7-27-2011

Uncovering the Hidden Cognitive Processes and Underlying Dynamics of Deception

Nicholas David Duran

To the University Council:

The Dissertation Committee for Nicholas D. Duran certifies that this is the final approved version of the following electronic dissertation: "Uncovering the Hidden Cognitive Processes and Underlying Dynamics of Deception."

Danielle S. McNamara, Ph.D.
Major Professor

We have read this dissertation and recommend its acceptance:

Rick Dale, Ph.D.

Arthur C. Graesser, Ph.D.

Eugene H. Buder, Ph.D.

Accepted for the Graduate Council:

Karen D. Weddle-West, Ph.D.
Vice Provost for Graduate Programs

UNCOVERING THE HIDDEN COGNITIVE PROCESSES AND UNDERLYING DYNAMICS OF DECEPTION

by

Nicholas D. Duran

A Dissertation

Submitted in Partial Fulfillment of the

Requirement for the Degree of

Doctor of Philosophy

Major: Psychology

The University of Memphis

August 2011

Acknowledgement

The trajectory of my academic career has been masterfully guided by several amazing people, for whom I am truly grateful. First and foremost, I am indebted to my two co-advisors, Danielle McNamara and Rick Dale. Their patience, insight, support, and friendship have proven to be invaluable over the last several years. The work presented in this dissertation would not have been possible without them.

I would also like to thank the Institute for Intelligent Systems (IIS) and the many colleagues who have challenged my thinking and pushed me in new directions. Special recognition goes to Phil McCarthy who, from the earliest stages of my training, has been both a mentor and a confidant.

I would also like to recognize the National Science Foundation for supporting me with a Graduate Student Fellowship. This Fellowship was a huge boon to my academic development, and allowed me to carve out a unique niche in the research landscape.

And finally, my utmost gratitude goes to my family for their enduring love, particularly to my wonderful wife who is the best thing that has ever happened to me.

Abstract

Duran, Nicholas D. PhD. The University of Memphis. August/2011. Uncovering the Hidden Cognitive Processes and Underlying Dynamics of Deception. Danielle McNamara.

This dissertation examines the processing demands associated with motor responding and verbal statements during deceptive (or deceptive-like) behavior. In the first set of studies presented in Chapter 2, participants' motor movements in a false response paradigm revealed signatures of competition with the truth. In a second set of studies presented in Chapter 3, deceptive participants used language that reflected cognitive and social demands inherent to various types of deception. In evaluating both motor and verbal cues, this dissertation provides a comprehensive, multi-modal approach to better understanding the cognitive processes underlying deception.

In conducting the motor responding studies, participants' arm movements were analyzed as they navigated a motion tracking device (computer-mouse, Nintendo Wiimote) to visually co-present response options, where the "true" option acts as a competitor to a false target. In an initial study, competition during deceptive responding was shown to be much greater than during truthful responding. In two follow-up studies, the introduction of various task-based cognitive demands was shown to systematically modulate response performance. Specifically, these studies suggest that an intention to false respond early in question presentation will amplify competition effects, and that false responding to information in autobiographical memory is much more difficult than responding to information in general semantic memory.

In the studies analyzing verbal statements, the focus is turned to large-scale linguistic analyses using automated natural language processing tools. In the first study, changes in language use were identified between deceptive and truthful narratives using six psychologically relevant categories. A major finding was that the language of deception is adapted to facilitate ease of cognitive processing.

In a second study, the indicative phrasing and semantic content of deceptive texts was extracted using a *contrastive corpus analysis*, whereby indicative features are defined by their frequent use in one corpus while being infrequent in a comparative corpus. Two contexts of deception were evaluated. In the first context of *computer-mediated conversations*, deceivers used a range of unique thematic elements, as in avoiding personal involvement in their narrative accounts. In the second context of *attitudes towards abortion*, unique thematic elements once again emerged; for example, participants tended to position their arguments in terms of formal law.

Preface

This dissertation is comprised of a collection of published journal articles of which I am the primary author, as well as unpublished data that supplements the published work. This original research was conducted while I was a research assistant in two laboratories; one led by Dr. Danielle McNamara and the other by Dr. Rick Dale. This research was also partly funded by a National Science Foundation Graduate Student Fellowship.

There are two major sections of this dissertation, the first, *Cognitive Dynamics and Deception*, is built upon an article published in the peer-reviewed journal *Psychonomic Bulletin & Review*:

> Duran, N.D., Dale, R., & McNamara, D. S. (2010). The dynamics of overcoming the truth. *Psychonomic Bulletin & Review, 17*, 486-491.

This first section also includes three additional unpublished studies that extend Duran, Dale, et al. (2010).

The second section, *Discourse Analysis and Deception*, is built upon an article published in the peer-reviewed journal *Applied Psycholinguistics*:

> 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.

An additional unpublished study is also included that extends Duran, Hall, et al. (2010). This additional study was written with the intention of being submitted for publication, following the guidelines of the journal *Discourse Processes*.

All studies in this dissertation are preceded by a short note indicating whether the reported work has been previously published, and if so, a citation is provided. No major modifications have been made to the original articles (e.g., original formatting preserved). However, changes to figure and table number have been made to maintain consistency across chapters.

It should also be noted that each section includes a self-contained literature review and justification. However, in the primary *Introduction*, I also provide additional review material that supplements the information in each section. I briefly canvass previous research regarding the general cognitive underpinnings of deception. While many of the studies are germane to the current study, the overarching goal is to provide a general theoretical landscape in which to situate the current work. This introductory material is also weighted toward the Chapter 2: *Cognitive Dynamics and Deception*, where the literature review from the original published article was somewhat limited.

