Expanding a Catalogue of Deceptive Linguistic Features with NLP Technologies

Nicholas D. Duran,* Scott Crossley, Charles Hall,*** Philip McCarthy,*** and Danielle McNamara***

Department of Psychology*
University of Memphis
Memphis, TN, 38152
{nduran, dsmcnamr} @ memphis.edu

Department of English**
Mississippi State University
Mississippi State, MS, 39762
sc544@msstate.edu

Department of English***
University of Memphis
Memphis, TN, 38152
{cehall, pmmccrth} @ memphis.edu

Abstract

We evaluate conversational transcripts of deceptive speech using a sophisticated natural language processing tool called Coh-Metrix. Coh-Metrix is unique in that it tracks linguistic features based on social and cognitive factors. The results from Coh-Metrix are compared to linguistic features reported in previous independent deception research, which used a natural language processing tool called LIWC. The comparison provides converging validity for several linguistic features, and establishes new insights on deceptive language.

Introduction

In this study, we are concerned with establishing links between deception patterns in written discourse and linguistic indices. We build from the work of Hancock et al. (2008) and their analysis of the linguistic features characterizing the deceptive and truthful conversations of native English speakers.

To conduct their analysis, Hancock et al. (2008) used an automated natural language processing tool called Linguistic Inquiry Word Count (LIWC, Pennebaker, Francis, and Booth, 2001). In our study, we replicate and extend Hancock and colleagues with another computational tool, Coh-Metrix (Graesser et al. 2004), which has a rapidly growing reputation for text analysis (Crossley et al. 2007; Duran et al. 2007; McCarthy et al. 2006; McNamara et al. 2006).

Through our approach, using Coh-Metrix, we hope to build a more complete descriptive catalogue of text characteristics that are indicative of deception. Additionally, by comparing two linguistic tools at the forefront of computational linguistics research, we intend to provide (and question) the converging validity of computational algorithms that purportedly measure linguistic features of deceptive speech.

Linguistic patterns of deception in conversation

Our interest here is to objectively observe deception as it naturally unfolds in conversation. According to Zhou et al. (2004), deception often occurs as a dialogue between interlocutors, and as such, the linguistic cues that identify deceptive competence emerge from the joint contribution of both conversational partners (hereafter referred to as the SENDER and RECEIVER of deception). However, the very goals of deception - a behavior designed to defeat detection - renders the enterprise of selecting identifiable linguistic cues very difficult. Indeed, even receivers and novice judges are close to chance in detecting a sender's deception (Vrij et al. 2000). It is for these reasons we present a quantitative analysis that attempts to uncover the subtle, but salient cues that might be useful for deception identification. To do so, we are not completely without guidelines. There are clear hypotheses about the social and cognitive constraints that influence culturally-specific deceptive behavior of native English speakers in an American context. By using these guidelines, relevant sets of linguistic indices can be identified beforehand. Winnowed sets of indices ensure we do not take haphazard liberties with statistical error, and further, the importance of hypothesized social and cognitive influences can be empirically tested.

For this study, we follow the lead of Hancock et al. (2008) in using computational natural language processing tools to identify linguistic patterns of deception. Hancock and colleagues were largely motivated by the research of Burgoon and Buller (Buller and Burgoon 1996; Burgoon, Buller, and Floyd 2001), who argue that deception is concomitant to maintaining plausibility in social interaction. The linguistic behavior of both speaker and receiver is predicated upon the unique context of social and linguistic coordination in conversation. Hancock and colleagues were most interested in whether the receiver engaged in lexical matching with the deceptive sender or instead changed linguistic behavior because of an implicit suspicion of the senders sincerity. Of course, the conclusions reached by Hancock and colleagues are contingent on the usefulness of the specific natural language processing tool they used to detect deception within a conversation. For their study, they used *Linguistic Inquiry and Word Count* (LIWC) to analyze a large corpus they collected of deceptive and truthful conversations from native English speakers.

The availability of the deception corpus allows us to proceed with our current analysis without having to collect new data. Moreover, it also provides the opportunity to compare and contrast the Coh-Metrix software (with a reputation for text analysis) used in the current study with the LIWC software (with a reputation for deception studies) used by Hancock et al. (2008). Before doing so, we first review and critique their method for collecting the deception corpus. We then present the data analysis of Hancock and colleagues alongside our expanded approach.

