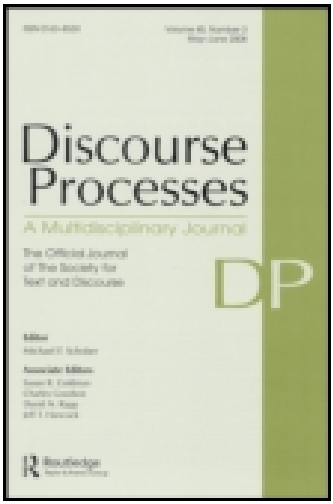

Discourse Processes
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hdsp20

His Lips Are Moving: Pinocchio Effect and Other Lexical Indicators of Political Deceptions

Michael T. Braun[a], Lyn M. Van Swol[b] & Lisa Vang[c]
[a] Department of Communication, Millikin University
[b] Department of Communication Arts, University of Wisconsin-Madison
[c] University of Wisconsin-Madison
Accepted author version posted online: 16 Jul 2014.Published online: 24 Nov 2014.

Discourse Processes, 52:1–20, 2015

ISSN: 0163-853X print/1532-6950 online
DOI: 10.1080/0163853X.2014.942833

Routledge
Taylor & Francis Group

His Lips Are Moving: Pinocchio Effect and Other Lexical Indicators of Political Deceptions

Michael T. Braun
Department of Communication
Millikin University

Lyn M. Van Swol
Department of Communication Arts
University of Wisconsin-Madison

Lisa Vang
University of Wisconsin-Madison

Using the software program LIWC (Linguistic Inquiry and Word Count), this study used political statements classified as truths and lies by website Politifact.com and examined lexical differences between statement type (lie or truth) and the setting (interactive or scripted) in which the statement was given. In interactive settings (where statements given are prompted by questioning), politicians used shorter sentences, simpler words, and fewer causation statements than statements given in scripted settings (where statements are prepared in advance). Lying politicians used more words and negation statements than truth-tellers in both interactive and scripted settings. In interactive settings, politicians used more first-person pronouns. There were no main effects of deception on use of first- and third-person pronouns. Further, there are no effects of deception on use of negative emotion words or profanity. Results are discussed in terms of the importance of setting for studying lexical differences in deception and implications for the study of deception in political communication.

Correspondence concerning this article should be addressed to Michael T. Braun, Department of Communication, Millikin University, 1184 W. Main Street, Decatur, IL 62522, USA. E-mail: michaeltbraun@gmail.com

"How do you know when a politician is lying? His lips are moving."
—Favorite Old Joke

INTRODUCTION

The appeal of a politician to voters is based partly on his or her ability to build a reality for supporters: one in which the politician's platform will lead to prosperity and success, whereas the opponent's policies will lead to decline and failure (Davis & Farrentino, 1996). This process of social influence, attempted by both politicians and those with a political agenda, can result in lies to achieve favorable results. These motivated lies represent a rich source of naturally occurring deception and an opportunity to examine lexical differences between lies and truths, a growing area of research (Buller & Burgoon, 1996; Buller, Burgoon, Buslig, & Roiger, 1996; Hancock, Curry, Goorha, & Woodworth, 2008; Newman, Pennebaker, Berry, & Richards, 2003; Toma & Hancock, 2012; Van Swol & Braun, in press-a; Van Swol, Braun, & Malhotra, 2012; Zhou, Burgoon, Nunamaker, & Twitchell, 2004). Software programs like Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, & Francis, 2007), CohMetrix, Wmatrix (Rayson, 2008), and SiNLP (SoLET Lab, 2012), which classify words into multiple categories and reports percentages of those words appearing in each analyzed file, facilitate such investigation.

To make use of political lies in the context of lexical analysis, this study uses LIWC to analyze lexical factors in statements of public figures the website Politifact.com had classified as true and false. Using Politifact.com offers several advantages that help build on previous research and theory on deception. First, all statements are freely given by the source; this is different from much past laboratory research on deception in which participants are instructed to lie (for review, see DePaulo et al., 2003). Second, Politifact's extensive work to categorize the statements, as well as their close link with the news media, suggests an objective view on which statements are true and which are false. Although Politifact does not choose statements to analyze in an objective way, according to Bill Adair, Politifact's editor (personal communication, February 22, 2012), it still provides a valuable source of objectively categorized and naturally occurring deceptive and truthful statements. Finally, the lies are consequential and have high stakes, as evidenced by Politifact's determination that they are newsworthy (B. Adair, personal communication).

Despite the ubiquity of deception in politics, there have been few attempts (although see Markowitz, Hancock, & Bazarova, 2011) to merge study of lexical differences between deceptive and truthful statements and political communication. This article is an attempt to begin filling that gap and to use these

freely given statements as a way to understand more about the proposed mechanisms driving lexical differences between lies and truths. In the following sections we review the theory behind the hypotheses that liars show different lexical patterns from truth-tellers and offer predictions for analyzing the truths and lies gathered from Politifact.com.

Lexical Patterns in Deception

Interaction management. For the political actors in our sample, words are key tools in efforts of persuasion and argumentation central to the electoral process. For example, words can be used to control an interaction if the speaker refuses to yield the floor or interrupts someone speaking out in dissent. The simplest use of words to this end is measured by word count. Liars have been found to use more words than truth-tellers, a phenomenon called the Pinocchio effect (Van Swol et al., 2012). According to the Pinocchio effect, liars strategically use more words than truth-tellers for two reasons. First they use their verboseness to create a believable reality through justifications and evidence for their statements, and second, they talk more than truth-tellers to manage the interaction between themselves and their audience and to enhance a connection with their audience to help reduce suspicion (Burgoon, Buller, Dillman, & Walther, 1995; Van Swol et al., 2012).