Table of Contents

CHAPTER PAGE

List of Tables

TABLE PAGE

List of Figures

Uncovering the Hidden Cognitive Processes

and Underlying Dynamics of Deception

Chapter 1: Introduction

Deceptive behavior can take on many forms and serve a variety of goals. Some lies are morally reprehensible, but are nevertheless effective in advancing personal gain (Bok, 1999). In other cases, an expected "white lie" can maintain social pleasantries, executed for the benefit of a group (DePaulo et al., 2003). Although the expressions are many, what all types of deception have in common is that the person has a willingness to violate what is considered to be reality, and to ostensibly maintain a falsity as truth (Ekman, 1997). To achieve such behavior, sophisticated cognitive processes are likely involved, such as the inhibition of a truth bias (Gilbert, 1991; Muraven & Baumeister, 2000), executive control in delimiting truth from lies (Spence et al., 2004), the generation of imagined events (Schacter, Normam, & Koutstaal, 1998), and the monitoring of another's mental state (Burgoon, Buller, Guerrero, Afifi, & Feldman, 1996; Frith & Frith, 2003). It should then come as no surprise to learn that deception is neurally instantiated in highly developed cortical regions (e.g., dorso- and ventro-medial prefrontal regions, anterior cingulate cortices; Sip, Roepstorff, McGregor, & Frith, 2007), as it is with other species that are capable of rudimentary types of deception (Premack, 2007). But to be clear, it is only with humans that a remarkable degree of flexibility and variation is revealed. Human deception is special, largely due to the richness of the cognitive processes involved.

The aim of this dissertation is to better understand the cognitive processes underlying deception. I do so within the context of multimodal channels of behavior, homing in on two major areas: continuous perceptomotor dynamics and low-level analysis of language. The emphasis is a departure from the predominant "applied" approach that deals primarily with the accuracy of detecting deception. Instead, the developments discussed in this dissertation are organized around two basic research questions. The first, *Research Question 1*, is summarized as: "What do the motor dynamics generated during deceptive behavior reveal about the underlying cognitive processes and strategies of deception?"; the second, *Research Question 2*, is summarized as: "What does the linguistic output of deceptive speakers reveal about the underlying cognitive processes and strategies of deception?" Both questions are explored in greater detail across several empirical studies. Before delving in, a short synopsis and supplementary justification of each study is first provided (see Preface for organizational rationale).

Research Question 1: Motor

The rationale for Research Question 1 is motivated by recent work on the dynamics of cognition as they relate to the dynamics of perception and action. I present evidence that complex cognitive states, such as those needed to perpetrate deception, will have particular signatures in the movements of the body. At the core of this theoretical perspective is the idea that much of the cognitive system relies heavily on perceptual and motor mechanisms. Rather than each mechanism "communicating" separately, where the output of one serves as the input of another, the components instead interact simultaneously (Spivey, 2006; Spivey, Richardson, & Dale, 2008). The important implication here is that continuous action can serve as the direct manifestation of real-time cognitive processing. In the studies that will be presented, I take advantage of one particularly relevant class of actions in the form of reaching movements during decision tasks.

This approach deviates from previous attempts that have exclusively employed simple reaction time measures (Gregg, 2006; Sartori, Agosta, Zogmaister, Ferrara, & Castiello, 2008; Walczyk et al., 2005; Walczyk, Mahoney, Doverspike & Griffith-Ross, 2009; Vendemia, 2005). Such measures capture only the duration of processing, rather than the moment-by-moment competition that might occur during false responding. To capture this competition, I follow the lead of action dynamics researchers that have used motion tracking devices like the Nintendo Wiimote controller and a simple computer-mouse (Dale, Roche, Snyder, & McCall, 2008; Spivey & Dale, 2006). These devices are useful in assessing how the hand goes about selecting critical "targets" amidst visually co-present "distractors." By surreptitiously measuring the position of the hand during the course of movement (via the position of a cursor on a computer screen), periods of instability, stabilization, and modulation can be identified. Such a fine-grained examination will shed new light on how the truth interferes with an intent to express the counterfactual.

Apart from some work in neuroscience, there have been very few attempts to characterize the mechanisms involved in deception. The theories that do exist are typically aligned with traditional information processing accounts (Vendemia, Buzan, & Green, 2005; Walczyk et al., 2005; Walczyk, Roper, Seemann, & Humphrey, 2003). These accounts involve discrete stages of processing that are largely independent of contextual constraints (Evans, 2008; Fodor, 1983; Fodor & Pylychyn, 1981, Miller, 1956). In contrast, for a dynamical account, response behavior is closely tied to emergent order, constrained and guided by task conditions, which continuously evolve over overlapping and varied time scales (Port & van Gelder, 1995; Spivey & Dale, 2006; Van Orden, Holden, & Turvey, 2003). Such an approach allows for a flexible depiction of cognitive processes that are themselves characterized by flexibility and change.

In the first of four studies (Study 1a), the x,y coordinates of a handheld Nintendo Wiimote are tracked while participants make true and false responses to yes-or-no autobiographical questions. Unlike a computer-mouse, the Wiimote has a notable advantage of increasing the range of arm motion during response performance. Rather than being confined to a computer desk (a stabilizing surface), the Wiimote is held with an outstretched arm toward the direction of presented stimuli.

In this first study, a question stimulus was presented at the bottom of a large screen and participants moved the Wiimote cursor from the bottom of the screen to "YES" and "NO" labels displayed at the top of the screen. Critically, half of the questions were preceded by a prompt that required a deceptive response. This prompt appeared with the last word of the question. The trajectories were assessed across multiple timescales and for a number of trajectory properties, including overall response time from start to finish, latency time to initiate a movement, and trajectory curvature toward visual attractor regions (e.g., target and competitor regions) in short slices of ongoing time (slices around 20 ms). Trajectory properties like instability and attractor strength were also measured. Together, these variables provide a vastly more detailed depiction of the fine-grained changes of cognitive dynamics (via the behavioral dynamics) of a decision process.