Conversational deception

The dynamic maintenance of conversational deception has unique challenges. Although a receiver may be unaware of the veracity of the sender's false statements, the sender must continually stay committed to preserving the receiver's presumption of truth. In doing so, senders must process and comprehend the speech of the receiver while simultaneously planning their own response (Greene et al. 1985); they must actively monitor the receiver's understanding in order to establish and maintain conceptual common ground (Clark and Schaefer 1987); and senders must adjust pragmatic strategies on-the-fly when addressing different audiences (e.g., employer vs. grandparent).

Hancock et al. (2008) hypothesized that the sender's maintenance of both their own false reality and the receiver's ostensible reality comes at the price of cognitive resources, thereby creating compensatory linguistic behavior on the part of the sender. Hancock and colleagues also hypothesized that the senders behavior may elicit a subtler, but nevertheless unique pattern of linguistic behavior in the receiver. Indeed, many researchers claim the mutual influence between interlocutors creates an inter-dependent relationship in language use (Clark 1996; Pickering and Garrod 2004). It is theoretically important, therefore, to consider the linguistic profiles of both sender and receiver.

Deception in interactive contexts such as conversation also increases the risk of being discovered as a fraud, resulting in *face loss* that is often associated with negative social standing (Brown 1977). These social factors are embedded in the influences of the culture at large, and are inextricably linked to the cognitive demands outlined. Based on these characterizations of conversational deception, we operationalized the linguistic indices into six categorical constructs that will be explored in further detail: (a) Quantity, (b) Immediacy, (c) Specificity, (d) Accessibility, (e) Complexity, and (d) Redundancy.

We proceed by briefly explaining the theoretical motivation for each of the six categorical constructs. For each category, we report the results from the Coh-Metrix analysis and interpret the results within a framework of conversational deception. Where possible, we also compare and contrast our results with those of Hancock et al. (2008).

Method

Participants

We used the original cohort of participants from the Hancock et al. (2008), which included 30 male and 36 female upper-level undergraduate students from a private university in the northeastern United States. The 66 participants were randomly paired to create 33 same-sex interlocutor pairs that were unacquainted with each other prior to their participation in the study.

Procedures

The experiment was conducted within a text-based, computer-mediated conversational environment (CMC). Participants were randomly assigned the role of *receiver* or *sender* for each dyad. The sender's role was to initiate and maintain a conversation; the sender was also responsible for introducing deception to the conversation. Senders were informed that it would be necessary to deceive their partners on two of the four topics pre-selected by the researchers, and to tell the truth on the other two topics. These experimental topics included: *Discuss the most significant person in your life*; *Talk about a mistake you made recently*; *Describe the most unpleasant job you have ever had to do*; and *Talk about responsibility*. The receivers, blind to the senders deception, were merely instructed to stay engaged and responsive to the ongoing conversation. The presentation of topics, as well as the order of deception, were counterbalanced across all participant pairs. Prior to initiation, participants were allowed 5 minutes to reflect upon the topics, thus allowing the sender (i.e., the deceiver) to prepare the fabricated responses. Participants were also allowed as much time as possible to discuss each topic.

The recorded messages were converted into sender and receiver transcript files according to topic. A total of 264 transcripts were produced, with each dyad generating eight different transcript files.

Data preparation and index selection

In our current study, we used the same transcripts as Hancock et al. (2008), but employed the Coh-Metrix software. Coh-Metrix integrates lexicons, syntactic parsers, part-of-speech classifiers, semantic analysis, and other advanced tools in natural language processing. Algorithms include referential overlap, proportion of situational dimensions (e.g., causal dependencies), latent semantic similarity, density of connectives, and syntactic complexity.

We selected linguistic indices that were guided by the many principles established in Hancock et al. (2008) and elsewhere in the communication and deception literature (Burgoon et al. 1996; Zhou et al. 2004). In this study, we attempt to replicate the LIWC finding of Hancock et al. as well as extend the indices to characterize novel components of deceptive speech. In the sections that follow, we address the criteria for selecting and computationally instantiating the relevant linguistic indices.