In the case of politicians and others whose words are carefully scrutinized, increased word count gives more context for the statement; certain qualifiers or exceptions may add nuance to the politician's expressed view. Further, given politicians' tendency toward high self-monitoring and interest in the reaction of their audience (Briggs & Cheek, 1988), they desire to manage the interaction to connect with their audience and reduce suspicion. Several studies have found that liars use more words (Duran, Hall, McCarthy, & McNamara, 2010; Hancock et al., 2008; Van Swol et al., 2012; Zhou et al., 2004). However, other research has found that when the goal of concealment is salient, rather than the goal of creating a believable reality, deceivers use fewer words to be intentionally vague and avoid contradicting themselves or contradicting evidence (Burgoon, Buller, Grerrero, Afifi, & Feldman, 1996; Hartwig, Granhag, & Stromwall, 2007; Pennebaker, 2011; Porter & Yuille, 1996; Toma & Hancock, 2012; Van Swol & Braun, in press-a). Given that the nature of statements easily analyzed as truth or lie tends toward expressed views or positions (e.g., that healthcare legislation creates "death panels") rather than attempts to keep something secret or hidden (e.g., denying claims of infidelity), we expect that liars in the contexts we are investigating will use more words than truth-tellers.

Verbal nonimmediacy. Liars may wish to distance themselves from their false statements through a strategy of verbal nonimmediacy to avoid or reduce the

association with the negative and socially disfavored action (Buller & Burgoon, 1994; Burgoon, Buller, Blair, & Tilley, 2006; Mehrabian & Wiener, 1966; Toma & Hancock, 2012). This theory has been used to explain differences in pronoun use between liars and truth-tellers. Specifically, in some contexts liars have been hypothesized to use fewer first-person pronouns (e.g., I, me, we) as a way to distance themselves from their lies. Additionally, they have been predicted to use more third-person pronouns (e.g., she, his, they) to shift the focus of deception from self to other and increase vagueness about the agent of the deception (Burgoon, Blair, Qin, & Nunamaker, 2003; DePaulo et al., 2003; Hancock et al., 2008; Ickes, Reidhead, & Patterson, 1986; Larcker & Zakolyukina, 2010; Newman et al., 2003; Pennebaker, 2011). For example, Van Swol et al. (2012) found that liars use significantly more third-person pronouns when deceiving in a dictator negotiation game. In this type of game, one participant divides an amount of money between herself and a partner; her partner must decide whether to accept this dictated amount or to refuse the offer and cause both players to forfeit all the money.

Emotional language. Researchers have also investigated cues to deception that are thought to arise because of negative emotions (Buller & Burgoon, 1994; Buller et al., 1996; Toma & Hancock, 2012). These cues involve three dimensions (Vrij, Semin, & Bull, 1996): nonverbal behavior (e.g., fidgeting), vocal tone (e.g., higher pitched voice), and lexical production (e.g., emotion words). This study focuses only on lexical cues.

As argued by several scholars, liars may feel guilt and anxiety about the deception, the violation of social norms, and the possibility of getting caught (Buller et al., 1996; DePaulo et al., 2003; Vrij, 2000; Zhou, Burgoon, Nunamader, & Twitchell, 2004; for evidence to the contrary, see Van Swol & Braun, in press-a). Given the association between lying and negative emotion, some researchers have predicted this feeling to carry over into word choice, leading to hypotheses that liars will use more negative emotion words (e.g., hurt, ugly, nasty) (Knapp & Comadena, 1979; Newman et al., 2003; Vrij, 2000; Zhou et al., 2004) and more negations (e.g., no, not, never) (Toma & Hancock, 2012). In the case of negative emotion words, the associated feeling of guilt may manifest itself in choosing words that convey, unintentionally, a more negative tone. Negations also allow liars to distance themselves from their statements; for example, by describing something as "not bad" rather than "good," liars avoid commitment to their own statements.

Contextual Differences Perspective

Some patterns of lexical indicators to deception have consistently emerged in published research, but there are also differences in findings and predictions. For example, differing results have emerged for negative emotion words, word

count, and negations (Duran et al., 2010; Hancock et al., 2008; Larcker & Zakolyukina, 2010; Pennebaker, 2011; Toma & Hancock, 2012; Van Swol & Braun, in press-a; Van Swol et al., 2012). These differences must be considered with two qualifications: First, the field is new and developing, and second, the lies and scenarios used to elicit deception have varied considerably across studies. The challenge in finding consistent results across lexical studies may be partly explained by contextual differences in deception production (Buller et al., 1996; Giordano & George, 2005; Giordano, Stoner, Brouer, & George, 2007; Pennebaker, 2011; Van Swol & Braun, in press-a; Van Swol et al., 2012; Vrij, 2008). As James Pennebaker states (2011, p. 142), "Perhaps most important is the context in which lies are told." Lexical cues can have multiple meanings, and without information about context or sender's intentions, it can be difficult to understand the cause of strategic lexical cues (Burgoon et al., 2006). This contextual perspective does not suggest that deception is unrelated to lexical differences; instead, it argues that contextual effects should guide hypothesizing and that, at times, contextual effects may be greater than deception effects.

One contextual effect that may interact with deception effects is the setting in which the lie is told (Toma & Hancock, 2012). Some settings allow the liar the freedom to construct a lie without immediate time pressures or potentially hostile audiences; this is more often true in cases where the lie is written rather than spoken. For example, Toma and Hancock (2012) argue that the construction of an online dating profile takes place in an environment with high editability (the option to revise a statement before posting it) and asynchronicity (the option to take time in message construction); this "may reduce the cognitive burden of lying" (p. 82) and thus change patterns of language use. In other settings, however, the liar must construct his or her lie in person. For example, when being interviewed or participating in a debate, the liar is under greater pressure to keep the details of his or her story straight to avoid potential contradictions; the liar must also monitor his or her own behavior to maintain a truthful demeanor (Hartwig et al., 2007; Levine et al., 2011) and monitor his or her partner for signs of suspicion (Burgoon, Buller, Dillman, & Walther, 1995; Duran et al., 2010).

Because of these differences, we consider lexical differences between truths and lies as well as between different settings. We wish to learn if truths and lies interact with setting to show different lexical patterns. The consideration of both deception and contextual effects of the setting in which the statement was given allow us to separate out deception from context, whereas the two factors have been confounded in much past deception research (Vrij et al., 1996; although see Giordano & George, 2005; Giordano et al., 2007). This distinction adds to theory in the expanding field of deception and lexical analysis.

