

# Individual Differences in Strategic Deception

Scott Appling and Erica Briscoe

Georgia Tech Research Institute

Georgia Institute of Technology

Atlanta, GA, United States

scott.appling@gtri.gatech.edu

erica.briscoe@gtri.gatech.edu

## Abstract

While text-based deception in computer mediated communication has been studied, e.g., Zhou (2005) and Duran et al. (2010), there has been less focus on the differentiation of strategies for deception, especially those which may manifest in modern communication, such as found in social media. In this paper, we extend our previous work on the evaluation of linguistic indicators to strategic deception (Appling et al., 2015), by evaluating the relationship between personality and deceptive strategy use and the utilization of linguistic features for inferring both (personality and deception). We find that even with a relatively small corpus, there is evidence that personality is related to particular deception strategies, though in short social media communications, these personality traits are difficult to infer using standard linguistic measures (e.g., LIWC). We also describe the corpus we collected from an experiment in which subjects engaged in deception through a social media platform.

## 1 Introduction

The ubiquitous nature of social media has significantly magnified the communicative ability of individuals and groups and changed the way many individuals receive information (Lenhart et al., 2010). Consequently, this increased access has enlarged the influence opportunity space, where deception is a commonly utilized influence method (Buller et al., 1994). A theme of our research is focused on the evaluation of how findings from previously studied communication methods, e.g., Burgoon et al.

(1996), differ as a result of both changing communication styles and the affordances provided by modern communication platforms (i.e., social media).

Research into “tells” that indicate when a person is lying have been a subject of interest for many years, e.g. (Ekman and Friesen, 1969), and has resulted in a substantial list of potential cues (DePaulo et al., 2003). These cues to deception are usually described as those that occur significantly more or less frequently when a communicator expresses an untruth as compared to the truth. These indicators may differ according to the point of view of the receiver and sender (Anderson et al., 1999) and have been identified in widespread communications, such as the 2016 United States election coverage (Scott, 2016). While our previous work evaluated the detection of deceptive statements and deception strategies using linguistic cues (Briscoe et al., 2014; Appling et al., 2015), here we describe our evaluation of how individual differences, specifically as represented through different personalities, may affect strategic deception and be inferred from an experimentally collected corpus.

## 2 Eliciting Deception Strategies

Past work has characterized multiple deception strategies (Rubin, 2010; Turner et al., 1975; Burgoon et al., 1996), roughly organized into the four main types described below.

- **Falsification** - e.g. lies, contradictions, or “distortions”
- **Exaggeration** - e.g. more or modified information via superlatives

- **Omission** - e.g. secrets (missing information), half-truths (less or modified information), or ‘concealment’
- **Misleading** - e.g. topic changes, irrelevant information, equivocation, or ‘diversionary responses’

To evaluate the use of the previously described set of strategies, we conducted human subject experimentation to simulate an online social network environment where participants were tasked with engaging in a conversation involving a decision making task where they were encouraged to employ deception. We also administered a personality inventory to assess each participant’s standing on the dimensions contained within the Five Factor Model (FFM) of personality (McCrae and Costa Jr, 1999): openness, conscientiousness, extroversion, agreeableness, neuroticism. Then we extracted linguistic features that were found to be statistically significant predictors of related personality traits. We then report on an analysis on the use of linguistic features for predicting deception strategies based on individual-trait level psychological constructs (personality).

The rest of the paper is organized as follows. First we describe the experimental setup and the collected data. Next, we discuss the deception strategies corpus that was created. We then discuss the results of two analyses: one related to predicting deception strategy on personality trait scores and a second on using psycholinguistic features to predict personality. We conclude with a discussion of the implications of the results and future directions

### 3 Deception Experiment

To study deception strategies in social media, we designed and ran an in-lab experiment where individuals were asked to use a social media platform for a group decision making task. The task allowed us to control the scenario as well as topics of discussion within the conversations.

