Qin, & Nunamaker, 2003; Hancock, Curry, Goorha, & Woodworth, 2008). Of primary interest in this study are the verbal or linguistic properties of deceit which have been shown to be significant (Newman, Pennebaker, Berry, & Richards, 2003; Hancock, Curry, Goorha, & Woodworth, 2004).

The Linguistics of Deception

The body of literature on the linguistics of deceit is relatively small. However, there are four main types of linguistic cues that have been associated with deception: 1) Word counts, 2) Pronoun usage, 3) Words pertaining to emotions, and 4) the Use of exclusionary words (Burgoon, Buller, Floyd, & Grandpe, 1996; Burgoon, Blair, Qin, & Nunamaker, 2003; Newman, Pennebaker, Berry, & Richards, 2003; Pennebaker, Mehl, & Neiderhoffer, 2003; Hancock, Curry, Goorha, & Woodworth, 2004).
Word Count. Much of the traditional research suggests that deceivers will offer fewer details about themselves when lying, and thus will reduce their overall word count (Burgoon et al, 2003; DePaulo et al, 2003; Porter & Yullie, 1996; Vrij, 2000).

However, recent research has suggested that deceptive communication actually contains more words if conducted in an asynchronous or near-synchronous computer-mediated communication where participants send each other text-based messages (Zhou, Burgoon, Nunamaker, & Twitchell, 2004). Three factors may explain this increased word usage. First, because of the asynchronous and text-based nature of the CMC medium, participants may have been able to take longer to plan and edit their messages. Such planning is not available in synchronous communication, allowing the participants using asynchronous communication to create better stories (Clark, 1996; Hancock & Dunham, 2001). Second, most previous research has focused around deceivers lying about discernable facts. Zhou et al. (2004) instead asked deceivers to lie about a non-verifiable opinion. Thus liars may be willing to speak longer and more verbosely if they don’t have to be as careful about contradicting their own facts. Finally, the nature of the interaction between liars and their partners may play a role. If the interaction is of a conversational nature, deceivers may use more words to manage information flow, enhance mutuality with their partners, and to decrease their partners’ level of suspicion (Burgoon, Buller, & Floyd, 2001).

Based on these three factors, deception in text-based CMC should be more verbose (more words, more messages, and longer average message length) than its truthful counterpart.

Pronoun Use. In a series of five studies Newman et al (2003) observed that deceivers tend to use first-person pronouns much less frequently when lying than when telling the truth. The use of first person singular pronouns ("I," "my," "me," etc.) is thought to be a subtle proclamation of one’s ownership of a statement. Knapp, Hart, and Dennis (1974) hypothesized that liars would avoid statements of
ownership to distance themselves from their words due to a lack of personal experience. Wiener and Mehrabian (1968) also suggested that liars should be more non-immediate and reference themselves less in their stories than truth-tellers. Thus, liars should use fewer first-person singular pronouns than their truthful counterparts.

The data for second- and third-person pronoun use is less consistent. Some studies have found that liars also use fewer second- and third-person pronouns (Newman et al, 2003), while other have found them to use more (Ickes, Reidhead, & Patterson, 1986; DePaulo et al, 2003). At this point, the data is inconclusive.

Emotional Words. In general, it is assumed that liars feel guilty about deceiving their conversation partners (Knapp and Comadena 1979; Vrij, 2000). Research has suggested that such guilt manifests itself in the deceiver using more disparaging statements and negative emotion words such as “hate,” “worthless,” and “sad” during the course of a conversation (Newman et al, 2003; Vrij, 2000; Zhou et al, 2004). These findings are consistent with the work of Burgoon et al. (2003) who found that deceivers tend to be more expressive than their truthful counterparts.

Exclusionary Words. Finally, Newman et al. (2003) predicted that the process of creating a false story should consume cognitive resources, leading liars to tell less complex stories than if they were telling the truth. Thus, liars are likely to focus on what they did and what they did not do. They should be wary of using any distinction markers (i.e. exclusive words — “but,” “without,” “except,” etc.) which require a deceiver to be more specific and precise. The more specific and precise deceivers are in the details, the more likely they are to be caught in a contradiction. Thus, Newman et al. (2003) predicted that deceivers would use fewer exclusionary words. Newman et al. found support for this hypothesis in a dialogue study, however, Hancock et al. (2008) did not find support in a conversational setting.
Turn-based Analysis

As noted above, previous research on the linguistics of deception has almost entirely been limited by study methodology. When assigning research participants to a deceptive role (say, asking them to lie to an unknowing receiver) previous studies have generally given the deceivers a topic to lie about and then asked them to go ahead and communicate with their partners. The deceivers were then assumed to be lying whenever they were talking about the topic on which they were supposed to be deceiving their partners (Newman et al., 2003; Hancock et al, 2004; Burgoon et al, 2003). But Hirschberg et al. (2005) point out that not all of the utterances within a deceptive conversation topic are going to be lies. Some are going to be truths used to support the overall lie. Hirschberg terms this distinction “Big Lie” (the overall topical lie) "Little Lie" (the deceptive messages told in support of a topic). Along similar lines, not all of the utterances in truthful (honest) topic will be true; some will also be deceptive. Based on this classification, there are four types of utterances in the context of telling lies and truths:

1) Deceptive utterances that serve an overall lie topic \( (tLuD) \)
2) Truthful utterances that serve an overall lie topic \( (tLuT) \)
3) Deceptive utterances that serve an overall honest topic \( (tHuD) \)
4) Truthful utterances that serve an overall honest topic \( (tHuT) \)

See Figure 1 for a graphical representation of the above based on Hirshberg et al.‘s (2005) data. For the sake of clarity, also note the parenthetical abbreviation for each of the above utterance types. The notation will be used throughout the rest of this paper to refer to each of the four utterance types. This list has also been repeated in Figure 2 for quick reference.
Previous research grouped the Little Truths and Little Lies in with the Big Lies and the Big Truths and ran automated linguistic analyses against the topics as a whole. This could prove problematic, because what if the Little Truths are hiding significant linguistic differences between truthful and deceptive utterances by adding noise to the deceptive data? Or, conversely, what if there is no significant difference in language use between deceptive and truthful utterances, but there is a difference between truthful utterances told in support of lie topics and truthful utterances told in support of honest topics?

In the 2005 paper, Hirschberg et al. found initial evidence that suggests approximately 40% of utterances within lie topics are actual truthful, while the remainder were deceptive utterances. However, their data was based on face-to-face interviews and did not occur in a conversational setting. The present study seeks to look at the linguistic differences between deception and truth at the level of individual utterances, rather than at the more traditional topical level.