Hypotheses

With these factors in mind, we suggest the following hypotheses.[1] First, we assess how the setting may affect the complexity of speech. This serves both as a useful manipulation check on our operationalization of some settings placing different demands on speakers and also as a test of the added effect of setting on lexical differences hypothesized to occur because of deception. We separate statements into two setting categories, following the work of Biber (1986, 1988, 1995, as cited in Jahandarie, 1999): interactive and scripted. In interactive settings, like a political debate, the speaker must interact with others in real time. In scripted settings, like delivering a speech or writing an op-ed, the speaker has time to prepare and edit his or her statement.

The two types of settings investigated here may alter speech patterns. Communicating in an interactive setting, in which speakers have no options to edit their speech and must interact with others in real time, may affect lexical content. For example, some research has investigated the factor of cognitive load and lexical patterns (Duran et al., 2010; Graesser, Zhiquiang, Louwerse, & Daniel, 2006). Duran and colleagues (2010) hypothesized a higher cognitive load will lead people to select a vocabulary that is simpler and more accessible from memory. Although their results failed to support this hypothesis, neither did they find support for the opposite contention. We propose similar hypotheses related to the setting. In interactive settings, sentences should consist of fewer words and words used should consist of fewer letters, operationalized here as the number of words of length six letters or longer (a standard category in LIWC). Additionally, sentence structure should be less complex (Duran et al., 2010), which we test using causation words (e.g., because, effect, hence) and conjunctions (e.g., and, but, whereas). We also test these same factors with deception to see if the act of lying similarly alters these lexical patterns. This leads to the following hypotheses:

- **H1–H4**: Political actors in interactive contexts and telling lies will produce speech with shorter sentences, shorter words, fewer causation words, and fewer conjunctions compared with those speaking in scripted contexts and telling truths. Setting and statement type will interact such that liars in interactive settings will show the most extremes in these four categories.

Second, we address the "Pinocchio effect," the finding that liars use more words than truth-tellers. As argued above, we reason that inconsistent findings in

[1]Our hypotheses are based on past research that focuses on falsification and bold-faced lies. These types of deception may be different from equivocations (Buller et al., 1994, 1996; Van Swol et al., 2012) and self-deception (Pennebaker, 2011). For further comment on how this may affect the results of this study, please refer to Discussion.

word counts between liars and truth-tellers can be explained by motivation. Liars motivated to conceal (as in online dating profiles; Toma & Hancock, 2012) may use fewer words than truth-tellers, whereas liars motivated to craft a believable reality (as in a negotiation game, where the allocator is trying to get the receiver to accept his or her offer; Van Swol & Braun, in press-a; Van Swol et al., 2012) will use more words. Here, we expect the political actors telling lies in our sample to work toward the latter and propose this hypothesis:

- **H5**: Political actors telling lies will use more words than those telling truths. We also ask if there will be a main effect of setting or an interaction between the two.

Third, because lying has been associated with negative emotions (but see Van Swol & Braun, in press-a), we expect that these emotions will manifest themselves lexically. However, campaigns often depend on a separation between attacks on opponents and the politician's own public persona (Dalton & McIlwain, 2011). This could suggest that politicians try to maintain a positive public persona and may use emotion words more strategically than has been found in past lexical research, but this should be easier in contexts with high editability. Candidates may not wish to appear excessively negative, especially as campaigns try to stress that things will improve if the candidate is elected; this desire, however, may not be possible when interacting in synchronous settings. As synchronous settings may make it harder for the political actor to respond as strategically as possible, then setting may be an important additional factor to consider, especially as it could make it difficult to maintain proper decorum (Van Swol et al., 2012). Thus, we suggest the following hypotheses:

- **H6 and H7**: Political actors telling lies will use higher rates of negations and negative emotion words than political actors telling the truth. Because these words may be easier to control than pronouns and may in some cases be strategically used (Toma & Hancock, 2012), we also ask whether there will be a significant effect of setting or a significant interaction effect.

Finally, we assess how deception and setting interact with the use of pronouns, specifically first- and third-person pronouns. Past research and theory (Burgoon et al., 2003; DePaulo et al., 2003; Hancock et al., 2008; Ickes et al., 1986; Larcker & Zakolyukina, 2010; Newman et al., 2003) has suggested that liars distance themselves from their lies by using fewer first-person and more third-person pronouns than truth-tellers. We consider this a reasonable basis for a hypothesis but note that setting-specific factors like time pressure (which exists in interactive settings) reduces opportunities for deliberative and systematic response and increases reliance on personal views without consideration for what is socially

appropriate. Research has found this to be true even when speakers are exposed to counter-attitudinal positions that are suggested to be normative (Sanbonmatsu & Fazio, 1990; Smith & Terry, 2003). Smith and Terry (2003) demonstrated that individuals who must decide under time pressure are less swayed by the attitudes of others and more guided by their own initial attitudes. Thus, in interactive settings where the speaker must respond immediately, political actors may respond based on their own attitudes rather than consider the most favorable response to a question. This may manifest itself through increased use of first-person pronouns in more interactive settings.

Similarly, some researchers have suggested that under conditions of higher self-awareness (Vorauer & Ross, 1999), individuals may use more first-person pronouns (e.g., Davis & Brock, 1975). In interactive settings, where a political actor needs to monitor his or her own and his or her audience's behavior closely (Duran et al., 2010; Levine et al., 2011), the same outcome may result. We believe this research on time pressure and self-awareness may apply especially to politicians, given the high stakes scrutiny of candidates by the media, intensified in the 24/7 news cycle and YouTube era (e.g., George Allen's infamous "Macaca moment"; Karpf, 2010). Because of these potential alternative effects of setting, we suggest the following research question:

- **RQ1**: Will there be differences in rates of first- and third-person pronoun use between scripted and interactive settings, between deceptive and truthful statements, or an interaction between the two?