The second study (Study 1b) replicates the Wiimote study, but data are now captured via mouse movements. Furthermore, the study is implemented in an Adobe Flash environment that can be played as an online game by participants across the United States. By doing so, a diverse range of demographics can be represented that are typically not found in the average university laboratory.[1] Another advantage of the current approach is that it is easily parameterizable and provides a baseline for comparing the following two studies (Study 2a and 2b). The study is also at the forefront of a growing trend in cognitive science of using “crowdsourcing” technologies, like Amazon’s Mechanical Turk, to collect human data (see Munro et al., 2010).

In the third and fourth studies (Study 2a and 2b), task-based cognitive demands of deception are examined. Deception is a very present force in virtually all social interactions. Whether it is to enhance one’s self-image (Feldman, Forrest, & Happ, 2002) give another a false impression of one’s feelings (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996), or even fabricate one’s whereabouts, plans, or actions (DePaulo et al., 2003), lying occurs at a surprisingly high rate, and is considered by some to be an evolutionary-driven norm in human interactions (DePaulo et al., 1996; DePaulo et al., 2003; DePaulo & Kashy, 1998; Lippard, 1988). However, the prevalence of deception stands in contrast to the cognitive difficulty associated with lying, which should limit the attempts at deception given the increased odds of mishandling the execution of a lie. Why then are people so willing to engage in deception? One reason is that deception exists on a continuum of difficulty, where certain contexts make lying more or less difficult, and therefore more or less risky. Indeed, this has been exploited by researchers interested in

[1] A laboratory-based mouse movement analysis was also conducted and is presented in Appendix D.

devising interview techniques to catch deceivers. By ramping up the difficulty of deception, as in asking questions that are unexpected or by asking a suspected liar to repeat a fabricated story, the more pronounced the tell-tale indicators of deception will become (Vrij, Granhag, Mann, & Leal, 2011).

In studies that exploit the cognitive difficulty of lying, it is often unclear what the cognitive factors are that reside behind better "deception-eliciting" questions. Although a general notion of executive function is given, there is little in the way of an explanation for how this system interacts with parameters inherent to the questions asked, such as the memory demands associated with question content, or the time given to prepare an answer. Although subtle, these changes are likely to modulate the difficulty of lying, and are the focus of the current action dynamics approach.

Beginning with the specifics of the third study (Study 2a), I explore signatures of deception when the false response prompt occurs at the question initial position, that is, with the first word of a question. This setup primes the participant with a true or false response orientation before a response is actually required. The additional time allows participants to anticipate how they will respond (either truthfully or falsely) without knowing what form the response will take (either with an affirmative "yes" or a denial "no"). This setup simulates a situation where someone with concealed knowledge (as in knowledge of the details surrounding a store robbery) is regarded as a suspect, and has a heightened awareness of the possibility that he or she might have to lie. In other words, this suspect is oriented to a state of readiness for false responding. But for this scenario to conform to the conditions of the current study, the suspect must be asked a series of questions whose required response is difficult to predict (again, whether the question requires a “yes” or “no”). Given this set-up, there are two predictions for response behavior outcomes. One prediction, based on Walczyk et al. (2003; 2005), is that lying occurs in discrete steps, whereupon being asked a question, one must activate the truth, make a decision to falsify, and then generate a false answer. Given that a participant in the current study is told ahead of time to falsify information, the second step of “decision to falsify” is effectively bypassed, thereby minimizing the cognitive resources expended. As a result, responses should be faster and more stable, at least when compared to Study 1b, where the prompt occurs at the question final position (i.e., with the last word of the question). However, another prediction, more aligned with a dynamical systems account, is that a false response might instead be more difficult to execute. The priming of a false response deepens a false response attractor, eliciting a tendency to “deny” information. This stronger tendency to say “no” (i.e., deny) builds up during the time it takes to present the words in each question, and makes certain types of responses more difficult, like those requiring one to say “yes” falsely. Moreover, a false prime might also activate the “true” condition of false information, whereby the truth becomes a stronger distractor when attempting to fabricate information. This is similar to the effect of negation, where the non-negated (“true”) condition of a negated sentence interferes with interpretation, even when one has foreknowledge that to-be-read sentence requires negation (Mayo, Schul, & Burnstein, 2004). This latter account thus makes predictions that overall false behavior will take longer and be more unstable, but that false response behavior is also modulated by whether the eventual response takes on the form of “yes” or “no.” These predictions are not easily accommodated by “discrete-based” accounts.

Finally, for the fourth study (Study 2b), I examine the response dynamics involved in falsely responding to information that is autobiographical in nature and information that is based on general semantic knowledge. These question types draw on different memory sources that are likely to modulate the difficulty associated with deception. Given that recall fluency depends on how information is first encoded, as well as the number and strength of associations within a memory network (Koriat, Goldsmith, Pansky, 2000; Nunez, Casey, Egner, Hare, & Hirsch, 2005, Tulving, 2002), autobiographical information is expected to generally be more accessible, as it is linked to a greater number of salient experiences, emotions, intentions, and motivations (Harmann, 2001). But if the truth content of autobiographical information "comes to mind" with greater ease, it will be that much more difficult to falsify. Indeed, there is some evidence that patterns of inhibitory control in neural firings are more pronounced when lying about personal events compared to non-personal events (Piefke, Weiss, Zilles, Markowitsch, & Fink, 2003). Thus, the hypothesis presented here is that falsifying personal, autobiographical facts will be slower and exhibit greater instability than the falsification of semantic facts.

Research Question 2: Linguistic

Turning now to Research Question 2, I move from action dynamics to an entirely new domain of discourse analysis. The goal is to use theoretical advances in computational technology to forge novel links between language and thought. This focus allows for large-scale exploration of spontaneously expressed thought in a variety of pragmatic contexts. To proceed, a comprehensive set of linguistic features, ranging from word information variables (Hancock, Curry, Goorha, & Woodworth, 2008; Newman, Pennebaker, Berry, & Richards, 2003) to sentence and discourse level variables (Zhou, Burgoon, Nunamaker, & Twitchell, 2004) are used across two major studies. Subsets of conceptually related features will be used to uniquely characterize the mindset, motivations, and limitations of a person engaged in deceptive behavior.