Deception and CMC

The ability to look at these individual utterances is an area in which Computer-Mediated Communication (CMC) excels. Instant Messaging (IM) is a recent CMC medium that has become quite popular for communicating online. IM is a nearly synchronous medium of communication, making it an ideal medium for studying deception because conversations can occur quickly (in almost real time), and in a manner similar to that of traditional face-to-face interactions (Nardi, Whittaker, & Bradner, 2000). IM also forces the production of a single text-based message at a time. Each message is composed independently of the others, without input or feedback from other members of the conversation during its composition. This allows us to look at entire conversations on a message-by-message level rather than at a purely conversational or topical level.
In general, an IM message will contain one “idea unit” of text (a ‘message’ or ‘utterance’). In normal conversation an idea unit would be a single sentence. However, in IM it can be either an incomplete sentence or multiple sentences (i.e. “Really? I thought...” or “funny”). Each of these thought units are transmitted by a single instant message, creating an utterance. Although it is possible for multiple units to be contained within a single message/utterance, in general experience this is rarely the case. Thus, an individual message (generally a single thought or turn in a conversation) can be used to analyze occurrences of deception at a much finer granularity than has traditionally been possible.

**Present Study**

The present study seeks to replicate previous studies on the linguistics of deception in the hope that, when analyzed on a finer level, the linguistic indicators still hold true. To this end, four well-supported hypotheses have been selected from previous research to be examined at the utterance level. These four hypotheses are:

**H1:** Participants will use more words when telling deceptive statements than when telling truthful statements.

This hypothesis is based on the research of Zhou et al. (2004) and Hancock et al. (2008). The following hypothesis is derived from Newman et al.’s (2003) work on pronoun use:

**H2:** Participants will use fewer first-person pronouns when telling deceptive statements than when telling truthful statements.

The final two hypotheses were also derived from Newman et al.’s (2003) work:

**H3:** Participants will use more negative emotion words (i.e. “hate,” “worthless,” and “sad”) when telling deceptive statements than when telling truthful statements.

**H4:** Participants will use fewer exclusionary words (i.e. “but,” “without,” and “except”) when telling deceptive statements than when telling truthful statements.
Combined, these four hypotheses should give a good idea of whether these linguistic cues, which have all been supported at the topical level of deception, still hold true at the finer utterance level.

Finally, since this study is examining the phenomenon of deceptive linguistics from a relatively new angle, it will also look at the following general question:

**RQ1**: How do linguistic markers occur at the utterance level?

All combined, these four hypotheses and single open-ended question should allow for a deeper understanding of how linguistic cues and markers behave in a text-based CMC setting at both the topical and utterance level of deceptive communication.

**Method**

**Participants**

Participants \( N = 70 \) were students at a northeastern American university (male = 26, female =40, 4 unreported), recruited from undergraduate communication courses for extra-credit. They were recruited for a “study of how people communicate online.” Participants were randomly paired to form 35 unacquainted dyads. Three dyads (6 participants) were excluded because the participants failed to follow instructions.

**Procedure**

Upon reporting to the laboratory, participants were asked to sign an informed consent form and led separately to remote rooms.

The general procedure was adapted from Burgoon, Buller, and Floyd (2001). In this study, all participants were told that they would be having a conversation with an anonymous partner. The
participants were then instructed that they would discuss five topics, which were provided on a sheet of paper. The first topic was always a simple introduction designed to allow the participants to become comfortable interacting with their partners, and was not included in any analysis. After this topic, participants began a discussion of the four experimental topics that included the following: “Discuss a quality that you possess that will most likely lead to fulfillment and success in your academic performance,” “Discuss a quality that you possess that will most likely lead to fulfillment and success in your career,” “Discuss a quality that you possess that will most likely lead to success in making new friends,” and “Discuss a quality that you possess that will most likely lead to fulfillment and success in your family life.” Note that the topics were selected based on a protocol developed by Vohs, Baumeister, and Ciarocco (2005). There was no time limit, and participants were asked to discuss each topic until they had exhausted it and understood each other’s responses.

In each dyad, one of the two participants was randomly assigned to the role of liar, and the other to the role of partner. Liars were asked to deceive their partners. In particular, they were instructed to “NOT tell ‘the truth, the whole truth, and nothing but the truth’” on two of the topics, and to be truthful on the other two topics. The two topics in which the whole truth was not to be told were marked in bold on the liar’s topic sheet. It was emphasized that the liar should try to produce large magnitude lies rather than small, white lies in the deceptive topics. Partners were blind to the deception manipulation and were given the same list of topics, without any marked in bold, and were asked to simply make the conversation go as smoothly as possible.

The participant assigned the liar role was motivated by the chance to win $100 if they were able to convince their partners of their lies. In addition, the liars had the following false statement printed on their instruction sheet:
You should remember that the ability to deceive your partner in these conversation tasks represents an extremely important skill. Previous research has shown that the ability to lie predicts future success in business, counseling, and health. It also predicts your ability to make and maintain friends, and is a good indicator of intelligence that is not measured by conventional IQ tests. As such, you should try your best to completely deceive your partner.

These two incentives were designed to increase the liars’ motivation to deceive their partners based on previous research (DePaulo, Lanier, & Davis, 1983; Forrest & Feldman, 2000).

The sequence in which the topics were discussed, and the order in which the liar lied, was counterbalanced across 16 orders. After the initial introduction topic, liars were instructed to lie on either the next two topics or on the last two topics.

Participants performed the task at isolated computer terminals. Participants used one of two desktop computer stations while the experimenter monitored and recorded the interaction from a third station. Once participants were seated at their terminals, the experimenter briefly demonstrated the use of the computer interface, AOL Instant Messenger, in which participants typed their messages in a private composition window and hit enter to send their messages to a shared window. Note that participants could edit their messages before transmitting it to their partner.

Once participants finished the discussion task, they were asked to complete a series of questionnaires based on their conversation, which included items assessing how truthful the liar had been and how truthful the partner believed the liar had been. In addition, participants were given a printed transcript of their conversation. The participants who had been the liar were asked to go back and mark each individual IM message in which they had actually lied, while the participants who had been the partner were asked to mark any message in which they believed the liar had been deceptive.
After completing these tasks, the participants were brought into a common room where they were introduced to their partners and fully debriefed.

**Automated Linguistic Analyses**

Conversation transcripts were converted into text files and each of the liars’ messages were coded for topic (whether the liar was supposed to be deceiving or telling the truth), truthfulness (whether that message contained a lie or not based on the liar’s self-report), and focus (was the liar talking about himself or his partner). For this study, data from the conversation partner were ignored. In addition, data from messages where the liar was not talking about himself (i.e. the message was coded with a focus was on the partner) were also ignored to make the results more comparable to previous studies on deception (see Newman et al, 2003; Zhou et al, 2004; DePaulo et al, 1996).