METHODS

Procedure

In the fall of 2010, seven undergraduate research assistants visited the website Politifact.com and gathered information from all currently classified statements coded as "truth," "lie," or "pants-on-fire," Politifact's label for an especially egregious lie; for our analysis, the two deception categories were combined. Statements that contained a degree of truth or falsehood (e.g., "half-true" "mostly false") were not coded. A total of 579 statements were coded; 294 were truths and 285 were lies (see Table 1 for sample statements).

To prepare the text for lexical analysis, undergraduate research assistants (blind to study hypotheses) first visited the Politifact page for the statement and attempted to locate the original source of the message (e.g., a transcript of the speech or interview, the video advertisement, the Twitter message, etc.) to verify that Politifact had included all truthful/deceptive remarks. In some cases Politifact reported an abbreviated version of the statement, removing, for

TABLE 1
Examples of Deceptive and Truthful Statements

Speaker	*Statement Type*	*Statement Text*
Hilary Clinton	True	"A ham and cheese sandwich on one slice of bread is the responsibility of the U.S. Department of Agriculture, which inspects manufacturers daily. But a ham and cheese sandwich on two slices of bread is the responsibility of the Food and Drug Administration, which inspects manufacturers about once every five years."
Fred Thompson	True	"Five percent of Americans pay over half the income taxes in this country. Forty percent of Americans pay no income taxes at all."
John McCain	True	"I forgot to mention last night that following World War II war crime trials were convened. The Japanese were tried and convicted and hung for war crimes committed against American POWs. Among those charges for which they were convicted was waterboarding."
Barack Obama	False	"If we went back to the obesity rates that existed in 1980, that would save the Medicare system a trillion dollars."
Mike Huckabee	False	"I stood by him in the war. I stood by him in the surge. I wasn't a latecomer like Mitt Romney was to believing that the surge was effective."
Joe Biden	False	"John doesn't have a record in the Senate. John's only passed four bills. They're all about post offices."

example, some words in the middle of a sentence. In these cases the full text was used when available. Each statement was then (when needed) edited for proper analysis in the LIWC software (e.g., punctuation added, misspellings corrected, etc.). The authors reviewed all captured statements for completeness.

Measures

Each statement's context was recorded and these contexts were put into five categories: spoken statement (e.g., speeches; $n = 211$), written statements (e.g., blog post, editorial, Twitter message; $n = 82$), advertisements ($n = 10$), debates ($n = 102$), and interviews ($n = 169$); context could not be coded for five statements. These categories were then split into two settings. The first setting was for written comments, including any situation in which the speaker delivers comments prepared in advance, including spoken statements, written statements, and advertisements (coded as 0; $n = 303$). The second setting was for synchronous, interactive interactions, where the speaker can rehearse (as political actors obviously do quite a bit) but must face a live questioner whose behavior changes based on the statements made by the speaker. Thus, the speaker has less of an ability to plan his or her exact words in advance. This dataset included statements made in debates and interviews (coded as 1; $n = 271$).

Each speaker was also coded as a Democrat or Democratic-Supporter ($n = 264$), a Republican or Republican-Supporter ($n = 304$), or Other ($n = 11$). Others are international political figures like former Iranian President Mahmoud Ahmadinejad, noted independents (e.g., former New York mayor Michael Bloomberg), or nonpartisan news hosts.

RESULTS

Rates of Deception

First, we considered if some demographics or contexts contained significantly more deception than others. Although we do not hypothesize differences in deception based on gender or political leaning, it was important to assess these categories to consider whether differences need to be accounted for in subsequent analyses. Additionally, significant differences in this category could reflect selection bias by Politifact.

There were no differences in rates of deception by gender. Men did not deceive more in the statements selected by Politifact than women, $\chi^2\,(1) = .031$, *ns*. There was a significant difference by politics. In statements selected by Politifact, Democrats or Democratic-supporters lied less than Republicans or Republican-supports, $\chi^2\,(1) = 10.296$, $p = .001$. There was no difference in rates of deception by context. No context showed a significantly higher rate of deception than any other, $\chi^2\,(1) = 7.769$, *ns*. Additionally, there were no meaningful lexical differences among lies when comparing Democratic and Republican lies. Table 2 shows rates of deception by context, political affiliation, and gender.

Hypotheses

Our first four hypotheses predicted that political actors speaking in interactive settings and lying will use shorter sentences, shorter words, fewer causation words, and fewer conjunctions. Additionally, we expected a significant interaction between setting and statement type such that liars in interactive settings would show the most extreme results. To test these hypotheses, two-way analyses of variance were used. Three categories showed main effects of setting but no main effect of deception or significant interaction effects. Specifically, speakers in interactive settings used shorter sentences ($F\,(1, 569) = 6.310$, $p < .05$, $\eta^2 = .011$), fewer words of six letters or longer ($F\,(1, 569) = 8.232$, $p < .01$, $\eta^2 = .014$), and fewer causation words ($F\,(1, 569) = 4.610$, $p < .05$, $\eta^2 = .008$). No significant main or interaction effects were revealed for conjunctions. Hypotheses 1 through 3 were supported for setting but not for

TABLE 2
Rates of Deception by Context, Politics, and Gender

	Statement Type	
Context	*Truth*	*Lie*
Speech or public statement	103	108
Advertisement	7	3
Written statement	36	46
Debate	62	40
Interview	82	87
Politics		
Democrat or Democratic supporter	152	112
Republican or Republican supporter	134	170
Other	8	3
Gender		
Male	241	232
Female	53	53

deception. Hypothesis 4, predicting differences for conjunctions, was not supported. Table 3 shows lexical category means by statement type and setting.

Next, we predicted that word count would be higher among liars than among truth-tellers but that there would be no difference by setting or a significant interaction. This was supported. A main effect of statement type was found (F (1, 569) = 5.551, $p < .05$, $\eta^2 = .010$). Political actors telling lies ($M = 52.0$, $SD = 43.13$) used significantly more words than those telling the truth ($M = 44.3$, $SD = 32.16$), supporting the Pinocchio effect and hypothesis 5.