Coh-Metrix analysis and LIWC comparison

Each category above (quantity, immediacy, specificity, accessibility, complexity, and redundancy) is represented by 2 to 3 Coh-Metrix indices that were chosen to provide converging construct validity. At least one of these indices was selected to be computationally similar to a LIWC index. We added this similarity constraint for the purpose of comparing to Hancock et al. (2008), as well as for establishing measurement reliability. However, several categories do not have a representative and/or a computationally similar LIWC index. These omissions are addressed accordingly.

Quantity. In both Hancock et al. (2008) and this study, the number of words and number of words per conversational turn were computed and compared between deceptive and truthful conversation transcripts. This measure is theoretically important for assessing the willingness of deceptive senders to proffer information. On the one hand, senders may use fewer words to minimize the opportunities to incriminate themselves (Colwell, Hiscock, and Memon 2002). As such, senders' overall word count and number of words per conversation turn should be significantly less in deception than when telling the truth. On the other hand, senders want to appear socially involved so as not to violate a social norm of reciprocity that might otherwise raise suspicion (Burgoon et al. 1996). Senders, therefore, may maintain their word count across truthful and deceptive interactions.

Immediacy. Wiener and Mehrabian (1968) have suggested that deceptive statements are marked by "distancing strategies" that minimize personal involvement with the content of the message. One such distancing strategy is the decreased use of first person personal pronouns. Related to this decrease, deceptive messages are expected to have a greater number of second and third person pronouns (e.g., you, s/he, it, they) to divert attention from themselves.

For the analysis of pronoun use, Hancock et al. (2008) computed the percentage of first, second, and third person pronouns in deceptive and truthful conversations. In a similar approach, we also used Coh-Metrix to assess pronoun use as a distancing strategy. However, Coh-Metrix computes a measure for all personal pronouns. A statistically significant Coh-Metrix result would suggest that the first, second, and third person pronouns all change in the same direction.

Specificity. According to Reality Monitoring theory (Johnson and Raye 1981), the temporal and spatial characteristics for each event will differ in terms of specificity. Events that originate in actual perception will have greater temporal and spatial detail than events that originate from internal simulations. For this analysis, we chose two Coh-Metrix indices that capture the linguistic features of temporal and spatial characteristics. The temporal features index is computed as a ratio score that divides the summed occurrence of all temporal words in a conversational transcript by the total number of words in the transcript. For the Coh-Metrix spatial index, the number of locational prepositions (e.g., here, on, in) are counted for each transcript and normalized for differences in transcript length by converting to an incidence score (out of 1000 words).

There are no equivalent measures for temporal and spatial specificity in Hancock et al. (2008). However, in terms of a general specificity, Hancock and colleagues hypothesized that there might be a decrease in general specificity, thus prompting the receiver of a lie to ask more questions for clarification or detail. As such, the number of questions asked by receivers will increase as the sender is lying. To infer an asked question, Hancock and colleagues used LIWC to compute the percentage of sentences ending with question marks. In similar fashion, we used Coh-Metrix to compute a proportion score of Wh- adverbs (e.g., why, what) to assess possible changes in receiver's linguistic behavior.

Accessibility. We hypothesized that deceivers would select from a vocabulary that is easier to retrieve from memory. Based on the seminal work of Paivio (1969) and Underwood and Schulz (1960), word retrieval accessibility is modulated by experiential influences of word meaningfulness, familiarity, and concreteness. We thus analyzed word familiarity, meaningfulness, and concreteness scores from Coh-Metrix, which provides an average score based on human ratings of over 150,000 words compiled in the MRC database (Coltheart 1981). Word meaningfulness is operationalized by the number of associations that a word invokes for native English speakers. More associations increase word meaningfulness and the ease of retrieval for that word. Word familiarity is the familiarity of the orthographic form of a word. The more familiar a word is the more likely it will be retrieved. Finally, word concreteness refers to how easy it is to explicitly ground a word in perceptual experiences. For example, a word like *house* is more easily grounded than an abstract word like *interesting*. As such, concrete words are more easily recalled than abstract words. There are no equivalent indices for accessibility in Hancock et al. (2008).

Complexity. Another linguistic predictor of conversational deception is change in the syntactic complexity of sentential structures. Based on our general hypothesis of cognitive and social demands, deceivers will minimize or compensate for the demand by avoiding sentences with difficult syntactic composition. In Coh-Metrix, a standard measure of sentence complexity is the number of words before the main verb of the main clause. Assuming that a deceiver's memory resources would be taxed by the process of lying, we can expect a decrease in words before the main verb (i.e., lower complexity) compared to the truth-telling condition.