The transcript data were then analyzed using the Wmatrix program (Rayson, 2008). This Natural Language Processing program is a software tool for corpus analysis and comparison, providing an interface to the USAS and CLAWS corpus annotation tools. In addition, it is able to report standard corpus linguistic methodologies such as part-of-speech and semantic tagging, frequency analysis, and concordances. The program allows for easy statistical comparison between two English-language datasets. To make the results of the current study more directly comparable to previous deception research, the default Wmatrix semantic lexicon was not used. Instead, a custom lexicon was loaded based on the Linguistic Inquiry and Word Count program (Pennebaker, Francis, & Booth, 2001). Wmatrix is able to use this custom lexicon to report the log-likelihood statistical comparison between two corpora (see Rayson, Berridge, & Francis, 2004; Rayson & Garside, 2000).
Liar’s conversations were broken down into six separate corpora. The first two corpora divided the message data along topic lines, separating the data into a collection of messages where the liars were supposed to deceive their partners and a collection where they were supposed to tell the truth. This “topical” division is consistent with previous deception research (i.e. Newman et al, 2003; Hancock et al, 2004). These topical corpora are further broken down to the message level: Deceptive Messages within Honest Topics, Truthful Messages within Honest Topics, Deceptive Messages within Lie Topics, and Truthful Messages within Lie Topics. These six corpora were then used to study deception at both the traditional topical level and at the deeper message level.

**Results**

Overall, 1788 utterances (IM messages) were sent by the participants assigned the role of liars. Once coded and corrected for message focus and topical distinctions that number dropped to 805 messages on which the following analyses were run. The following analyses are broken down into sections: the traditional topical division, utterances (messages) within the lie topic alone, and utterances across all topics. The topical division is included to show how the current study results compare with past research. All data are reported in Tables 1-4.

Also of note is the Log Likelihood (LL) statistic reported in Tables 2-4. The greater the LL value the more significant the difference between the two frequencies reported (Rayson et al., 2004). For corpora comparison the critical LL value for the 95th percentile ($p < 0.05$) is 3.84, for the 99th percentile is 6.63, and for the 99.9th percentile is 10.83 (Rayson et al., 2004).

**Across Topics (Topical Division)**

Referring to Table 1, 463 messages (5342 words) were sent under a lie topic while 342 messages (4357 words) were sent under an honest topic. In this study participants used more words (were more
verbose) when discussing lie topics than when discussing honest topics \(p < 0.05\). Participants also sent more messages when talking about lie topics, however the average message length was marginally shorter \(p = 0.062\).

Unlike previous research, the current study found no difference in first-person pronoun use nor the frequency of negative emotion word use across lie and honest topics. In fact, participants produced very few negative emotion words. However, as previous research has suggested, the current study did find that participants used more exclusionary words when discussing honest topics \(p < 0.05\).

**Within the Lie Topic**

Having looked at deception through the traditional comparison approach, the topics can now be broken down into their individual utterances. Remember, every utterance within a given topic was coded as being either deceptive or truthful by the participants. Since the intent of this study is to look at how deception works at the utterance level, it will be examined it separately.

Participants used marginally more words when sending deceptive messages than truthful messages when sent in support of an overall lie topic \(p = 0.053\). While there was no difference in the number of truthful and deceptive messages, the average message length was significantly longer for deceptive messages within the lie topic \(p < 0.001\).

Contrary to the findings at the topical level, the message-level analysis revealed that participants used significantly *more* first-person pronouns in deceptive messages than in truthful messages when told in support of an overall lie topic \(p < 0.001\). However, as with the topical division, there was no difference in the use of words signaling negative emotions. Interestingly, in contrast to the
findings at the topical level, deceptive messages used more exclusive words than did truthful messages (LL = 4.31, p < 0.05) when told in support of an overall lie topic.

**Message Level Across Topics**

Having looked at utterances told within lie topics in isolation, we can now begin comparing deceptive and truthful utterances across all topics. Overall word and message frequencies are reported in Table 1 and graphically in Figure 3. In general, there are significant differences across all four utterance types (tLuD, tLuT, tHuD, and tHuT) in terms of word and message counts, with two exceptions. As noted above, there is no difference between the number of deceptive messages sent in support of an overall lie topic and the number of truthful messages sent in support of an overall lie topic (p = 0.53). In addition, there was no difference between the number of words sent in deceptive messages told in support of overall lie topics and the number of words sent in truthful messages told in support of overall honest topics. All other utterance types were significantly different, or nearly so, across both lie and honest topics (see Table 1 for values).

In terms of first-person pronoun use, there are significant differences between deceptive and truthful messages told in support of lie topics (p < 0.01), between deceptive messages told in support of lie topics and truthful messages told in support of honest topics (p < 0.05), and between truthful messages told in either topic (honest or lie) (p < 0.0001). The last finding is of special interest because of the magnitude of the difference (LL = 22.37). It is significant at well above the 99.99th percentile, indicating a large difference in how truthful messages are told across topics. In contrast, there is no difference in first-person pronoun use in deceptive messages across topics.
Even at the utterance level, the current study finds no difference in negative emotion word frequency across topics or within topics.

As previously noted, there is a significant difference ($p < 0.05$) in exclusionary word frequency between deceptive and truthful messages told in support of an overall lie topic. Likewise, there is also a large difference between deceptions and truths told in support of an honest topic ($p < 0.01$). In addition, there is a large difference ($LL = 12.00, p < 0.001$) between deceptions told in support of a lie topic and truths told in support of a honest topic. Even more interesting is that, like in the first-person pronoun findings above, there is a significant difference between truthful messages told in either topic ($LL = 26.90, p < 0.001$). Again, there is no difference between deceptive messages across topics.

**Discussion**

The goal of the present study was to test the findings of previous research on linguistic cues to deception at the utterance level. By examining the linguistic cues present in individual messages in a controlled CMC-based experiment, it is hoped that a better overall picture of how text-based deception occurs will emerge. The results show that not only are the cues are not always significant at the traditional topical level, but they can even show the opposite effect when the topics are further broken down into truthful and deceptive utterances.

**Word Count**

Hypothesis 1 predicted that deceptive messages would be more verbose than their truthful counterparts. Consistent with previous research, this study found that indeed, at the topical level deceivers do use more words and send more messages. More interesting is the significant difference between truths and deceptions when looking only at messages within the lie topic. Remember, traditionally these messages would have all been grouped together as “deceptive,” but now that
ground-truth information at the individual utterance level is available, they can be examined more closely. In fact, the difference is significant on two of the three verbosity indicators (word count and average message length), indicating that deceptive messages when told to support an overall lie are likely to be both wordier and longer than truthful messages told to support an overall lie.

It gets more interesting when honest topics are factored in. As expected, there are many more truthful messages and words told in support of the overall honest topic. However, the average message length of a \( tHuD \) message is still significantly longer than the average length of a \( tHuT \) message. Thus, the average message length, regardless of topic, is a good indicator of deception at the message level (it only approached significance at the topical level, \( p = 0.053 \)). This is contrary to Hancock et al.’s (2008) findings which found that there was no difference in message length between the deceiving participant’s deceptive and truthful communication.

The most interesting finding is the differences between truthful messages across topics. There is a significant difference between truthful messages, regardless of topic, on all three indicators of verbosity. This is especially apparent when examining the average message length. The average length of a \( tLuD \) message is much closer to the average length of a \( tHuT \) message than it is to the length of a \( tHuD \) message. While the difference between a \( tLuD \) and a \( tHuT \) is still significant, it is not as large as the difference between a \( tHuT \) and a \( tDmT \).