Seventy-five subjects recruited from the student body at Georgia Tech participated in the experiment. Each subject was seated at a computer station and asked to use a mock online social media site (FaceFriend, see Figure 1). The mock site was intended



**Figure 1:** Example screenshot of the *FaceFriend* web-based interface used in the experiment. The interface was designed to mimic commonly used social media platforms.

to resemble a popular social networking site. Participants completed three pre-experiment surveys, including a personality test and a survey regarding their social media and email usage, so as to ensure consistency among familiarity with social media. The procedure and results reported below are from one of a three part experiment, the other two parts of which are not reported here (as they centered on the perception of deception and credibility, see Briscoe et al. (2014) and Briscoe et al. (2013) for details of the other parts of the study. Subjects were paid for their participation.

Participants were asked to read a scenario involving a group decision making task called the “sub-arctic survival problem” (Eady and Lafferty, 1969). Participants were instructed to log in to a mock social media platform, *FaceFriend*, and to communicate with another individual to decide the best ranking of items to take from a crashed plane site in an Arctic environment. Participants were further instructed to advocate for a specific ranked list of items, also provided, and to do so using deception, when possible. To ameliorate differences in creative ability, subjects were provided with expert explanations for the advantages provided by each item; these could be used as the basis for creating deceptive statements. For example, the list of items

included an Axe, which the expert noted could be used to chop wood. Lastly, participants were ostensibly informed that the other conversation participant would be unaware of any potential deception. The participant’s conversational partner was a confederate member of the research team, sitting in another room. During the conversation, to minimize variation in the confederate’s response language, the confederates communicated using a list of statements conceived beforehand as suitable decision making conversation responses (e.g., “What about the canvas?”). After discussing the items for 12 minutes, the interaction was concluded, and subjects were asked to log out of FaceFriend.

After the end of the conversation, participants were shown the list of statements they made during the task in a browser windows and asked to identify which ones were deceptive using a check box. Upon indicating a particular statement was deceptive, participants were prompted to categorize their strategy for that particular message. The possible strategies to select were the same types described in Section 2. Multiple strategies were allowed to be designated. An additional category, “Other”, was available that allowed free response space for any strategies that the subject felt did not fit into one of the four provided categories. Table 3 displays example deceptive statements, as identified by subjects, in each category.

## 4 Deception Corpus

The deception corpus contains all statements made by participants during the deception experiment described earlier. These include truthful and deceptive statements. For each statement we provide a boolean indicator as to whether or not the participant annotated that statement as enacting one or more of the deception strategies described in Section 2.<sup>1</sup>

Table 4 lists the frequencies of each identified strategy in the corpus as indicated by the subjects. Statements which subjects identified as containing multiple strategies are not included. On average, the misleading strategy was the most likely strategy to be employed by participants and the omission strategy the least likely. Table 1 lists summary statistics about the statements in the corpus. Sentiment and

<sup>1</sup>The collected corpus is available by emailing the authors.

subjectivity scores were calculated for each statement, using the method described by De Smedt and Daelemans (2012). Table 2 provides the descriptive statistics of the distribution of personality variables across our subject pool, as well as relevant demographic information.

## 5 Analysis

### 5.1 Linguistic Cues and Deception Strategies

After excluding statements falling below a minimum one word threshold<sup>2</sup>, 393 statements (including truthful statements) were used in a one-way multivariate analysis of variance to determine the effect of deception strategy on linguistic cue levels. The 8 linguistics cues studied are listed in Table 1. A statistically significant difference was found between the different deception strategies and truthful statements on the combined linguistic cues dependent variables  $F(32, 1406.654) = 1.93, p < 0.001$ ; Wilks’  $\Lambda = 0.852$ ; partial  $\eta^2 = 0.039$ , though no specific differences were found between specific strategies. Follow-up univariate ANOVAs showed a statistically significant difference between the use of deception strategies and no strategy using a Bonferroni adjusted  $\alpha$  level of 0.025. Tukey post-hoc tests showed: there were less average verb cues for truthful statements compared to exaggerations ( $p < 0.001$ ) and compared to misleading statements ( $p < 0.019$ ) ( $M = 4.695$ ;  $SD = 0.274$ ;  $M = 6.657$ ;  $SD = 0.419$ ;  $M = 6.222$ ,  $SD = 0.413$ , respectively); lower mean word count for truthful statements compared to exaggeration statements ( $p < 0.019$ ;  $M=20.024$ ;  $SD=0.899$ ;  $M=25.086$ ;  $SD=1.376$ , respectively); and lower mean Fleish-Kincaid score cues for truthful statements compared to exaggeration statements ( $p < 0.005$ ;  $M = 3.228$ ;  $SD = 0.438$ ;  $M = 5.993$ ;  $SD = 0.670$ , respectively).