In the first of two studies, a natural language processing (NLP) tool Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004) is used to evaluate deceptive and truthful conversations that occur within a context of computer-mediated communication. This study builds on the work of Hancock et al. (2008) who evaluated this conversational corpus with an NLP tool called Linguistic Inquiry and Word Count (LIWC, Pennebaker, Francis, & Booth, 2001). Both analyses take advantage of the expectation that, despite a deceiver's best attempt to avoid exposing the truth, the cognitive and social demands brought on by deception will be closely bound to subtle changes in language use (Pennebaker, Mehl, & Niederhoffer, 2003). Indeed, such changes have been observed in previous research, and have been used to develop training programs for improving people's abilities to detect deception. One of the most well-known programs, *Criteria-based Content Analysis*, involves 19 criteria based on the notion that fabricated reality will be qualitatively distinct from actual reality. Human judges are asked to evaluate verbal content for features like breakdowns in logical structure, quantity of details, and spontaneous corrections (Steller & Kohnken, 1989; Vrij, 2005). In another approach known as *Reality Monitoring*, that was originally developed to experimentally explore phenomenological experiences, imagined reality is shown to lack the sensory richness that is evident in actual memories (Johnson, 1988; Johnson & Raye, 1981). Deceptive narratives have also been found to lack spatial and temporal details that are typically found in the truth (Masip, Sporer, Garrido, & Herrero, 2005). However, a major disadvantage in the *Criteria-based Content Analysis* and *Reality Monitoring* approaches is that the verbal features under analysis must be identified by human raters. Not only is this a time-consuming process, but it leaves open rater biases that are likely to generate a greater number of false alarms or misses (Burgoon, Blair, & Strom, 2008). And verbal cues that are obvious enough to be processed by human judges, might also be easier to manipulate by deceivers.

In the current NLP approach, many of the concerns with previous verbal analysis are largely reduced with the use of sophisticated computational algorithms. These algorithms allow for automatic evaluations of low-level linguistic features that would be nearly impossible for human judges to calculate. Furthermore, these features lose none of their psychological plausibility, as they are grounded in theories of memory, pragmatics, and social psychology (including many of the categories of the aforementioned *Criteria-based Content Analysis* and *Reality Monitoring* approaches). In total, there are six major categories of indices that correspond to the following core areas: a) the amount of information that is offered, b) the degree of personal identification with message content, c) the amount of sensory detail of descriptions, d) the semantic accessibility of words chosen for message content, e) the difficulty of syntactic phrasing, and f) the novelty of message content.

In addition, because there are many linguistic features that conceptually overlap between Coh-Metrix and LIWC, convergent validity between NLP tools is also examined. And finally, because the corpus used here has transcripts for both sender (i.e., liar) and receiver, the coordination of linguistic features between sender and receiver during deceptive and truthful exchanges can be evaluated.

In the second and final study of this section, *Discourse Analysis and Deception*, an NLP tool called the Gramulator (Duran & McCarthy, in review; Min & McCarthy, 2010) is used to examine deception in an emotionally-salient context. The Gramulator is based on contrastive corpus analysis (Granger, 1998; Min & McCarthy, 2009), and is useful for comparing short phrases and sequences of text that are statistically more likely to occur in one corpus relative to another. The result is an inductive evaluation of language elements, where the evaluation is conducted by returning to the local context in which the features are used, and using this context to draw meaningful conclusions.

In summary, this dissertation is organized around two major research questions concerned with the motor and linguistic indicators of deception. This comprehensive examination of multiple behavior channels paves a way for a holistic approach to deception detection. Such approaches have been advocated as a crucial next step in improving detection techniques (Porter & ten Brinke, 2010; Vrij & Mann, 2004), and to date, few have taken up the challenge. One exception, however, is the work conducted by Judee Burgoon and her research group at the University of Arizona. These researchers have experienced a great deal of success pursuing a multi-channel exploration of deceptive behavior (e.g., Burgoon et al., 2005; Jensen, Meservy, Burgoon, & Nunamaker, 2009). The research presented here pursues many of the same goals, but is uniquely situated to do so within the theoretical and methodological perspective of cognitive psychology, incorporating domains of action dynamics, executive processes, memory, language representation, and pragmatics. Thus, there is continuity between the two major sections of this dissertation (i.e., *Cognitive Dynamics and Deception* and *Discourse Analysis and Deception*), despite the sections largely constituting independent areas of research.

Chapter 2: Cognitive Dynamics and Deception

NOTE: This following introduction supplements Study 1a (below) by providing additional justification and detail that was omitted from the original publication.

Action Dynamics

The cognitive activity that occurs during deception is largely understood by reaction time measures in "guided lie" or "intentional false responding" paradigms (Spence et al., 2001; Vendemia, Buzan, & Green, 2005; Walcyzk et al., 2003). A consistent finding is that responses incompatible with the truth are executed more slowly than compatible responses. The observed increase in reaction time is typically taken as evidence for an increased workload in executive function. While executive function is a rather broad term, it generally relates to brain activity involved in the control and coordination of some task performance (Baddeley, 1996; Botvinick, Braver, Barch, Carter, & Cohen, 2001), including the inhibition of distractors while attending to a primary task (Carter et al., 1998; Garavan, Ross, Murphy, Roche, & Stein, 2002), and allocating attentional resources across multiple tasks to permit rapid transitions between tasks (Braver, Reynolds, & Donaldson, 2003). In the same manner, deceptive behavior is hypothesized to require greater executive function in the inhibition of true responses while selecting and transmitting false responses in real time (Nunez et al., 2005).