Moreover, there is no significant difference in word count between \( tLuD \) and \( tHuT \) (although the average message length is significant). This discounts the notion that word count as a linguistic factor is due to a difference between “pure” lies (deceptions told in support of lie topics) and “pure” truths. Thus, something else must be contributing to the overall difference observed at the topical level. While the difference between deceptions told within each topic could contribute to this discrepancy, the
overall small number of deceptions told in honest topics leads us to disregard this as a major factor. Instead, it seems that there is a greater difference between truthful messages across topics, supporting the observation above. That is, the difference in word counts between honest and lie topics is mostly due to the difference in truthful utterance word counts at the message level. This, combined with the message length findings above, could indicate that there are actually linguistic differences between types of truth (truths serving deceptive topics and truths serving truthful topics). This finding will be explored further momentarily.

In general, Hypothesis 1 finds mixed support. The average length of a message is a good indicator of whether the message is deceptive or not, but word count and number of messages are not as conclusive. Word count is only indicative as a cue within a given topic (and switches direction depending on which topic) and number of messages is inconclusive overall. Thus, verbosity is a linguistic cue only if defined as the average length of a message.

First-Person Pronoun Use

Hypothesis 2 predicted that participants would use fewer first-person pronouns in deceptive messages. Unlike in previous research (Newman et al, 2003), the current study found no difference in first-person pronoun use at the topical level. However, within the lie topic, participants actually used more first-person pronouns when sending deceptive messages than when sending truthful messages in support of the overall lie. This is also contrary to previous findings (Newman et al, 2003; Hancock et al, 2008). However, the previous findings do hold true when comparing tLuD with tHuT, the “pure” lies and truths. In addition, there is no difference in first-person pronoun use within honest topics. Thus, it is not clear that first-person pronoun use is a good linguistic indicator of deception at the message level.
DePaulo et al. (2003) and Ickes (1986) theorized that deceivers would use fewer first-person pronouns in an effort to shift the focus of a lie away from themselves. Since the present study specifically asked deceivers to talk about themselves in personal situations, even during the lie topics, it may have pre-selected for first-person pronoun use. Further study with more varied conversation topics is needed to verify whether first-person pronouns are actually a good linguistic marker at the utterance level.

However, it is interesting to again note the significant difference in truthful messages between topics. Like with word count, there is a larger difference between tHuT and tLuT than there is between tHuT and tLuD. This adds more credence to the possibility that there is a significant linguistic difference between the two types of truth.

*Negative Emotion Words*

Hypothesis 3 predicted that participants would use more negative emotion words in deceptive messages. The data show no difference in the use of negative emotion words at any level, consistent with Hancock et al.’s (2008) findings, but contrary to Newman et al.’s (2003) previous work. However, the nature of the current study may have played a role in this outcome. Negative emotion words were used at a very low frequency overall in the current study (less than 0.5% across both topics) compared to previous studies (Hancock et al., 2008: a little less that 2% use across topics; Newman et al., 2003: about 1.4%). The nature of the current study’s conversational topics, all dealing with fulfillment and success in life, may have encouraged the participants to lean more towards positive emotions. Hypothesis 3 thus remains unsubstantiated at the utterance level.
Exclusionary Words

Finally, Hypothesis 4 predicted that fewer exclusionary words would be used in deceptive messages. This is supported at the topical level, consistent with Newman et al.’s (2003) findings. However, within the lie topic the opposite effect is observed: deceptive messages included more exclusionary words than did their truthful counterparts. And, much like in first-person pronoun use, that finding is reversed when looking at deceptive and truthful messages within an overall truthful topic. Thus, Hypothesis 4 does not have support at the utterance level.

It is interesting to once again note the significant difference in truthful messages between topics. Like with word count and first-person pronouns, there is a larger difference between \( tHuT \) and \( tLuT \) than there is between \( tHuT \) and \( tLuD \). There is no difference in deceptive messages across topics.

Truthful Messages

In three of the four hypotheses examined in this study it was found that there was a significant difference in truthful utterances between lie and honest topics. In all three cases \( tHuT \) were closer in value to \( tLuD \) than they were to the \( tLuT \) used to support deceptive statements. Also, in two of the cases there was no difference in deceptive utterances between topics while there existed an overall difference between honest and deceptive topics at the topical level. This brings up an interesting possibility: that the observed cues in the language of deception are more often due to differences in the language of truthful statements rather than deceptive ones. Such a finding would contradict previous research, which has, almost exclusively, focused on deceptive text to find the linguistic markers of deception. Of course, such deceptive text has always been at the topical level and thus inadvertently included the truthful text as well.
If linguistic markers are mostly present in truthful messages told in support of a lie topic, rather than in deceptive messages told in support of a lie topic, much of the existing theorization would be incorrect. For example, the current prevailing theory of first-person pronoun use is that individuals try to cognitively distance themselves from a lie (Newman et al, 2003). However, it seems that individuals are actually trying to distance themselves from truths told in support of a lie. One possible explanation for this action is that instead of distancing themselves from the lie, as the current theory assumes, they instead distance themselves from the truth associated with that lie in order to reduce the mental tension between telling lie-truths and honest-truths. The psychological distancing that has been demonstrated in past research (Buller, Burgoon, Buslig, & Roiger, 1996; Knapp & Commadena, 1979) has now shifted away from the distinction between lies and truths, which should be relatively clear in the individual's mind, to keeping the distinction between lie-truths and honest-truths clear which are likely to be much fuzzier. By keeping this distinction clear, individuals will be lowering the level of cognitive complexity inherent in the overall lie and thus will be better able to focus cognitive resources on other aspects of the deception (Vrij, 2000). Additional research is necessary to test this possibility.

Limitations of the Current Study

The current study has three major limitations. First, it has a relatively small number of participants compared to similar CMC-based deception studies. Once the liar’s partner is factored out and the data are limited to messages focused on the liar, the total corpus drops down to just a few thousand words. Increasing this sample size would likely yield a clearer picture of deceptive utterances. Second, the topics selected for conversation may well be pre-disposed to certain attitudes. As was mentioned in the discussion of negative emotion word frequencies, each of the topics dealt primarily with success and fulfillment: positive emotions. Thus, the discussion is much more pre-disposed to positive emotion
words than to negative ones. It would be interesting to pick more neutral topics and re-run the study to see if the same results are found for negative emotion word frequency. Finally, no conversation exists in isolation. In this study only the “liar” participant data are examined, and only those messages which were coded as having a “self” focus. This precludes a number of messages, and also hides any Linguistic Style Matching (LSM) effects that may play a role in the deceptive and truthful messages. Research has shown that individuals engaged in conversation often attempt to match their partner’s linguistic style in order to appear more credible and more attractive, and thus more persuasive (Burgoon, Stern, & Dillman, 1995; Giles & Coupland, 1991; Niederhoofer & Pennebaker, 2002). By limiting the current study to only one side of the interaction, we miss any sort of interaction effect that may be present (i.e. Hancock et al, 2008).