A second Coh-Metrix index of complexity that is common to LIWC is the number of negation connectives (e.g., did not, is not, but, except) that appear in each conversational transcript. Newman et al. (2003) argued that deceptive speakers will tend to avoid using negation connectives because they risk presenting incriminating contradictions and muddled detail. Negation connectives require speakers to contrast events that actually occurred with events that did not occur. Negative connectives help clarify event depictions, although the speaker must also recall additional detail from memory. Of course, for deceptive speakers, that detail must be conjured at that moment. As such, deceptive speakers might have additional challenges because they are "recalling" false details from an already distorted reality; a reality that may be loosely constructed in spontaneous conversation. Thus, the deceiver may sacrifice clarity and use fewer negation connectives to avoid accidental contradictions.

Redundancy. An important component of event narration, in both deceptive and truthful conversations, is the coherence of statements and ideas. The question remains as to whether the coherence of a speaker's mental event representation influences the cohesion of their speech. Deceptive speech can potentially address this question because deceivers' mental representations of false events are likely to be less coherent than representations of truthful events. If this is the case, the less coherent deceptive representation may result in less cohesive speech.

We evaluated the cohesion of deceptive and truthful conversations with two widely used indices in text analysis that are incorporated in Coh-Metrix: argument overlap and Latent Semantic Analysis (LSA) given/new value (Dufty et al. in press). Both indices are broad indicators of between-sentence conceptual redundancy. This redundancy reinforces information by keeping it focal in a developing narrative. Argument overlap computes explicit overlap between two sentences by tracking the common nouns in either single or plural form. These measures are unique to Coh-Metrix and there was no equivalent in Hancock et al. (2008).

Results

Quantity

In the current Coh-Metrix analysis, a significant main effect of message type (deceptive vs. truthful) was observed for total word count, $F(1, 33) = 8.87$, $p = .005$. More words were produced during deceptive conversation (M = 159.38, SE1 = 9.97) than truthful conversation (M = 122.76, SE = 9.23). Senders increased word use from 123.15 words (SE = 10.21) in truthful conversations to 158.16 words (SE = 12.01) in deceptive conversations. Receivers increased word use from 122.37 words (SE = 10.39) in truthful conversations to 160.59 words (SE = 16.12) in deceptive conversations. These patterns of results were virtually identical to Hancock et al. (2008), who also found a statistically significant main effect for message type.

Immediacy

For pronoun use, no statistical differences were noted. As such, our analysis lends indirect support to Hancock et als original finding.

Specificity

For our specificity analysis, Coh-Metrix indices of temporal and spatial specificity were not statistically significant. However, for the Coh-Metrix index of general specificity in terms of questions asked, there was a significant interaction between message type and speaker type for number of wh-adverbs used, $F(1, 33) = 6.83$, $p = .01$. An analysis of wh-adverb use at each level of speaker type for deceptive and truthful messages revealed that senders used fewer wh-adverbs, and presumably asked fewer questions, when being deceptive (M = 6.53, SE = .98) than when telling the truth (M = 9.04, SE = 1.09), $F(1, 33) = 4.19$, $p = .05$; conversely, receivers used marginally more wh-adverbs when being deceived (M = 10.34, SE = 1.23) than when told the truth (M = 7.33, SE = 1.02), $F(1, 33) = 3.30$, $p = .08$. These patterns of results suggest that receivers ask more questions when being deceived while senders ask fewer questions when being deceptive.

Accessibility

There was a statistically significant main effect of message type for word meaningfulness in conversations, $F(1, 33) = 7.88$, $p = .008$. The words used in deceptive conversations were more meaningful (M = 418.47, SE = 1.23) than words used in truthful conditions (M = 412.76, SE = 1.75). Senders' use of meaningful words increased from a rating of 415.21 (SE = 2.30) in truthful conversations to a rating of 418.15 (SE = 1.47) in deceptive conversations. Receivers increased from a rating of 410.31 (SE = 2.60) in truthful conversations to a rating of 418.78 (SE = 2.00) when they were being deceived. No interaction was observed between message type and speaker type.