Indeed, the role of executive function has been implicated as one of the major reasons why young children have difficulty in communicating false information (Carlson & Moses, 2001; Carlson, Moses, & Hix, 1998). For example, in temptation resistance paradigms, children are told not to peek under a cup that might contain a desirable object, such as a toy or piece of candy. However, for children who inevitably do (with rates as high as 90%), they often fail to falsely deny having peeked, or are unable to avoid naming the object that they had seen. But children who are better able to deny (usually children older than 4 to 5 years of age), are also those who perform better on tests that measure inhibitory control and working memory, presumably because of a superior ability in suppressing the truth and holding false information in memory (Evans, Xu, & Lee, 2011; Talwar & Lee, 2008). Furthermore, in domains of comparative psychology, the inability to demonstrate intentional, strategic lying in primates is also linked to difficulties with executive function. In one study, chimpanzees were given incentives to avoid pointing to the location of a hidden food source, for example, by introducing a human confederate who did not know about the food location, but once informed, "stole" and ate the food without sharing. Chimpanzees in this study showed an initial tendency to point to a misleading "deceptive" location, but could not maintain this action, and always redirected their pointing to the hidden location (Woodruff & Premack, 1979).[2] This behavior is taken as evidence for an inability to inhibit the truth and maintain false information in working memory.

In the adult studies of deception, it is reaction time that exposes the breakdown of executive function in deception. However, it is also a surprisingly coarse and indirect measure of cognitive difficulty. To provide greater detail, researchers have begun to employ neuroimaging techniques to supplement their understanding of the processes involved in deception (Langleben et al., 2002; Vendemia & Buzan, 2004). For example,

---

[2] It should be noted, that in both cases involving children and primates, there are also other sophisticated cognitive processes that underlie difficulty associated with lying, most notably issues with mind-reading and perspective-taking (Byrne, 2010; Carlson, Moses, Breton, 2002). Such processes are also likely to play a significant role in adult's ability to deceive, but have not yet been explored in great detail (see Carrion, Keenan, & Sebanz, 2010; Ybarra, Winkielman, Yeh, Burnstein, & Kavanagh, 2010 for initial work in this area).

event-related brain potentials (ERPs) and functional magnetic resonance imaging (fMRI) have been used in conjunction with reaction time to isolate the control processes involved in monitoring and choosing between false and true responses (Spence et al., 2004). A general finding is that the neural activity underlying false responding is similar to the neural activity underlying conflict monitoring and distractor inhibition in Stroop interference (Nuñez et al., 2005) and task-switching paradigms (Christ, Van Essen, Watson, Brubaker, & McDermott, 2008). These neuroimaging studies provide direct evidence that executive function is increasingly strained during false responding (Nunez et al., 2005). The studies also augment the speculations of many reaction time experiments: increased processing time involves the inhibition of a true response, the selection of a false response, and the transmission of the falsity (Vendemia, 2005; Walczyk et al., 2003; Walczyk et al., 2005).

Going beyond reaction time is an important step in clarifying the cognitive aspects of false response behavior. For example, by evaluating the brain-state activity that accompanies a false response, additional insights can be had as to why processing is more difficult. However, in drawing conclusions, researchers assume for methodological convenience that processes of inhibition, selection, and transmission proceed with discrete and minimally interactive ordered components. These general assumptions have a long history in cognitive psychology (Donders, 1868; Wundt, 1874), and form the basis of the subtractive method of measurement, an analytic technique most recently instantiated in Sternberg's (1969) *Additive Factors Method*. In neuroimaging studies of deception, neural activity recorded during a deceptive response (i.e., experimental condition) is subtracted from the neural activity during a control behavior. The "deceptive" brain activity is the linear difference between these two response behaviors. Accordingly, any variables affecting deception cannot interact (because of the linear assumptions) and are regulated to separate processing stages. By doing this, there is an assumption that stages proceed in sequential order, in which one component gives way to the next in a single chain of processing (Van Orden & Papp, 1997).

An alternative to the assumptions of discrete cognitive processing is the assumption of continuous cognitive processes. Accordingly, component processes do not occur in a discrete sequence, but can be activated in an overlapping manner and resolved in parallel. Indeed, there is mounting evidence that continuous cognitive processing is conveyed through the continuous movement of the body (Spivey, 2006; Spivey & Dale, 2006). In studies of deception, the continuity is often obscured by reaction time measures that limit motor execution to ballistic movements, thereby collapsing intermediary cognitive processes to a single discrete behavior (Gregg, 2006; Walczyk et al., 2003). Continuity is also obscured by neuroimaging studies that often hold *de facto* assumptions about discrete processing, as is the case for neuroimaging studies focused on deception (e.g., Johnson, Barnhardt, & Zhu, 2004). Through tracking continuous action dynamics, we loosen such methodological and theoretical restrictions. As cognitive activity flows into the body, the continuous movements of the arm act as a direct conduit of the cognitive activity.

Continuous mind-body covariance has been repeatedly demonstrated in decision tasks that require responses to be made in the visual co-presence of target and distractor choices. In an example related to the current study, participants evaluated the truth of simple propositions by navigating a computer-mouse to true or false response options in the uppermost corners of a computer screen (McKinstry, Dale, & Spivey, 2008). Propositions that had a high level of uncertainty, such as *Is murder sometimes justifiable?*, were answered with slower arm movements and with greater moment-to-moment fluctuations than propositions with a high level of certainty of being true (e.g., *Should you brush your teeth everyday?*). These unstable arm movements suggest that greater cognitive effort is involved in evaluating ambiguously true information. Moreover, the continuous movements of the arm reveal an immediate and persistent influence of a “positive confirmation” bias throughout the response (see also Gilbert, 1991). This confirmation bias was most salient while answering no to propositions that were clearly false, such as “Is the mother younger than the daughter?”. For these propositions, there was a statistically significant tendency for the arm to gravitate relatively slowly toward a “yes” response (positive confirmation) during a no response movement. This pattern of movements suggest that both target false responses and a competitor confirmation bias were simultaneously activated, with the activation from the target response eventually prevailing.