Conclusion
This research advances our understanding of how linguistic cues and markers behave in a synchronous CMC setting at both the topical level and at the utterance level. Understanding deception at this level and in this medium is becoming increasingly important as the internet becomes more ubiquitous around the world. With such instant communication available across borders, law enforcement, military, and business uses abound for automated deception detection. This research helps take us one step closer to making such detection possible through linguistic analysis of communicative text.
References


Hancock, J. T., & Dunham, P. J. (2001). Language use in computer-mediated communication: The role


Appendix A

Conversing Online Study

Consent Form

[For Participants in the Liar (Sender) Condition]

You are invited to participate in a research study on interpersonal communication patterns. We ask that you read this form and ask any questions you may have before agreeing to be in the study.

Background Information: The purpose of this study is to understand how people communicate online. We will be examining how people converse about several different topics.

Procedures: If you agree to be in this study, we will ask you to do the following:

1. Engage in four topic discussions with another student (each discussion may take up to 5-15 minutes)
2. You will be asked to lie in some of the conversation tasks that take place in the study. In particular, you will be asked to not "tell the truth, the whole truth, and nothing but the truth" during parts of the conversation.
3. You will complete some forms (may take up to 30 minutes)
4. You will be debriefed, and you will have an opportunity to ask any questions you may have about the study.

The total time for this experiment is usually about 1 hour, but may take up to 1 hour and 30 minutes.

Risks and Benefits of being in the Study: We do not anticipate any risks for you participating in this study, other than those encountered in day-to-day life. However, some people may feel uncomfortable being deceptive during their conversation.

There are no direct benefits associated with participating. One indirect benefit for participating is learning about how studies examining interpersonal communication are conducted.

Compensation: You will receive class credit for participation.

Voluntary Nature of Participation: Your decision whether or not to participate will not affect your current or future relations with the University or the Communication department. If you decide to participate, you are free to withdraw at any time without affecting those relationships. If you decide not to participate after signing up or during the study, notify your coordinator and you will be free to leave the study (and still receive your credit).

Confidentiality: The records of this study will be kept private. In any sort of report we might publish, no information identifying you will be available. Research records will be kept in a locked file; only the researchers will have access to the records. Transcripts of the online discussions do not contain any identifying information about you. Note, if you communicate with us via email, there is a chance that your email could be read by a third party.

Contacts and Questions: The primary researchers conducting this study are Professor Jeff Hancock, Natalie Bazarova, and Barrett Amos. Please ask any questions you have now. If you have questions later, you may contact them via email at jth34@cornell.edu or mnb8@cornell.edu or bea4@cornell.edu, or by phone, at 607-255-4452. If you have any questions or concerns regarding your rights as a participant in this study, you may contact the Institutional Review Board for Human Participants (IRB) / www.irb.cornell.edu <http://www.irb.cornell.edu/> / irbhp@cornell.edu / 607-255-5138. You will be given a copy of this form to keep for your records.

Statement of Consent: I have read the above information, and have received answers to any questions I asked. I consent to participate in the study. I also agree for my conversations to be recorded for research purposes only.

Signature ___________________________ Date ______________________

This consent form will be kept by the researcher for at least three years beyond the end of the study and was approved by the IRB on October 10, 2007
Appendix B

[Sender] Instructions

Sometimes situations arise where it is NOT in one’s best interests to tell ‘the truth, the whole truth, and nothing but the truth’—for instance, to present your best image, to protect another’s feelings, to avoid unpleasant circumstances. A certain level of communication skill is necessary to be able to adapt to these situations. We would like you to practice these skills in the upcoming discussion and to allow us to determine how well others can detect these deceptions. **If you are able to convince the partner of your lies, you will be entered in a raffle for a $100 prize.**

After introducing yourself, you will discuss four different topics. The topics are numbered in the order in which we would like you to discuss these topics.

- You have to be truthful when answering the questions where indicated to “TELL THE TRUTH”.
- You have “to NOT tell the truth, the whole truth, and nothing but the truth” when answering the questions where indicated to “LIE”.

You might give clear but completely untrue answers. You might be vague, indirect, unclear, and ambiguous, or you might withhold, omit, or avoid discussing relevant information. The primary aim is to use your own technique and communication style to give answers which fall short of being the ‘truth, the whole truth, and nothing but the truth.’

*You should remember that the ability to deceive your partner in these conversation tasks represents an extremely important skill. Previous research has shown that the ability to lie predicts future success in business, counseling, and health. It also predicts your ability to make and maintain friends, and is a good indicator of intelligence that is not measured by conventional IQ tests. As such, you should try your best to completely deceive your partner.*

Your conversation will be with an anonymous partner and will take place over a computer network (much like an Internet chat room). The experiment will give you quick instructions on how to use the computer to interact with your partner.

You should discuss each conversation topic until you are both clear on each other’s viewpoints. Typically, this takes about 5-10 minutes per conversation topic, although there is no time-limit for your discussion. Once you have finished discussing one topic, please write “FINISHED” and move onto the next topic with your partner.

When you have completed all of the conversation topics, you will have several questionnaires to complete, which will take about 15-30 minutes.

The Experimenter will go over these instructions with you before the study begins. Please wait until the Experimenter instructs you and your partner to begin before starting your conversation.
Appendix C

[Receiver] Instructions

After introducing yourself, you will converse with a partner on four different topics listed on the sheet supplied by the Experimenter. The topics are numbered in the order in which we would like you to discuss these topics.

We are interested in how people converse about different types of topics in computer-mediated interactions. Your instructions are to try to make the conversation go as smoothly as possible.

You should discuss each conversation topic until you are both clear on each other’s viewpoints. Typically, this takes about 5-10 minutes per conversation topic, although there is no time-limit for your discussion. Once you have finished discussing one topic, please write “FINISHED” and move onto the next topic with your partner.

Your conversation will be with an anonymous partner and will take place over a computer network (much like an Internet chat room). The Experimenter will give you brief instructions on how to use the computer to interact with your partner.

When you have completed all of the conversation topics, you will have several questionnaires to complete, which will take about 15-30 minutes.

The Experimenter will go over these instructions with you before the study begins. Please wait until the Experimenter instructs you and your partner to begin before starting your conversation.
Appendix D

Conversation Topics [Counterbalanced]

After introducing yourself, please discuss the following topics in the order that they are presented below. You should discuss each conversation topic until you are both clear on each other’s viewpoints. Typically, this takes about 5-10 minutes per conversation topic, although there is no time-limit for your discussion. Once you have finished discussing one topic, please write “FINISHED” and move onto the next topic with your partner.

Introduction
Take a few minutes to introduce yourself to your partner. Say a few words about where you are from, what you major is, what year you are in, etc.

Topic I
Discuss a quality that you possess that will most likely lead to fulfillment and success in your academic performance.

Topic II
Discuss a quality that you possess that will most likely lead to fulfillment and success in your career.