For the analysis of word concreteness there was a significant interaction between message type and speaker type, $F(1, 33) = 5.42$, $p = .02$. An analysis of word concreteness at each level of speaker type for deceptive and truthful messages suggest that senders use more concrete words when deceptive (M = 340.63, SE = 3.31) than when they are telling the truth (M = 332.99, SE = 2.69), $F(1, 33) = 3.25$, $p = .05$. There was no difference for receivers in deceptive conversations (M = 337.49, SE = 2.34) or truthful conversations (M = 337.77, SE = 3.28).

The third accessibility measure of word familiarity was not statistically significant.

Complexity

A significant main effect of message type was observed for this syntactical complexity measure, $F(1, 33) = 5.63$, $p = .02$. More words were used before the main verb in deceptive conversations (M = 7.14, SE = .46) than in truthful conversations (M = 5.79, SE = .37). Senders use more words before the main verb when deceptive (M = 6.79, SE = .71) than when telling the truth (M = 6.16, SE = .61. Receivers use more words before the main verb (M = 7.50, SE = .60) when they are being deceived than in truthful conversations (M = 5.41, SE = .43). No interaction was observed between message type and speaker type.

For the occurrence of negation connectives no statistically significant effects were found. This agrees with the findings from Hancock et al., (2008).

Redundancy

For the first redundancy analysis of argument overlap, we did not find any statistically significant effects. However, the more subtle measure, in the LSA given/new value, revealed a statistically significant main effect for message type, $F(1, 33) = 9.32$, $p = .004$. In the deceptive conversations, there was a higher given/new value (M = .25, SE = .005) compared to truthful conversations (M = .23, SE = .007). Senders' given-new value was higher when they were deceptive (M = .26, SE = .007) compared to when they were telling the truth (M = .24, SE = .01). Receivers' given-new values were higher when they were being deceived (M = .25, SE = .008) compared to when they were being told the truth (M = .22, SE = .01).

Discussion

The overall results of our study demonstrate that the linguistic features that characterize deceptive conversations are substantially different from those that characterize truthful conversations. From the perspective of Coh-Metrix, we can describe deceptive conversations as involving: (a) more words overall, but fewer words used per conversational turn, (b) words that are more meaningful, (c) utterances of each conversational turn being more syntactically complex, and (d) less unique information introduced during the course of the conversation.

The above effects changed in the same direction for both sender and receiver. However, other changes in linguistic behavior were specific to just the sender or receiver. For example, word concreteness increased only for senders while they were being deceptive. For receivers, word concreteness did not increase or decrease between deceptive or truthful conversations. Additionally, for the number of questions asked in deceptive conversations, receivers asked marginally more questions while the senders asked fewer questions.

In terms of the comparison between Coh-Metrix and LIWC, five indices were comparable, and of these five, total word count and negation had the same result. This convergence confirms that more words are used in deceptive conversations and that there are no differences for the use of negation. The multi-method alignment lends greater credibility to the Coh-Metrix and LIWC indices, as well as to the quantity and immediacy constructs in general.

The two remaining indices, words per conversational turn and questions, did not completely converge. For the words per conversational turn index, the Coh-Metrix analysis revealed that both sender and receiver used fewer words in each utterance during deceptive conversations. With LIWC, only receivers used fewer words in each utterance during deception. For the questions index, the Coh-Metrix analysis revealed that receivers asked more questions while being deceived and senders asked fewer questions while being deceptive. With LIWC, only the receiver asked fewer questions during deception. In general, for both of the nonconverging indices, the Coh-Metrix analysis found a statistically significant effect that was not found in the LIWC analysis.

Taken as whole, both this study and Hancock et al. (2008), demonstrate that at least one type of deception is detectable and analyzable through sophisticated natural language processing tools. Through our approach, we were able to evaluate the effectiveness of each NLP tool in a common context of social interaction. In addition, we were also able to use Coh-Metrix to build a more complete catalogue of the linguistic features that emerge during deception.

Acknowledgments

This research was supported in part by the Institute for Education Sciences (IES R305G020018-02; IES R305a080589), Counter-intelligence Field Activity (CIFA H9c104-07-C-0014) and the National Science Foundation (NSF) Graduate Research Fellowship awarded to the first author. The views expressed in this paper do not necessarily reflect the views of the IES, CIFA or NSF. The authors would like to thank Jeff Hancock for providing us with the conversational transcripts used in this study. The authors also acknowledge the contributions to this project made by Arthur Graesser and Zhiqiang Cai.