TABLE 3
Means and Standard Deviations for Each Category by Statement Type and Setting

	Truth		*Lie*	
Setting	*Scripted* (n = *144*)	*Interactive* (n = *146*)	*Scripted* (n = *157*)	*Interactive* (n = *127*)
Words per sentence	24.46 (14.12)	20.32 (10.14)	23.74 (14.56)	22.38 (12.90)
Six-plus letter words	22.13 (8.54)	20.01 (9.05)	22.12 (8.66)	20.19 (7.24)
Conjunctions	4.41 (3.33)	4.07 (3.27)	4.59 (3.46)	4.43 (3.24)
Causation	2.02 (2.32)	1.37 (2.28)	1.78 (2.27)	1.61 (2.32)
First-person pronouns	3.30 (4.18)	4.08 (3.91)	3.30 (5.35)	4.72 (4.76)
Third-person pronouns	1.88 (3.11)	2.77 (4.05)	2.29 (3.33)	1.96 (3.20)
Negations	1.81 (2.70)	1.94 (2.74)	2.26 (3.28)	2.60 (3.75)
Negative emotions	1.99 (3.04)	1.26 (2.26)	1.68 (2.35)	1.42 (2.24)
Word count	44.29 (26.61)	44.30 (37.02)	53.55 (42.52)	50.06 (43.70)

Political actors showed no differences in word count between interactive and scripted settings, and a significant interaction effect was not found.

Hypotheses 6 and 7 predicted differences in indicators of negative emotion: Liars will use higher rates of negations and negative emotions words than will truth-tellers; differences in setting were also considered as a research question. Negations showed significant differences by statement type, F (1, 569) = 4.461, $p < .05$, $\eta^2 = .008$. Liars used a significantly higher rate of negation words ($M = 2.41$, $SD = 3.495$) than truth-tellers ($M = 1.87$, $SD = 2.718$). Liars and truth-tellers did not differ in their use of negation words by setting, and an interaction between the two factors was not found. Negative emotion words, however, did not show a significant main effect for statement type. In this case, there was only a significant main effect for setting, with speakers in scripted settings ($M = 1.83$, $SD = 2.706$) using significantly more negative emotion words than speakers in interactive settings ($M = 1.34$, $SD = 2.304$), F (1, 569) = 5.466, $p < .05$, $\eta^2 = .010$. There was no significant interaction effect. In summary, liars used more negation words, consistent with hypothesis 6. A significant effect of setting on negative emotion words was found but in opposite the predicted direction; thus, hypothesis 7 was not supported.

Finally, we questioned whether rates of first- and third-person pronoun use would differ between interactive and scripted setting, between deceptive and truthful statements, or an interaction of the two. All tests were performed using two-way analyses of variance with setting and statement type predicting pronoun use. For first-person pronouns, there was only a main effect of setting, F (1, 569) = 8.180, $p < .01$, $\eta^2 = .014$. Speakers in interactive settings used significantly more first-person pronouns ($M = 4.378$, $SD = 4.330$) than speakers in scripted settings ($M = 3.297$, $SD = 4.815$). There were no differences in statement type or a significant interaction. For third-person pronouns, there were no significant main effects, but there was a significant interaction effect, F (1, 569) = 4.527, $p < .05$, $\eta^2 = .008$. Use of third-person pronouns did not differ between liars and truth-tellers or between settings, but truth-tellers in interactive settings ($M = 2.774$, $SD = 4.048$) used more third-person pronouns than truth-tellers in scripted settings ($M = 1.879$, $SD = 3.113$). Lying political actors in scripted ($M = 2.289$, $SD = 3.327$) and interactive ($M = 1.955$, $SD = 3.203$) settings did not differ in their rates of third-person pronouns.

DISCUSSION

This study examined lexical differences between political lies and truths, as determined by organization Politifact.com, and attempted to demonstrate the effects of setting and statement type on lexical patterns. In confirmation of the "old favorite joke," politicians engaging in deception were indeed moving their

"lips" more than truthful politicians, the Pinocchio effect (Van Swol et al., 2012) was supported, and political actors telling lies used more words than when telling truths. In addition, by separating out statements by their context (made in scripted or interactive settings), we were able to assess whether the setting or the production of deception are more strongly associated with lexical differences. We found support for both contextual effects of setting and effects of deception, with stronger support for contextual effects. This supports the arguments of Van Swol and Braun (Van Swol & Braun, in press-a; Van Swol et al., 2012) and other researchers like Giordano and colleagues (Giordano & George, 2005; Giordano et al., 2007), who argue that the context in which the statement occurs shapes lexical patterns more than the deceptive or truthful nature of the statement given. This study also combines the growing field of lexical cues to deception with that of political speech, a valuable context of study for deception and a potentially useful method for future political research.

Interaction Control

The Pinocchio effect posits that liars use more words than truth-tellers and that this effect is because liars use more words to control the interaction and provide justifications to create a believable reality (Van Swol & Braun, in press-a; Van Swol et al., 2012). The lexical patterns found in this study support the Pinocchio effect. First, liars did indeed use more words than truth-tellers. Second, and equally important, there was no significant effect of setting or an interaction effect between the two predictors. This supports the contention that increased word count may be a strategic decision. Liars may be driven to add more details to their statements, either to increase believability or to add nuance for later deniability (perhaps especially important in the political context). Future study should investigate the decisions liars make about what to say.

Other studies have found the opposite pattern, with liars using fewer words than truth-tellers (Toma & Hancock, 2012). In the online dating context considered by Toma and Hancock, daters may have, to paraphrase Abraham Lincoln, sought to write little and appear liars rather than write more and remove all doubt. Toma and Hancock theorize that in cases of verifiable lies, liars motivated by concealment will use fewer, not more, words. Because all lies in our study are, by nature, verifiable, we must reconcile the predictions of the Pinocchio effect with past research that has found liars to use fewer words. First, we note that the traditional political lie is often not straightforward. Although we included only statements that were classified as either fully true or fully false (no "half-truths" were included), even the lies in our study can be considered part of the exaggerated claims that make up much political rhetoric (Davis & Farrentino, 1996). Under this framework, lies may be uncovered eventually, but the benefits in the short term (e.g., a positive audience reaction) may be more

important; Davis and Farrentino argue that politicians must carefully balance short-term gains against long-term costs.