<sup>2</sup>Exclusion of one-word statements was based on the fact that these statements were almost exclusively one-word responses (i.e., “yes” or “no”) to questions posed by the confederate. While these responses may exhibit untruths, the authors feel that the reflection of deception is mostly based in the content of the interrogation, rather than the response itself.

**Table 1:** Deception Conversation Corpus Summary Statistics per Subject

| Variable                  | Mean  | Std. Dev. | Min.  | Max.  |
|---------------------------|-------|-----------|-------|-------|
| <b>Per Subject</b>        |       |           |       |       |
| Statement Count           | 6.00  | 3.62      | 1.00  | 17.0  |
| Strategy: Exaggeration    | 1.20  | 1.07      | 0     | 4.00  |
| Strategy: Misleading      | 1.31  | 1.21      | 0     | 6.00  |
| Strategy: Omission        | 0.80  | 1.01      | 0     | 5.00  |
| Strategy: Falsification   | 1.03  | 1.37      | 0     | 8.00  |
| <b>Per Comment</b>        |       |           |       |       |
| Avg. Modifiers            | 2.85  | 1.09      | 0.75  | 6.00  |
| Avg. Pausality            | 2.38  | 1.02      | 0     | 4.66  |
| Avg. Verbs                | 5.68  | 2.19      | 2.00  | 12.33 |
| Avg. Words                | 23.43 | 7.33      | 8.00  | 41.00 |
| Avg. Word Length          | 3.82  | 0.27      | 3.10  | 4.79  |
| Avg. Fleish-Kincaid Score | 4.91  | 3.40      | -3.62 | 12.20 |
| Avg. Sentiment            | 0.17  | 0.12      | -0.17 | 0.41  |
| Avg. Subjectivity         | 0.40  | 0.12      | 0.17  | 0.75  |

**Note:** Sentiment scores range between -1 (extremely negative) and 1 (extremely positive); Subjectivity scores range between 0 (completely objective) and 1 (completely subjective).

**Table 2:** Participant Demographic and Personality Statistics

| Variable          | Mean          | Stdev. | Min.        | Max. |
|-------------------|---------------|--------|-------------|------|
| Age               | 22.5          | 4.8    | 18          | 45   |
| Openness          | 25.1          | 4.8    | 15          | 35   |
| Conscientiousness | 35.9          | 6.4    | 18          | 49   |
| Extroversion      | 29.8          | 6.2    | 17          | 42   |
| Agreeableness     | 28.7          | 3.0    | 21          | 35   |
| Neuroticism       | 19.6          | 6.3    | 8           | 34   |
| Gender            | <b>Female</b> | 35     | <b>Male</b> | 40   |

## 5.2 Predicting Deception Strategies using Personality

Hierarchical multiple regression analyses were conducted using each deception strategy as the dependent variable with each FFM personality trait as a predictor including gender as a blocking variable. See Table 5 for details on each regression model. From the analysis, we find that only *extroversion* was a predictive trait, found to influence the *omission* deception strategy during deceptive communications,  $R^2 = .139$ ,  $F(2, 72) = 5.823$ ,  $p < .01$ . The direction of the relationship suggests that less extroverted individuals are more likely to use an

omission strategy.

## 5.3 Predicting Personality using Psycholinguistic Lexical Features

Given that there was a relationship found between *extroversion* and the *omission* strategy, we are then motivated leverage other work centered on automatically deriving personality from text, in order to derive indicators for extra version. To create our indicators, we utilize previously identified psycholinguistic features from (Mairesse et al., 2007) related to predicting extroversion from text. A hierarchical multiple regression analysis was conducted using the *omission* deception strategy as the dependent variable with the following LIWC features (found to be indicative of extroversion by Mairesse (2007): 1st Person Plural, 1st Person Singular, Articles, Body, Family, Filler, Inclusion, Negations, Positive Emotion, Social Processes, Tentativeness, Total Pronouns.