Despite the evidence of graded activation in the parallel processing of truth evaluations, it is still unclear if similar processing is involved while contradicting the truth in a prompted false responding paradigm. The greater difficulty associated with prompted false responding may alter processing from continuous activation to that of discrete and serial operations. Indeed, as discussed earlier, the discrete hypothesis plays a significant role in the theoretical and applied assumptions of deception detection. In the current study, I pursue an alternative hypothesis that posits false responding is carried out more continuously. In doing so, I intend to show that measurements based on continuous processing can reliably distinguish between true and false response behavior (and does so with even greater accuracy than a reaction time measurement; see Appendix D for this and other supplementary analyses).

Study 1a: Psychonomic Bulletin & Review

Action Dynamics During False Responding (Wiimote)

NOTE: The following study was originally published in Psychonomic Bulletin & Review; Duran, Dale, & McNamara, 2010.

Most people can easily confirm or deny an assertion that they once met Elvis Presley or that they collect clocks in their spare time. Nonetheless, evidence suggests that prior to a final confirmation or denial of such assertions, people temporarily and perhaps non-consciously believe the assertions to be true (Gilbert, 1991; Gilbert, Tafarodi, & Malone, 1993). The nature of belief appears to be biased towards initially accepting a proposition as true, even when that proposition is unequivocally false. It stands to reason that for someone to *dishonestly* confirm or disconfirm any proposition, as in agreeing to having met Elvis when that was not the case, the person will accept the assertion as true, then assess whether the initial acceptance is correct (which it is not), and then respond in order to violate the conclusion of the assessment (by falsely saying *yes*). In such deceptive behavior, competition exists between the initial belief, the assessment of the belief, and the intention to deceive. While the initial belief is involuntary, this acceptance of the truth is an impediment that requires active processing to overcome.

By many accounts of deception, the competition and resolution involved in false responding is a far more challenging and time-consuming process than confirming the truth (Vendemia & Buzan, 2004; Walczyk et al., 2005). Accordingly, researchers interested in deception detection have designed clever tasks that exploit this increased processing time (e.g., Gregg, 2006; Sartori et al., 2008). Participants often respond honestly or deceptively to simple statements, and the time taken to respond, both by vocal onset or a manual key press, is recorded and analyzed. In general, these response time latencies are useful in discriminating certain deceptive behaviors, and have thus risen to prominence as a standard-bearer of detecting deception. Unfortunately, response time only captures the outcome of a completed cognitive process, and the real-time cognitive dynamics that occur during the process are lost.

To begin exploring these moment-to-moment changes of response selection, we turn to a growing body of *action dynamics* research. In this research, actions that occur in conjunction with a cognitive task often reflect ongoing characteristics of processing, ranging from low-level speech perception (Spivey, Grosjean, & Knoblich, 2005) to higher-level learning (Dale et al., 2008). The response activity involved in these tasks is usually recorded as arm movements within a set spatial region. Analysis of the arm movement provides insight into what information is important during processing, and when that information is most relevant. For example, in Dale, Kehoe, and Spivey (2007) participants used a computer-mouse to make typicality judgments on category membership. Each trial involved matching an animal exemplar (e.g., *whale*) to one of two visually co-present response options located in opposite corners of a computer screen. For some trials, classification was potentially ambiguous, such as matching *whale* to a *fish* or *mammal* option. During response movements, the streaming *x,y*-coordinates of the computer-mouse were recorded. The researchers found that computer-mouse trajectories for atypical animals (e.g., *whale*) curved more towards a featurally similar distractor (e.g., *fish*), suggesting that semantic categorization process also unfolds partly into the dynamics of response execution.

We use this cognition-action interplay to tap the dynamics of false responding. Indeed, there is well-established evidence that deception often “leaks” into a deceiver’s actions, such as facial movements and body posture (Vrij, 2001). Here, we employ an action dynamics technique to study response behavior as continuous competition from an initial belief and the goals of deception. To the extent that this competition is expressed in action, we can use *x,y*-coordinate trajectories to expose the dynamics of overcoming this initial true belief, and enacting the agenda of a false response.

In the current study, we expose hidden cognitive activity that is involved while falsely accepting or denying assertions about oneself. To do so, we use a modified version of the “guided lie” paradigm that is commonly employed in EEG and fMRI analysis (e.g., Spence et al., 2004; Vendemia & Buzan, 2005). Participants are prompted to respond falsely or truthfully to simple autobiographical facts, such as *Have you ever been to Asia?*. Rather than answering with a computer key press, participants use a Nintendo Wiimote and the *x,y*-coordinates of their arm movements are rapidly sampled (see Dale et al., 2008).

With this rich data output, we evaluate signatures of deception in terms of the shape of each movement trajectory and the location of the trajectory over time. We also quantify trajectory properties on dimensions of velocity, stability, and direction. As the results reveal, the “unpacking” of response time not only provides unique distinctions between false and true responses, but also in the more subtle distinction between false responses answered either with a “no” or “yes”.

Experiment: Revealing the Dynamics of False and True Responding

Participants

Twenty-six undergraduate students (19 female, 7 male) participated for extra credit. Only native English speakers with normal-to-corrected vision were eligible to participate. All participants were right-hand dominant.

Procedure

A trial began with a small bull's-eye-shaped circle appearing at the bottom-center region of a 3.8m x 1.8m screen positioned approximately 2.7m directly in front of the participant. The participants' task was to click on the circle with the Wiimote-controlled cursor, and by doing so, the first word in a biographical question would appear above the circle. With each click of the Wiimote, the current word was replaced with the next word in the question. This process continued until the final word of the question was encountered (akin to self-paced reading tasks; Just, Carpenter, & Woolley, 1982). At this point, a "NO" response box appeared in the top-right corner of the screen and a "YES" response box appeared in the top-left corner of the screen (each box was approximately 0.5m x 0.5m). Also at this time, the bull's-eye-shaped circle changed to the color green or to the color red. If the circle changed to the color green, the participant was simply instructed to answer the question truthfully. If the circle changed to the color red, the participant was to answer falsely. All responses were made by navigating the Wiimote cursor to the appropriate "NO" or "YES" box (see Appendix B for an example layout of the experimental design).