Can the online daters used in Toma and Hancock's (2012) study and public figures in this study be compared in this framework? Consider first the prospects of short-term success. This is of primary importance for both online daters and political actors: Online daters seek dates and political actors seek attention and news coverage. But online daters have greater certainty in the consequences of a lie in their dating profile; the lie may result in increased attention in the short-term but no lasting long-term relationships. For the speakers in our study, the connection is not nearly as clear. Although websites like Politifact.com and news organizations may spend time checking the veracity of political claims, not all statements can be checked, and those that are checked cannot all be broadcast. This results in a high likelihood that inflammatory statements result in short-term gains with lessened risk of long-term consequences. A politician who makes his or her name making bold claims about the opposition can temper the message as his or her popularity rises, a pattern seen repeatedly as candidates aim for their base during a primary election but then seek support from moderates and independents during the general election. Further, a speaker's target audience for his or her lie and the audience verifying the lie (Politifact) are different; even if Politifact verifies that a statement is a lie, this information may not reach or matter to the speaker's target audience (Garrett, 2011).

These differences suggest that all verifiability is not created equal. For online daters the process of uncertainty reduction (Berger & Calabrese, 1975) in romantic courtship suggests any lies will be quickly exposed (e.g., a dater can say he or she loves hiking until a request for a trail date reveals the dishonesty), and thus verifiability works to keep online daters from expanding widely on their deceptive claims, which leads to lower word count in cases of deception. As Toma and Hancock (2012) conclude, "online daters acted strategically to hide or distract attention from their deceptions" (p. 87). For political actors, however, their audience may be less motivated to uncover the specific truths behind the claims (Garrett, 2011), especially given the motivation for exaggeration and campaign promises (Davis & Farrentino, 1996) and tendency to believe false rumors about opposing party candidates (Garrett, 2011). The goals of the two groups are qualitatively different because of these motivations. Although online daters might be motivated to conceal information about themselves that could hurt their chances with a romantic partner, politicians' charisma is based partly on the necessity of building a reality for supporters, one in which the politician's platform will lead to prosperity and success, whereas the opponent's policies will lead to decline and failure (Davis & Farrentino, 1996). Just as Van Swol et al. (2012) found that lying by omission produced the fewest number of words, so too would someone trying to conceal a verifiable, unflattering truth. Thus, the goal of the statement (to create or to conceal), not the verifiable nature, may be the

primary determinant of the Pinocchio effect across all contexts. This argument deserves further research to build a theory of lexical differences.

Other Lexical Indicators

Although the strategic Pinocchio effect was not influenced by setting, several other key variables were. First, we confirmed that lexical patterns associated with word and sentence complexity do indeed differ between scripted and interactive settings, while showing no additional effects of deception. Sentence length was shorter for individuals in interactive settings, and they used fewer long words. Additionally, they used fewer causation words, possibly indicating differences in logical flow between different settings. No difference in conjunctions was found. Altogether, these results may suggest that setting pressures influenced the types of ideas expressed, with speakers in interactive settings less able to describe their ideas with complex, nuanced logic.

With this evidence of setting manifesting itself lexically, we used setting as an additional predictor of a variety of lexical patterns previously found to be associated with lies and truths. For pronouns, no main effect of deception was found for either first- or third-person pronouns. This failed to support past research (e.g., Burgoon et al., 2003; DePaulo et al., 2003; Hancock et al., 2008; Ickes et al., 1986; Larcker & Zakolyukina, 2010; Newman et al., 2003; Toma & Hancock, 2012; Van Swol et al., 2012). Setting, on the other hand, was a significant predictor of first-person pronouns, with individuals in interactive settings showing increased use. This may occur because under pressure to respond in interviews and debates, individuals may find it easier to think and talk about themselves (e.g., Davis & Brock, 1975).

For negative words, we also considered both statement type and setting as predictors. Negations were higher for liars than truth-tellers, supporting hypothesis 6 and past research (Hancock et al., 2008; Toma & Hancock, 2012). Additionally, past research has argued that negations may be more controllable than pronouns (Hancock et al., 2008). This is supported in our results, because there was no main effect of setting or a significant interaction effect.

The pattern was markedly different for other indicators of the negative emotions believed to be associated with lying. Negative emotion words showed no main effect of statement type but did show significant differences by setting. Unexpectedly, the differences manifested themselves in the opposite pattern expected by past claims: Individuals in scripted settings used more negative emotion words than individuals in interactive settings. We can only offer a post-hoc explanation for this observed difference. Perhaps political actors, who are extensively coached for debate and interview appearances, know stakes are much higher for these publicized events and thus resist "going negative." Toma and

Hancock (2012) similarly found deceptive dating profiles had fewer negative emotion words than truthful profiles, possibly because participants with deceptive profiles were more concerned with appearing "datable."

Contributions, Limitations, and Future Research

The results of this study emphasize several key points: that setting plays a role in lexical patterns, that contextual information is necessary to make predictions about lexical differences in deception and truth, and that setting may have a larger effect on lexical cues to deception that are less controllable than consciously used strategic cues.

This study is limited in several important ways. First, political lies may differ from lies told in everyday life for several reasons. Political motivations driven by high stakes interactions may mean the dynamics that affect the production of lies and truths may be different from those produced in other contexts. These politicians, like many of their most fervent supporters, may be "true believers." Thus, alleging, for example, that the Affordable Care Act requires citizen healthcare decisions to go before a "death panel" may not register as a lie to the person who makes the claim. Although some previous research has found that self-deception has similar lexical qualities as deception (Pennebaker, 2011), others (Schober & Glick, 2011) have identified ways in which self-deceptive speech may differ. For example, Schober and Glick (2011) suggest that self-deception may be defensive speech and found that people using defensive speech had a higher word count, used more exclusive words, used more negations, and had more repetition. One limitation was that we could not distinguish between deliberate and self-deception to examine lexical characteristics of self-deception.

Second, we rely on a relatively small number of statements compared with the overall number of political statements made every day. Although our selection of statements is helped by reliable "true or false" classification, our study may still be underpowered to find lexical differences. Third, as with much research in this fast developing field, not all our significant findings matched our hypotheses and all our results were not consistent with past literature. We advise readers to interpret these results with appropriate caution; this is just one study among many in a rapidly growing and changing area of research.