The ANOVA omnibus F-test proved insignificant,  $F(13, 74) = 0.840$ ,  $p = 0.617$ , indicating no statistically significant relationships existed between the LIWC features for predicting omission; no follow-up analyses were conducted.

**Table 3:** Examples of Deception Strategies

| Identified Deception Strategy | Participant’s Statement During Experiment  | Participant’s Explanation  |
|-------------------------------|--|--|
| Exaggeration                  | “the axe is awesome, and should be the first priority. In that cold weather we’ll need definitely need something to chop firewood with”  | “in this statement, I’m playing along with a positively framed question. I agree with the other guy and exaggerate how awesome the axe is, and how it should definitely be the first priority”   |
| Misleading                    | “well I think the canvas should be number 4, it could be used for shelter protect us from the harsh winds, but should be below the Crisco because that could be used to store water and be a signal ”    | “This was misleading because the canvas would probably be more important than the crisco because the canvas could hlave been used as a blanket that produced warmth and the crisco would not be as useful because you could just eat snow” |
| Falsification                 | “I think the axe, newspaper, crisco, cavas, wool, and pistol should be brought in that order. I itemized them from most importance to least. we can disgard lower priority items if the trip gets rough” | “I was trying to get list across without really taking time to actually reason through why I am choosing that order.”  |
| Omission                      | “I agree that the wool should be ranked 5th, but the crisco is much more useful than the canvas because it can be used for fire which is essential.”   | “I didn’t compare the cost/benefit of the two items, only what I liked about the crisco”   |

**Table 4:** Frequency of Deception Strategies

|              | Exaggeration | Misleading | Omission | Falsification |
|--------------|--------------|------------|----------|---------------|
| <b>Count</b> | 67           | 69         | 41       | 45            |

## 6 Discussion

The goal of the work was to evaluate how personality (as a means of characterizing individuals) can be used to predict the types of deception strategies that those individuals might use in short, social media-based conversations. The secondary goal was to determine if, given a limited social media corpus, those personality characteristics found related to strategic deception could be derived based on linguistic features. While we did find a relationship between personality and deception strategies, where less extro-

verted individuals are more likely to use an *omission* strategy, we did not find evidence that the *extroversion* trait can be accurately derived from this same text using only the extroversion indicators identified by Mairesse et al. (2007). The difficulty in inferring traits from content-derived psycholinguistic features has been identified in other investigations (Adali and Golbeck, 2012; Schwartz et al., 2013); however, the use of additional features, both linguistic, e.g. (Appling et al., 2013)) and non-, e.g. (Adali and Golbeck, 2012), can improve trait prediction.

Additionally, we determined that several linguistic features (e.g., the number of verbs in a statement) are significantly different across truthful and deceptive statements, but only for certain types of deception strategies. In previous work we evaluated the use of discriminative models for classifying decep-