Topic III
Discuss a quality that you possess that will most likely lead to success in making new friends.

Topic IV
Discuss a quality that you possess that will most likely lead to fulfillment and success in your family life.
Appendix E

Post-Interaction Questionnaire [Receiver]

The following questionnaires are designed to assess the degree to which you are able to determine whether your partner was telling “the truth, the whole truth, and nothing but the truth” for each particular topic, as well as to gather some information on how you went about determining whether your partner was attempting to deceive you or not.

There are 5 questionnaires that follow. The first four are specific to each conversation topic, while the last questionnaire contains general questions. You will only be answering questions about the four conversation topics, after the introduction. The Experimenter will allow you to review each conversation topic, in order to help you answer these questions.

For each questionnaire, please read each item carefully and circle the number that best describes your response. Describe yourself, your behaviors and your feelings as honestly and as accurately as possible.

Conversation Topic #1

Please complete these questions after reviewing your discussion of the first conversation topic, after the introduction. Base your answers on your discussion of this topic only.

1. To what degree do you believe your partner was truthful about this topic (circle the most appropriate response)?

| Not at all truthful | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Completely truthful |

2. If you had to make a decision, was your partner LYING or TRUTHFUL (circle one)?

Please indicate how much you agree or disagree with the following statements by circling the most appropriate choice.

3. I am confident in my judgment about whether my partner was truthful or not.
4. This part of the discussion went smoothly.

5. I felt comfortable during this part of the discussion.

6. I truly got to know my partner during this part of the discussion.

Now turn to the transcript and identify every utterance in this part of the discussion that you thought was deceptive by circling the time stamps.

Conversation Topic #2

Please complete these questions after reviewing your discussion of the second topic. Please base your answers on your discussion of this topic only.

1. To what degree do you believe your partner was truthful about this topic (circle the most appropriate response)?
2. If you had to make a decision, was your partner LYING or TRUTHFUL (circle one)?

Please indicate how much you agree or disagree with the following statements by circling the most appropriate choice.

3. I am confident in my judgment about whether my partner was truthful or not.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

4. This part of the discussion went smoothly.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

5. I felt comfortable during this part of the discussion.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

6. I truly got to know my partner during this part of the discussion.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
Now turn to the transcript and identify every utterance in this part of the discussion that you thought was deceptive by circling the time stamps.

**Conversation Topic #3**

Please complete these questions after reviewing your discussion of the third topic. Please base your answers on your discussion of this topic only.

1. To what degree do you believe your partner was truthful about this topic (circle the most appropriate response)?

   Not at all truthful 0 1 2 3 4 5 6 7 8 9 10 Completely truthful

2. If you had to make a decision, was your partner LYING or TRUTHFUL (circle one)?

   Please indicate how much you agree or disagree with the following statements by circling the most appropriate choice.

3. I am confident in my judgment about whether my partner was truthful or not.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

4. This part of the discussion went smoothly.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
5. I felt comfortable during this part of the discussion.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

6. I truly got to know my partner during this part of the discussion.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

Now turn to the transcript and identify every utterance in this part of the discussion that you thought was deceptive by circling the time stamps.

**Conversation Topic #4**

Please complete these questions after reviewing your discussion of the fourth topic. Please base your answers on your discussion of this topic only.

1. To what degree do you believe your partner was truthful about this topic (circle the most appropriate response)?

<table>
<thead>
<tr>
<th>Not at all truthful</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Completely truthful</th>
</tr>
</thead>
</table>

2. If you had to make a decision, was your partner LYING or TRUTHFUL (circle one)?

Please indicate how much you agree or disagree with the following statements by circling the most appropriate choice.
3. I am confident in my judgment about whether my partner was truthful or not.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

4. This part of the discussion went smoothly.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

5. I felt comfortable during this part of the discussion.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

6. I truly got to know my partner during this part of the discussion.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

Now turn to the transcript and identify every utterance in this part of the discussion that you thought was deceptive by circling the time stamps.

General Questions

Please indicate how much you agree or disagree with the following statements by circling the most appropriate choice.

1. I find it hard to imitate the behavior of other people.
2. At parties and social gatherings, I do not attempt to do or say things that others will like.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

3. I can only argue for ideas which I already believe.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

4. I can make impromptu speeches even on topics about which I have almost no information.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

5. I guess I put on a show to impress or entertain others.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

6. I would probably make a good actor.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

7. In a group of people I am rarely the center of attention.
8. In different situations and with different people, I often act like very different persons.

9. I am not particularly good at making other people like me.

10. I’m not always the person I appear to be.

11. I would not change my opinions (or the way I do things) in order to please someone or win their favor.

12. I have considered being an entertainer.

13. I have never been good at games like charades or improvisational acting.
14. I have trouble changing my behavior to suit different people and different situations.

Strongly Disagree  1  2  3  4  5  6  7  Strongly Agree

15. At a party I let others keep the jokes and stories going.

Strongly Disagree  1  2  3  4  5  6  7  Strongly Agree

16. I feel a bit awkward in public and do not show up quite as well as I should.

Strongly Disagree  1  2  3  4  5  6  7  Strongly Agree

17. I can look anyone in the eye and tell a lie with a straight face (if for a right end).

Strongly Disagree  1  2  3  4  5  6  7  Strongly Agree

18. I may deceive people by being friendly when I really dislike them.

Strongly Disagree  1  2  3  4  5  6  7  Strongly Agree

Finally, please answer some demographic questions:
Are you MALE or FEMALE (circle one)

What is your AGE (in years)? __________

Is English your FIRST language: YES NO (circle one)

How much time a day do you spend online communicating with others? (circle one)

a) Less than 1 hour
b) More than 1 hour but less than 2
c) More than 2 hours but less than 3
d) More than 3 hours but less than 4
e) More than 4 hours but less than 5
f) More than 5 hours but less than 6
g) More than 6 hours
Appendix F

Post-Interaction Questionnaire [Sender]

The following questionnaires are designed to assess the degree to which you “told the truth, the whole truth, and nothing but the truth” for each particular topic, as well as to gather some information on how you went about deceiving your partner or tried to convince your partner that you were telling the truth.

There are 5 questionnaires that follow. The first four are specific to each conversation topic, while the last questionnaire contains general questions. You will only be answering questions about the four conversation topics, after the introduction. The Experimenter will allow you to review each conversation topic, in order to help you answer these questions.

For each questionnaire, please read each item carefully and circle the number that best describes your response. Describe yourself, your behaviors and your feelings as honestly and as accurately as possible.

**Conversation Topic #1**

Please complete these questions after reviewing your discussion of the second topic. Please base your answers on your discussion of this topic only.