Finally, the complexities of lies occurring in everyday life, and especially lies that may be as carefully constructed as those here, present challenges for researchers in selecting what parts of text to edit. A lie may be set up by a preceding sentence, told in the second sentence, and given context or nuance in a third. For our purposes, based on the assessment by Politifact, we analyzed only that second sentence (or however long the specific deceptive content ran). These choices may not reflect the full extent to which deception altered lexical content. For example, the guilt from a lie may manifest itself in more negative emotions

words in later truthful content. When we had access to the original transcript, we made the decision to only analyze the deceptive/truthful content and not later content on another topic, especially because Politifact had only selected the deceptive/truthful content to post on their site.

In our description of this issue, we echo McCornack, Morrison, Paik, Wisner, and Zhu (2014):

> There is no reason, in accounting for this message, to presume that it was constructed with a priori intent to deceive; nor is there any reason to presume that deceptive intent persisted for more than the few micromoments necessary to generate the one piece of false information. This [. . .] illustrates one of the most profound implications of adopting a problem-solving view of deceptive discourse production: *intent to deceive may occur before or during the production of discourse that is functionally deceptive—but most certainly need not precede discourse production*. (p. 356, emphasis in original)

Our political actors may intend to be truthful in interactive settings, find the truth to be problematic for their message or audience, insert some degree of deception, and then return to truthful information; this can make lies and truths difficult to demarcate and reveals a limitation in current knowledge about how widespread lexical differences between lies and truths may be. Do the differences manifest themselves in the buildup to the lie, in the lie itself, or in sentences after the lie? Future research can help to address this important question.

Despite these limitations, we believe our study contributes to the study of lexical differences in deception. This study demonstrates that setting, as operationalized based on editability and synchronicity of message production, is a more significant predictor of the lexical usage than the truthfulness of the statement. The predicted interaction between deception and setting was not supported. When setting is included as a predictor, presence or absence of deception had little additional effect. Deception main effects were seen only for word count (supporting the Pinocchio effect; Van Swol et al., 2012) and negations. Whether the speaker was producing a carefully constructed written statement or answering in an interview or debate, liars used more words and more negations than truth-tellers.

We add to the study of lexical differences and deception by including setting as a predictor of lexical differences. Multiple contexts have been studied as part of experiments on deception (e.g., Giordano & George, 2005; Giordano et al., 2007; Van Swol & Braun, in press-b; Van Swol, Braun, & Kolb, in press; Van Swol, Malhotra, & Braun, 2012), and we are pleased to continue this in the study of lexical cues to deception. A statement that is produced "on the fly" or in response to a difficult question from a reporter differs from a statement that is carefully assembled in an environment of high asynchronicity and editability.

Future research should examine how setting may influence cognitive load, perhaps in a controlled lab setting where cognitive load may be successfully manipulated and measured. We encourage more research testing some of the underlying explanations for lexical differences, such as psychological distancing or presence of negative emotions. As eloquently argued by McCornack (1997), these assumptions are far from safe when one reviews the deception literature. Further, other contextual effects may interact with deception, such as presence or absence of suspicion in one's partner. Overall, enough research on lexical differences has emerged that testing contextual effects and underlying assumptions seems a wise step for additional research to help validate past explanations and theory.

ACKNOWLEDGMENTS

We thank Andrea Carlson, Daniel Kaplan, Peter Moomjian, Lindsay Montgomery, Carla Pentimone, and Randi Russell for their help with data collection and coding.

REFERENCES

Berger, C., & Calabrese, R. (1975). Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication. *Human Communication Research*, *1*, 99–112.

Briggs, S. R., & Cheek, J. M. (1988). On the nature of self-monitoring: Problems with assessment, problems with validity. *Journal of Personality and Social Psychology*, *54*, 663–678.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, *6*, 203–242.

Buller, D. B., Burgoon, J. K., Buslig, A., & Roiger, J. (1994). Interpersonal deception VIII: Further analysis of nonverbal and verbal correlates of equivocation from Bavelas et al. (1990) research. *Journal of Language and Social Psychology*, *13*, 396–417.

Buller, D. B., Burgoon, J. K., Buslig, A., & Roiger, J. (1996). Testing interpersonal deception theory: The language of interpersonal deception. *Communication Theory*, *6*, 268–289.

Burgoon, J. K., Blair, J. P., Qin, T., & Nunamaker, J. F., Jr (2003). Detecting deception through linguistic analysis. In H. Chen, R. Miranda, D. D. Zeng, C. Demchak, J. Schroeder, & T. Madhusudan (Eds.), *Proceedings of the Symposium on Intelligence and Security Informatics* (pp. 91–101). New York, NY: Springer-Verlag.

Burgoon, J. K., Buller, D. B., Blair, J. P., & Tilley, P. (2006). Sex differences in presenting and detecting deceptive messages. In D. Canary & K. Dindia (Eds.), *Sex differences and similarities in communication* (2nd ed., pp. 263–280). Mahwah, NJ: Lawrence Erlbaum Associates.

Burgoon, J. K., Buller, D. B., Dilman, L., & Walther, J. B. (1995). Interpersonal deception. *Human Communication Research*, *22*, 163–196.

Burgoon, J. K., Buller, D. B., Guerrero, L. K., Afifi, W., & Feldman, C. (1996). Interpersonal deception: XII. Information management dimensions underlying deceptive and truthful messages. *Communication Monographs*, *63*, 50–69.

Dalton, P., & McIlwain, C. (2011). Third-party "hatchet" ads: An exploratory content study comparing third-party and candidate spots from the 2004 presidential election. *Atlantic Journal of Communication*, *19*, 129–151.

Davis, D., & Brock, T. C. (1975). Use of first person pronouns as a function of increased objective self-awareness and performance feedback. *Journal of Experimental Social Psychology*, *11*, 381–388.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*, 74–118.

Duran, N. D., Hall, C., McCarthy, P. M., & McNamara, D. S. (2010). The linguistic correlates of conversational deception: Comparing natural language processing technologies. *Applied Psycholinguistics*, *31*, 439–462.