References

Brown, B. R. 1977. Face saving and face restoration in negotiation. In Druckman, D., ed., *Negotiations: Social-Psychology Perspectives*. Beverly Hills, CA: Sage Publications. 275–299.

Buller, D. B., and Burgoon, J. K. 1996. Interpersonal deception theory. *Communication Theory* 203–242.

Burgoon, J. K.; Buller, D. B.; Floyd, K.; and Grandpre, J. 1996. Deceptive realities: Sender, receiver, and observer perspectives in deceptive conversations. *Communication Research* 23:724–748.

Burgoon, J. K.; Buller, D. B.; and Floyd, K. 2001. Does participation affect deception success?: A test of the interactivity principle. *Human Communication Research* 27:503–534.

Clark, H. H., and Schaefer, E. F. 1987. Collaborating on contributions to conversations. *Language and Cognitive Processes* 2:19–41.

Clark, H. H. 1996. *Using Language*, volume 23. Cambridge University Press.

Coltheart, M. 1981. The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology* 33A:497–505.

Colwell, K.; Hiscock, C. K.; and Memon, A. 2002. Interviewing techniques and the assessment of statement credibility. *Applied Cognitive Psychology* 16:287–300.

Crossley, S. A.; Louwerse, M. M.; McCarthy, P. M.; and McNamara, D. S. 2007. A linguistic analysis of simplified and authentic texts. *The Modern Language Journal* 91:15–30.

Dufty, D. F.; Hempelmann, C. F.; McCarthy, P. M.; Cai, Z.; McNamara, D. S.; and Graesser, A. C. in press. Givenness and newness of information: Automated identification in written discourse. *Discourse Processes*.

Duran, N. D.; Bellissens, C.; Taylor, R. S.; and McNamara, D. S. 2007. Quantifying text difficulty with automated indices of cohesion and semantics. In *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, 233–238. Austin, TX: Cognitive Science Society.

Graesser, A. C.; McNamara, D. S.; Louwerse, M. M.; and Cai, Z. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods Instruments and Computers* 36:193–202.

Greene, J. O.; O'Hair, H.; Cody, M. J.; and Yen, C. 1985. planning and control of behavior during deception. *Human Communication Research* 11:335–364.

Hancock, J. T.; Curry, L.; Goorha, S.; and Woodworth, M. T. 2008. On lying and being lied to: A linguistic analysis of deception in Computer-Mediated communication. *Discourse Processes* 45:1–23.

Johnson, M. K., and Raye, C. L. 1981. Reality monitoring. *Psychological Review* 88:67–85.

McCarthy, P. M.; Lewis, G. A.; Dufty, D. F.; and McNamara, D. S. 2006. Analyzing writing styles with Coh-Metrix. In *Proceedings of the 19th Annual Conference of the Florida Artificial Intelligence Society*, 764–769. Menlo Park, California: AAAI Press.

McNamara, D.; Ozuru, Y.; Greasser, A.; and Louwerse, M. 2006. Validating Coh-Metrix. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society, Mahwah, NJ: Erlbaum*, 573–578.

Newman, M. L.; Pennebaker, J. W.; Berry, D. S.; and Richards, J. M. 2003. Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin* 29:665.

Paivio, A. 1969. Mental imagery in associative learning and memory. *Psychological Review* 76:241–263.

Pennebaker, J. W.; Francis, M. E.; and Booth, R. J. 2001. *Linguistic Inquiry and Word Count: LIWC 2001*. Mahwah, NJ: Lawrence Erlbaum.

Pickering, M. J., and Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27:169–190.

Underwood, B. J., and Schulz, R. W. 1960. *Meaningfulness and Verbal Learning*. Philadelphia: Lippincott.

Vrij, A.; Edward, K.; Roberts, K. P.; and Bull, R. 2000. Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior* 24:239–263.

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

Zhou, L.; Burgoon, J. K.; Nunamaker, J. F.; and Twitchell, D. 2004. Automating Linguistics-Based cues for detecting deception in Text-Based asynchronous Computer-Mediated communications. *Group Decision and Negotiation* 13:81–106.