1. To what degree were you truthful about this topic (circle the most appropriate response)?

   Not at all truthful 0 1 2 3 4 5 6 7 8 9 10 Completely truthful

2. If you had to make a decision, were you LYING or TRUTHFUL (circle one)?

Please indicate how much you agree or disagree with the following statements by circling the most appropriate choice.
3. I am confident that my partner believed me in this part of the discussion.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

4. This part of the discussion went smoothly.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

5. I felt comfortable during this part of the discussion.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

6. I truly got to know my partner during this part of the discussion.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

7. If you had to lie on this topic, answer the following questions:
   What was the lie?
   ______________________________________________________

   What could have you said that was truthful?
   ______________________________________________________

8. Please describe how you mixed the deceptive messages with truthful messages either by circling one of the options below or describing in your own words:
   a) randomly inserted lies
   b) separated lies between truths
   c) told all the lies together
d) other (please explain): ____________________________________________________________

Now please turn to the transcript and identify every message in this part of the discussion that is explicitly deceptive by circling time stamps.

Conversation Topic #2

Please complete these questions after reviewing your discussion of the second topic. Please base your answers on your discussion of this topic only.

1. To what degree were you truthful about this topic (circle the most appropriate response)?

Not at all truthful 0 1 2 3 4 5 6 7 8 9 10 Completely truthful

2. If you had to make a decision, were you LYING or TRUTHFUL (circle one)?

Please indicate how much you agree or disagree with the following statements by circling the most appropriate choice.

3. I am confident that my partner believed me in this part of the discussion.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

4. This part of the discussion went smoothly.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
5. I felt comfortable during this part of the discussion.

| Strongly Disagree | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Strongly Agree |

6. I truly got to know my partner during this part of the discussion.

| Strongly Disagree | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Strongly Agree |

7. If you had to lie on this topic, answer the following questions:
   What was the lie?
   ____________________________________________________________

   What could you have said that was truthful?
   ____________________________________________________________

8. Please describe how you mixed the deceptive messages with truthful messages either by circling one of the options below or describing in your own words:
   e) randomly inserted lies
   f) separated lies between truths
   g) told all the lies together
   h) other (please explain): ______________________________________

Now please turn to the transcript and identify every message in this part of the discussion that is explicitly deceptive by circling time stamps.

Conversation Topic #3

Please complete these questions after reviewing your discussion of the second topic. Please base your answers on your discussion of this topic only.

1. To what degree were you truthful about this topic (circle the most appropriate response)?
2. If you had to make a decision, were you LYING or TRUTHFUL (circle one)?

Please indicate how much you agree or disagree with the following statements by circling the most appropriate choice.

3. I am confident that my partner believed me in this part of the discussion.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

4. This part of the discussion went smoothly.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

5. I felt comfortable during this part of the discussion.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

6. I truly got to know my partner during this part of the discussion.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

7. If you had to lie on this topic, answer the following questions:
What was the lie?
____________________________________________________

What could have you said that was truthful?
____________________________________________________

8. Please describe how you mixed the deceptive messages with truthful messages either by circling one of the options below or describing in your own words:

   i) randomly inserted lies
   j) separated lies between truths
   k) told all the lies together
   l) other (please explain): __________________________________________

Now please turn to the transcript and identify every message in this part of the discussion that is explicitly deceptive by circling time stamps.

Conversation Topic #4

Please complete these questions after reviewing your discussion of the second topic. Please base your answers on your discussion of this topic only.

1. To what degree were you truthful about this topic (circle the most appropriate response)?

   Not at all truthful 0 1 2 3 4 5 6 7 8 9 10 Completely truthful

2. If you had to make a decision, were you LYING or TRUTHFUL (circle one)?
Please indicate how much you agree or disagree with the following statements by circling the most appropriate choice.

3. I am confident that my partner believed me in this part of the discussion.

   Strongly Disagree  1   2   3   4   5   6   7   Strongly Agree

4. This part of the discussion went smoothly.

   Strongly Disagree  1   2   3   4   5   6   7   Strongly Agree

5. I felt comfortable during this part of the discussion.

   Strongly Disagree  1   2   3   4   5   6   7   Strongly Agree

6. I truly got to know my partner during this part of the discussion.

   Strongly Disagree  1   2   3   4   5   6   7   Strongly Agree

7. If you had to lie on this topic, answer the following questions:
   What was the lie?
   ________________________________________________________________

   What could you have said that was truthful?
   ________________________________________________________________

8. Please describe how you mixed the deceptive messages with truthful messages either by circling one of the options below or describing in your own words:
m) randomly inserted lies
n) separated lies between truths
o) told all the lies together
p) other (please explain): ________________________________

Now please turn to the transcript and identify every message in this part of the discussion that is explicitly deceptive by circling time stamps.

General Questions

Please complete the following questions as honestly and as accurately as possible by circling the most appropriate choice.

1. I find it hard to imitate the behavior of other people.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

2. At parties and social gatherings, I do not attempt to do or say things that others will like.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

3. I can only argue for ideas which I already believe.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

4. I can make impromptu speeches even on topics about which I have almost no information.
5. I guess I put on a show to impress or entertain others.

6. I would probably make a good actor.

7. In a group of people I am rarely the center of attention.

8. In different situations and with different people, I often act like very different persons.

9. I am not particularly good at making other people like me.

10. I’m not always the person I appear to be.
11. I would not change my opinions (or the way I do things) in order to please someone or win their favor.

12. I have considered being an entertainer.

13. I have never been good at games like charades or improvisational acting.

14. I have trouble changing my behavior to suit different people and different situations.

15. At a party I let others keep the jokes and stories going.

16. I feel a bit awkward in public and do not show up quite as well as I should.
17. I can look anyone in the eye and tell a lie with a straight face (if for a right end).

18. I may deceive people by being friendly when I really dislike them.

Finally, please answer some demographic questions:

Are you MALE or FEMALE (circle one)

What is your AGE (in years)? ________

Is English your FIRST language: YES  NO (circle one)

How much time a day do you spend online communicating with others? (circle one)

   a) Less than 1 hour
   b) More than 1 hour but less than 2
   c) More than 2 hours but less than 3
   d) More than 3 hours but less than 4
   e) More than 4 hours but less than 5
   f) More than 5 hours but less than 6
   g) More than 6 hours
Appendix G

Detecting Deception in CMC Conversations
Debriefing

As Computer-Mediated Communication (CMC) becomes more ubiquitous, it becomes increasingly important for us to understand how a variety of communication processes may be affected. For example, what are the characteristics of truthful and deceptive messages in CMC? How can we detect deception in computer-mediated forms of communication based on message linguistic characteristics? Previous research has not yet examined these questions and the goal of our research is to begin to address the impact of CMC on deception production and detection.

Rationale

In this experiment, we compare linguistic differences between deceptive and truthful messages in CMC. Not all messages produced within the Topic Lie Assignment are lies, and, vice versa, some messages produced within the Topic Truth Assignment may be lies. We asked participants assigned to the Sender role to indicate message truthfulness on a turn-by-turn basis so that we can examine language differences between and within Topic Lie and Topic Truth assignments.

Methods

Participants were randomly assigned to either the Sender role, who were instructed to not “tell the truth, the whole truth, and nothing but the truth” on two of the assigned conversation tasks, or to the Receiver role, who was instructed to try to detect the Sender’s deceptions after the conversations were completed.