Garrett, R. K. (2011). Troubling consequences of online political rumoring. *Human Communication Research*, *37*, 255–274.

Giordano, G. A., & George, J. F. (2005). Task complexity and deception detection in a collaborative group setting. In *Proceedings of the 38th Annual Hawaii International Conferences on System Sciences.* Piscataway, NJ: IEEE.

Giordano, G. A., Stoner, J. S., Brouer, R. L., & George, J. F. (2007). The influences of deception and computer-mediation on dyadic negotiations. *Journal of Computer-Mediated Communication*, *12*, 362–383.

Graesser, A. C., Zhiqiang, C., Louwerse, M. M., & Daniel, F. (2006). Question Understanding Aid (QUAID): A web facility that helps survey methodologists improve the comprehensibility of questions. *Public Opinion Quarterly*, *70*, 3–22.

Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2008). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, *45*, 1–23.

Hartwig, M., Granhag, P. A., & Stromwall, L. A. (2007). Guilty and innocent suspects' strategies during interrogations. *Psychology, Crime, & Law*, *13*, 213–227.

Ickes, W., Reidhead, S., & Patterson, M. (1986). Machiavellianism and self-monitoring: As different as "me" and "you." *Social Cognition*, *4*, 58–74.

Jackson, B., & Jamieson, K. H. (2004). Finding fact in political debate. *American Behavioral Scientist*, *48*, 229–237.

Jahandarie, K. (1999). *Spoken and written discourse: A multi-disciplinary perspective*. Stamford, CT: Ablex.

Karpf, D. (2010). Macaca moments reconsidered: Electoral panopticon or netroots mobilization? *Journal of Information Technology and Politics*, *7*, 143–162.

Knapp, M. L., & Comadena, M. A. (1979). Telling it like it isn't: A review of theory and research on deceptive communications. *Human Communication Research*, *5*, 270–285.

Larcker, D. F., & Zakolyukina, A. A. (2010). *Detecting deceptive discussions in conference calls* (Tech. Rep. No. 83). Stanford, CA: Stanford University, Rock Center for Corporate Governance.

Levine, T. R., Serota, K. B., Shulman, H., Clare, D. D., Park, H. S., Shaw, A. S., … Lee, J. H. (2011). Sender demeanor: Individual differences in sender believability have a powerful impact on deception detection judgments. *Human Communication Research*, *37*, 377–403.

Markowitz, D., Hancock, J. T., & Bazarova, N. (2011, November). *The language of presidential lies: How words can reflect lies about war, personal scandal, and state secrets*. Presented at the 97th Annual Meeting of the National Communication Association, New Orleans, LA.

McCornack, S. A. (1997). The generation of deceptive messages. Laying the groundwork for a viable theory of interpersonal deception. In J. O. Green (Ed.), *Message production: Advances in communication theory* (pp. 91–126). Mahwah, NJ: Lawrence Erlbaum Associates.

McCornack, S., Morrison, K., Paik, J. E., Wisner, A. M., & Zhu, X. (2014). Information manipulation theory 2: A propositional theory of deceptive discourse production. *Journal of Language and Social Psychology*, *33*(4), 348–377.

Mehrabian, A., & Wiener, M. (1966). Nonimmediacy between communicator and object of communication in a verbal message: Application to the inference of attitudes. *Journal of Consulting Psychology*, *30*, 420–425.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*, 665–675.

Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us*. New York, NY: Bloomsbury Press.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic inquiry and word count: LIWC 2007*. Austin, TX: LIWC.

Porter, S., & Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior*, *20*, 443–458.

Rayson, R. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, *13*, 519–549.

Sanbonmatsu, D. M., & Fazio, R. H. (1990). The role of attitudes in memory-based decision making. *Journal of Personality and Social Psychology*, *59*, 614–622.

Schober, M. F., & Glick, P. J. (2011). Self-deceptive speech: A psycholinguistic view. In C. Piers (Ed.), *Personality and psychopathology* (pp. 183–200). New York, NY: Springer.

Smith, J. R., & Terry, D. J. (2003). Attitude-behaviour consistency: The role of group norms, attitude accessibility, and mode of behavioural decision-making. *European Journal of Social Psychology*, *33*, 591–608.

SoLET Lab. (2012). *SiNLP: Simple NLP*. Retrieved from http://129.219.222.66/Publish/sinlp.html

Toma, C., & Hancock, J. T. (2012). What lies beneath: The linguistic traces of deception in online dating profiles. *Journal of Communication*, *62*, 78–97.

Van Swol, L. M., & Braun, M. T. (in press-a). Communicating deception: Differences in language use, justification, and questions for lies, omissions, and truths. *Group Decision and Negotiation.*

Van Swol, L. M., & Braun, M. T. (in press-b). Channel choice, justification of deception, and detection. *Journal of Communication*.

Van Swol, L. M., Braun, M. T., & Kolb, M. R. (in press). Deception, detection, demeanor and truth bias in face-to-face and computer-mediated communication. *Communication Research.*

Van Swol, L. M., Braun, M. T., & Malhotra, D. (2012). Evidence for the Pinocchio effect: Linguistic differences between lies, deception by omission, and truth. *Discourse Processes*, *49*, 79–106.

Van Swol, L. M., Malhotra, D., & Braun, M. T. (2012). Deception and its detection: Effects of monetary incentives and personal relationship history. *Communication Research*, *39*, 217–238.

Vorauer, J. D., & Ross, M. (1999). Self-awareness and feeling transparent: Failing to suppress one's self. *Journal of Experimental Social Psychology*, *35*, 415–440.

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. Chichester, UK: John Wiley.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Hoboken, NJ: John Wiley & Sons.

Vrij, A., Semin, G. R., & Bull, R. (1996). Insight into behavior displayed during deception. *Human Communication Research*, *22*, 544–562.

Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research*, *19*, 52–90.

Zhou, L., Burgoon, J. K., Nunamaker, J. F., & Twitchell, D. (2004). Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications. *Group Decision and Negotiation*, *13*, 81–106.