**Table 5:** Hierarchical Multiple Regression Predicting Strategy from Gender and Personality

| Variable          | Omission Strategy     |         |         |         | Misleading Strategy    |         |         |         |
|-------------------|-----------------------|---------|---------|---------|------------------------|---------|---------|---------|
|                   | Model 1               |         | Model 2 |         | Model 1                |         | Model 2 |         |
|                   | B                     | $\beta$ | B       | $\beta$ | B                      | $\beta$ | B       | $\beta$ |
| Constant          | 0.820**               |         | 1.982** |         | 1.295**                |         | -0.642  |         |
| Gender            | 0.295*                | 0.292   | 0.352*  | 0.349   | -0.180                 | -0.150  | -0.232  | -0.193  |
| Openness          |                       |         | 0.160   | 0.034   |                        |         | 0.006   | 0.024   |
| Conscientiousness |                       |         | 0.060   | 0.009   |                        |         | 0.020   | 0.108   |
| Extroversion      |                       |         | -0.039* | -0.239  |                        |         | 0.024   | 0.123   |
| Agreeableness     |                       |         | -0.046  | -0.137  |                        |         | -0.002  | -0.005  |
| Neuroticism       |                       |         | -0.024  | -0.151  |                        |         | -0.021  | 0.108   |
| $R^2$             | 0.085                 |         | 0.139   |         | 0.023                  |         | 0.055   |         |
| F                 | 6.807*                |         | 5.823*  |         | 1.680                  |         | 0.666   |         |
| Variable          | Exaggeration Strategy |         |         |         | Falsification Strategy |         |         |         |
|                   | Model 1               |         | Model 2 |         | Model 1                |         | Model 2 |         |
|                   | B                     | $\beta$ | B       | $\beta$ | B                      | $\beta$ | B       | $\beta$ |
| Constant          | 1.189**               |         | -1.926  |         | 1.021**                |         | 2.708   |         |
| Gender            | -0.161                | -0.152  | -0.167  | -0.158  | -0.079                 | -0.058  | -0.067  | -0.049  |
| Openness          |                       |         | -0.001  | -0.004  |                        |         | 0.031   | 0.110   |
| Conscientiousness |                       |         | -0.002  | -0.011  |                        |         | 0.032   | 0.154   |
| Extroversion      |                       |         | -0.004  | -0.021  |                        |         | -0.010  | -0.043  |
| Agreeableness     |                       |         | 0.106   | 0.297   |                        |         | -0.107  | -0.236  |
| Neuroticism       |                       |         | 0.014   | 0.084   |                        |         | -0.014  | -0.063  |
| $R^2$             | 0.010                 |         | 0.100   |         | 0.003                  |         | 0.066   |         |
| F                 | 1.715                 |         | 1.263   |         | 0.245                  |         | 0.802   |         |

Note: \*  $p < .05$ , \*\*  $p < .001$

tion strategies (Appling et al., 2015) based solely on linguistic features. There, we found that some features were more discriminative for some strategies. For example, the most predictive feature for discriminating falsification was in the LIWC negation category (e.g., using ‘no’, ‘not’, ‘never’), which we found unsurprising given that these statements are often disagreements in response to the statements made by the other participant. Those features, however, were not found to be significantly associated with any particular strategy in the current study. Also, in that study, the *omission* strategy was the most difficult to discriminate, using a train random

forest classifier (Breiman, 2001). This result was not surprising given that, by definition, the omission of information does not provide linguistic content from which to derive features. Interestingly, and perhaps unfortunately, it is this same strategy that we find most related to personality dimensions, i.e., extroversion, in the current study.

An obvious question from this and related studies focuses on the appropriate amount of text that is needed to infer individual-level traits, such as these inferences are as accurate as direct measurement. In Golbeck et al. (2011) the authors conducted analysis with 176 participants who averaged 42.6 words

from combined social media status updates, whereas Mairesse et al. (2007) captured on average 766.4 words per subject. Other works, e.g. Nie et al. (2014), give no indication concerning the amount of text being used per individual. These questions are especially relevant given that social media communications are likely highly variable, exhibiting fluctuations arising from a variety of factors, such as mood (Bradley and Mogg, 1994).

## 7 Future Work and Conclusion

Given the likely influence of modern communication platforms on human communication styles, utilizing linguistic-based indicators for deception is a difficult and evolving challenge. One, likely fruitful, method of addressing this challenge is to leverage work that has been done on characterizing individuals as a means of using that characterization for predictive tasks; however, given the brevity and informality of social media communications, much more work must be done to create methods for utilizing this kind of text. Related to the use of lexical psycholinguistic features, we see opportunity with recent work in vectorized word representations to develop semantic features based on lexical qualities that weight observations smoothly. Future work will continue to evaluate the efficacy of using informal text-based characterizations as means of predicting communication strategies.

## Acknowledgments

This material is based upon work supported in part by the Defense Advanced Research Projects Agency (DARPA) under Contract No. W911NF-12-1-0043.