Results

So, what do you think? If you were a Sender, do you think the communication setting you were in affected your ability to lie successfully? If you were the Receiver, do you think the conversation setting affected your ability to detect your partner’s deceptions? What cues did you use to try to identify your partner’s deceptions? If you would like to know how the results turned out, just email us and we’ll send you a report when we have collected and analyzed all the data.

Please note that it is very important that you do not discuss the manipulations or results of this experiment with others in your classes that have not participated in the experiment.

If you have any other questions, please feel free to contact Professor Hancock at jth34@cornell.edu or by phone (255.4452).

Thank you for your participation.
Figure 1

Truthful and Deceptive Utterances Across Topics as Found by Hirschberg et al, 2005

Lie Topic

Honest Topic

Deceptive Utterances

Truthful Utterances
Figure 2
Utterance Classification Abbreviations:

*tLuD* - **Deceptive** utterances told to support an overall **lie** topic
*tLuT* - **Truthful** utterances told in support of an overall **lie** topic
*tHuD* - **Deceptive** utterances told in support of an overall **honest** topic
*tHuT* - **Truthful** utterances told in support of an overall **honest** topic

How to read the abbreviations:
*tLuT* – **topic** *lie* utterance **Truth**
Figure 3

Truthful and Deceptive Utterances within Topics in Current Study
Table 1: Word Count

<table>
<thead>
<tr>
<th>Across Topics</th>
<th>Lie Topic</th>
<th>Honest Topic</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Words</td>
<td>5342</td>
<td>4357</td>
<td>&lt; 0.05*</td>
</tr>
<tr>
<td>Number of Messages</td>
<td>463</td>
<td>342</td>
<td>&lt; 0.05*</td>
</tr>
<tr>
<td>Avg. Message Length</td>
<td>11.54</td>
<td>12.74</td>
<td>0.062†</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Within Lie Topic</th>
<th>Deceptive Messages</th>
<th>Truthful Messages</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Words</td>
<td>3084</td>
<td>2258</td>
<td>0.053†</td>
</tr>
<tr>
<td>Number of Messages</td>
<td>210</td>
<td>253</td>
<td>0.2652</td>
</tr>
<tr>
<td>Avg. Message Length</td>
<td>14.69</td>
<td>8.93</td>
<td>&lt; 0.001*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Message Level Across Topics</th>
<th>Deceptive Msgs in Lie Topics</th>
<th>Truthful Msgs in Lie Topics</th>
<th>Truthful Msgs in Honest Topics</th>
<th>Deceptive Msgs in Honest Topics</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Words</td>
<td>3084</td>
<td>2258</td>
<td>3882</td>
<td>475</td>
<td>( ab = 0.053† ) ac = 0.109</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( ad &lt; 0.001* ) ( bc &lt; 0.001* ) ( bd &lt; 0.001* ) ( cd &lt; 0.001* )</td>
</tr>
<tr>
<td>Number of Messages</td>
<td>210</td>
<td>253</td>
<td>309</td>
<td>33</td>
<td>( ab = 0.265 ) ac = 0.05*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( ad &lt; 0.001* ) ( bc &lt; 0.05* ) ( bd &lt; 0.001* ) ( cd &lt; 0.001* )</td>
</tr>
<tr>
<td>Avg. Message Length</td>
<td>14.69</td>
<td>8.93</td>
<td>12.57</td>
<td>14.39</td>
<td>( ab &lt; 0.001* ) ac = 0.001*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( ad &lt; 0.001* ) ( bc &lt; 0.05* ) ( bd = 0.069† ) cd &lt; 0.001*</td>
</tr>
</tbody>
</table>

* = Correlation is significant at the 0.05 level (2-tailed), † = Approaches Significance
<table>
<thead>
<tr>
<th>Table 2: First-Person Pronoun Use</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Across Topics</strong></td>
</tr>
<tr>
<td>First-Person Pronouns</td>
</tr>
<tr>
<td><strong>Within Lie Topic</strong></td>
</tr>
<tr>
<td>First-Person Pronouns</td>
</tr>
</tbody>
</table>
| **Message Level Across Topics** | Deceptive Messages in Lie Topics\(^a\) | Truthful Messages in Lie Topics\(^b\) | Truthful Messages in Honest Topics\(^c\) | Deceptive Messages in Honest Topics\(^d\) | LL
| First-Person Pronouns           | 11.19 (%) | 8.64         | 13.01 | 10.32 |

\[^a\] = p < 0.05, **[^b\] = p < 0.01, 2-tailed test
### Table 3: Negative Emotion Word Frequency

<table>
<thead>
<tr>
<th></th>
<th>Lie Topic</th>
<th>Honest Topic</th>
<th>LL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Across Topics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neg. Emotion Words</td>
<td>0.49 (%)</td>
<td>0.41</td>
<td>0.29</td>
</tr>
<tr>
<td><strong>Within Lie Topic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deceptive Messages</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neg. Emotion Words</td>
<td>0.49 (%)</td>
<td>0.49</td>
<td>0.00</td>
</tr>
<tr>
<td>Truthful Messages</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neg. Emotion Words</td>
<td>0.49 (%)</td>
<td>0.49</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Message Level Across Topics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deceptive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neg. Emotion Words</td>
<td>0.49 (%)</td>
<td>0.49</td>
<td>0.56</td>
</tr>
<tr>
<td>Truthful</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neg. Emotion Words</td>
<td>0.49 (%)</td>
<td>0.56</td>
<td>0.42</td>
</tr>
<tr>
<td>Deceptive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neg. Emotion Words</td>
<td>0.49 (%)</td>
<td>0.42</td>
<td>0.42</td>
</tr>
</tbody>
</table>

* = p < 0.05, ** = p < 0.01, 2-tailed test
Table 4: Exclusionary Word Frequency

<table>
<thead>
<tr>
<th></th>
<th>Lie Topic</th>
<th>Honest Topic</th>
<th>LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Across Topics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exclusionary Words</td>
<td>3.35 (%)</td>
<td>4.18</td>
<td>4.41*</td>
</tr>
<tr>
<td>Within Lie Topic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exclusionary Words</td>
<td>3.90 (%)</td>
<td>2.83</td>
<td>4.31*</td>
</tr>
<tr>
<td>Message Level Across Topics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deceptive</td>
<td>Truthful</td>
<td>Truthful</td>
<td>Deceptive</td>
</tr>
<tr>
<td>Deceptive_msgs in Lie_topics</td>
<td>3.90 (%)</td>
<td>2.83</td>
<td>5.86</td>
</tr>
<tr>
<td>Deceptive_msgs in Honest_topics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truthful_msgs in Lie_topics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truthful_msgs in Honest_topics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LL</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* = \(p < 0.05\), ** = \(p < 0.01\), † = Approaches Significance, 2-tailed test

\[ab = 4.31*\]
\[ac = 12.00**\]
\[ad = 3.14\]
\[bc = 26.90**\]
\[bd = 0.40\]
\[cd = 11.78**\]