During each trial, the *x,y* coordinates of the cursor movement were continuously recorded (sampling at approximately 80Hz) and stored for later processing. Because the

Wiimote is held with an extended arm in front of the body and toward the screen, subtle directional changes in wrist and arm movement are captured, with 33 pixels traversed for each centimeter of lateral movement.

The participants responded to 84 question trials, including 4 practice trials. The 80 experimental trials were divided equally between false and true color prompts, and the order of the false and true prompts, as well as the order of the questions, was randomized for each participant. The position of the “NO” and “YES” response boxes was also reversed for a third of the participants (i.e., “NO” response box appeared in the top-left corner of the screen and the “YES” response box appeared in the top-right corner of the screen). After the initial set of questions had been answered, the participants completed a follow-up verification task that required participants to truthfully re-answer all the questions they had previously viewed.

Question Stimuli

The questions used in the study began with the stem *Have you ever…* and were completed with 120 possible statements (i.e., *Have you ever eaten pizza?*; stimuli available online[3]). Eighty of the 120 questions were randomly selected for each participant. The questions were also selected to elicit an equal number of false “no” responses, true “no” responses, false “yes” responses, and false “no” responses. A pilot study confirmed that the responses were approximately evenly distributed within individuals across our target population. All questions were completed with two to three words, following the pattern of *verb* + *object* or *verb* + *preposition* + *object.* The *object* of the question always occurred in the sentence-final position to prevent early guessing of sentence meaning.

[3] http://actiondeception.nickduran.com/

Results

Trials were excluded if responses in the verification task were incongruent with the original responses or if response time was above 3 standard deviations. This exclusion criterion eliminated one participant (incongruent responses exceeded 50% of total trials) and 92 additional trials (5% of the data). Of the 1908 trials remaining, 435 trials occurred in the "false/no" condition, 455 trials occurred in the "truth/no" condition, 503 trials occurred in the "false/yes" condition, and 515 trials occurred in the "truth/yes" condition.

Trajectory Shape. This analysis examines the shape of each trajectory as it moves from the bottom-center bull's-eye to the final response box at the top left or top right of the screen. To conduct this analysis, the response trajectories for each participant were initialized to x,y-coordinates (0, 0) and interpolated to 101 time steps (see Dale et al., 2007; Spivey et al., 2005). At each time step, the x- and y-coordinate positions were then averaged within condition for each participant. To compare conditions, we performed paired t-tests at corresponding x-coordinate time steps (a total of 101 t-tests). A consecutive run of statistically significant tests indicates that trajectories between conditions are diverging during response execution.

The false "no" and true "no" trajectories diverged for 29 time steps ($p < .05$) between the 59th and 88th steps, while the false "yes" and true "yes" trajectories statistically diverged for 58 time steps ($p < .05$) between the 40th and 98th steps (Figure 1a). The divergence for each comparison exceeds the minimum number of 8 consecutive time steps that bootstrap simulations have shown to be a standard for statistical significance (see Dale et al., 2007). Accordingly, the *trajectory shape* analysis reveals false and true responses that are conspicuously different. True response movements appear to travel a more direct route to the target response, whereas false response movements take a more curved route. The bend of the curve is always in the direction of the competing response option (i.e., the “true” response). This greater curvature suggests that competition is greater for false responses, whereas processing for true responses is relatively unaffected.

To further explore this competition, we examined divergence between falsely responding “no” compared to falsely responding “yes” by superimposing mirror-reversed false “yes” trajectories (see Figure 1c). A paired t-test analysis revealed that false “yes” responses diverged from false “no” responses for 16 time steps ($p < .05$) between the 55th and 70th time steps. During these time steps, false “yes” responses were closer to the competing true response option. Not only was there a greater competition for general false responding, but this effect was most pronounced with false “yes” responses.