## References

- Sibel Adali and Jennifer Golbeck. 2012. Predicting personality with social behavior. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, pages 302–309. IEEE Computer Society.
- D Eric Anderson, Bella M DePaulo, Matthew E Ansfeld, Jennifer J Tickle, and Emily Green. 1999. Beliefs about cues to deception: Mindless stereotypes or untapped wisdom? *Journal of Nonverbal Behavior*, 23(1):67–89.
- D Scott Appling, Erica J Briscoe, Heather Hayes, and Rudolph L Mappus. 2013. Towards automated personality identification using speech acts. In *Seventh International AAAI Conference on Weblogs and Social Media*.
- Darren Scott Appling, Erica J Briscoe, and Clayton J Hutto. 2015. Discriminative models for predicting deception strategies. In *Proceedings of the 24th International Conference on World Wide Web Companion*, pages 947–952. International World Wide Web Conferences Steering Committee.
- Brendan P Bradley and Karin Mogg. 1994. Mood and personality in recall of positive and negative information. *Behaviour research and therapy*, 32(1):137–141.
- Leo Breiman. 2001. Random forests. *Machine learning*, 45(1):5–32.
- Erica J Briscoe, D Scott Appling, Rudolph L Mappus IV, and Heather Hayes. 2013. Determining credibility from social network structure. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 1418–1424. ACM.
- Erica J Briscoe, D Scott Appling, and Heather Hayes. 2014. Cues to deception in social media communications. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, pages 1435–1443. IEEE.
- David B Buller, Judee K Burgoon, JA Daly, and JM Wiemann. 1994. Deception: Strategic and nonstrategic communication. *Strategic interpersonal communication*, pages 191–223.
- Judee K Burgoon, David B Buller, Laura K Guerrero, Walid A Affi, and Clyde M Feldman. 1996. Interpersonal deception: Xii. information management dimensions underlying deceptive and truthful messages. *Communications Monographs*, 63(1):50–69.
- Tom De Smedt and Walter Daelemans. 2012. Pattern for python. *The Journal of Machine Learning Research*, 13(1):2063–2067.
- Bella M DePaulo, James J Lindsay, Brian E Malone, Laura Muhlenbruck, Kelly Charlton, and Harris Cooper. 2003. Cues to deception. *Psychological bulletin*, 129(1):74.
- Nicholas D Duran, Charles Hall, Philip M McCarthy, and Danielle S McNamara. 2010. The linguistic correlates of conversational deception: Comparing natural language processing technologies. *Applied Psycholinguistics*, 31(03):439–462.
- Patrick M Eady and James Clayton Lafferty. 1969. *The subarctic survival situation*. Experimental Learning Methods.
- Paul Ekman and Wallace V Friesen. 1969. Nonverbal leakage and clues to deception. Technical report.

- Jennifer Golbeck, Cristina Robles, and Karen Turner. 2011. Predicting personality with social media. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 253–262, New York, NY, USA. ACM.
- Amanda Lenhart, Kristen Purcell, Aaron Smith, and Kathryn Zickuhr. 2010. Social media & mobile internet use among teens and young adults. millennials. *Pew Internet & American Life Project*.
- François Mairesse, Marilyn A Walker, Matthias R Mehl, and Roger K Moore. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research*, pages 457–500.
- Robert R McCrae and Paul T Costa Jr. 1999. A five-factor theory of personality. *Handbook of personality: Theory and research*, 2:139–153.
- Dong Nie, Zengda Guan, Bibo Hao, Shuotian Bai, and Tingshao Zhu. 2014. Predicting personality on social media with semi-supervised learning. In *Proceedings of the 2014 IEEE/WIC/ACM Conference on Web Intelligence and Intelligent Agent Technologies*, pages 158–165, Washington, DC, USA. IEEE Computer Society.
- Victoria L Rubin. 2010. On deception and deception detection: Content analysis of computer-mediated stated beliefs. *Proceedings of the American Society for Information Science and Technology*, 47(1):1–10.
- H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9):e73791.
- Leslie Scott. 2016. Across new jersey, bands of local islamophobes crusade to drive islam underground, Mar.
- Ronny E Turner, Charles Edgley, and Glen Olmstead. 1975. Information control in conversations: Honesty is not always the best policy. *Kansas Journal of Sociology*.
- Lina Zhou. 2005. An empirical investigation of deception behavior in instant messaging. *Professional Communication, IEEE Transactions on*, 48(2):147–160.