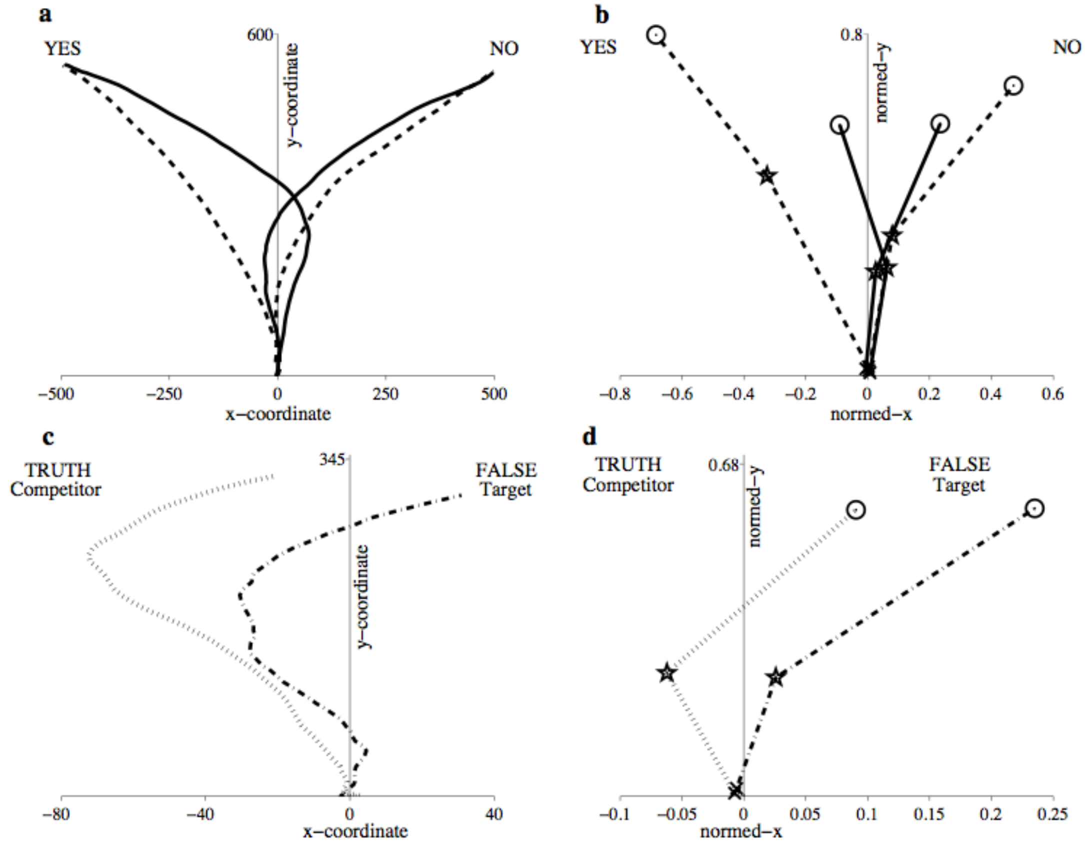

Figure 1. (*a*) Shape of Wiimote trajectories for each condition. False answers (solid lines) display a greater arc towards the competing response option than true answers (dashed lines). (*b*) Location of Wiimote trajectories for each condition. The *x*-coordinate position is plotted at 500 ms (cross), 1000 ms (star), and 1500 ms (circle). The false answer positions (connected by solid line) show slower movements toward the correct response location (e.g., upper-left corner for false "yes") and are closer to the competing response option (e.g., upper-right corner for false "yes") than true answers (connected by dashed line). (*c*) Shape of trajectories for false "yes" responses (mirror-reversed from that of Figure 1a) compared to shape of false "no" responses. The false "yes" responses (dotted line) are closer to the competing true option than false "no" responses (dash-dotted line). (*d*) Location of trajectories for false "yes" responses (mirror-reversed from that of Figure

1b) compared to shape of false "no" responses. The false "yes" responses (dotted line) are closer to the competing true option than false "no" responses (dash-dotted line).

Trajectory Location. This analysis compares the location of response trajectories after the first, second, and third 500 ms of processing (for 500, 1000, and 1500 ms). This information was lost in the previous analysis when temporal information was collapsed into fixed time steps. Now with *trajectory location*, we can answer the question of *when* trajectories begin to be statistically divergent. To conduct this analysis, the trajectory coordinates were first normalized to initiate at x,y-coordinate (0, 0) and end at (1, 1). Next, the normalized x-coordinate positions for false and true responses were captured at the 500, 1000, and 1500 ms processing mark and placed in corresponding "time bins". The average location of each time bin for each condition is plotted in Figure 1b.

A 2 (prompt type: true vs. false) x 3 (time bin: 500 vs. 1000 vs. 1500 ms) repeated-measures ANOVA was used to evaluate trajectory position in real-time. Beginning with the "no" trials, there was a statistically significant effect for prompt type, time bin, and the interaction between prompt type and time bin. To explore this interaction further, planned comparisons were conducted between prompt types at each time bin. There was a statistically significant difference of the x-coordinate position at the third time bin (1500 ms) between false (M = .23, SD = .27) and true (M = .47, SD = .23) trajectories, $F(1, 24) = 20.54$, $p < .001$.

The repeated-measures results for "yes" trials also showed a statistically significant effect for prompt type, time bin, and the interaction. Planned comparisons for the interaction revealed statistically significant differences at the x-coordinate position at the second time bin (1000 ms) between false (M = .06, SD = .10) and true (M = -.31, SD

= .20) trajectories, $F(1, 24) = 81.40$, $p < .001$, and at the third time bin (1500 ms) between false ($M = -.09$, $SD = .16$) and true ($M = -.68$, $SD = .19$) trajectories, $F(1, 24) = 203.29$, $p < .001$.

Taken together, these results indicate that false ”yes” trajectories diverge from true “yes” trajectories much earlier (at around 1000 ms) than when false “no” trajectories begin to diverge from true “no” trajectories (not until at least 1500 ms). Because divergence here refers to movement toward the competing response option, false ”yes” trajectories appear to be influenced by a truth competition much earlier than do false “no” trajectories.

As in the *trajectory shape* analysis, false “yes” and “no” trajectories can be directly compared to each other instead of using the true trajectories as a reference, which have their own idiosyncratic response biases (as in a “yes-bias” for “no” responses). As such, the trajectories for false “yes” responses were mirror-reversed (see Figure 1d) and paired t-tests at each time bin were assessed. Once again there was greater divergence for false “yes” responses compared to false “no” responses, with divergence recorded at the second time bin $F(1, 24) = 4.31$, $p = .05$ and third time bin $F(1, 24) = 7.51$, $p = .01$.

Trajectory Velocity. The velocity of response trajectories was evaluated by computing the distance (in terms of pixels) covered per second within a moving window of 8-x,y pixel coordinates across *total time*. Figure 2 shows the average velocity profile for each condition. This figure suggests that, on average, the initial increase in velocity (as participants committed to a response) and the subsequent decrease in velocity (as participants completed the response) occurred much later for false responses than for true responses. A repeated-measures ANOVA conducted on the moment of peak velocity confirms this observation, such that a significant interaction, $F(1, 24) = 25.51$, $p < .001$, and follow-up planned comparisons between prompt and response type reveal that false "no" responses peaked later in time than true "no" responses, $F(1, 24) = 22.29$, $p < .001$, and that false "yes" responses peaked later in time than true "yes" responses, $F(1, 24) = 93.32$, $p < .001$.

We also examined differences in the magnitude of peak velocity with the assumption that greater response activation would result in higher peaks. A statistically significant interaction was found between prompt and response type, $F(1, 24) = 16.87$, $p < .001$, showing that the peak for false "yes" responses is lower in magnitude than true "yes" responses, $F(1, 24) = 9.07$, $p < .01$. There were no significant differences for false "no" and true "no" responses.

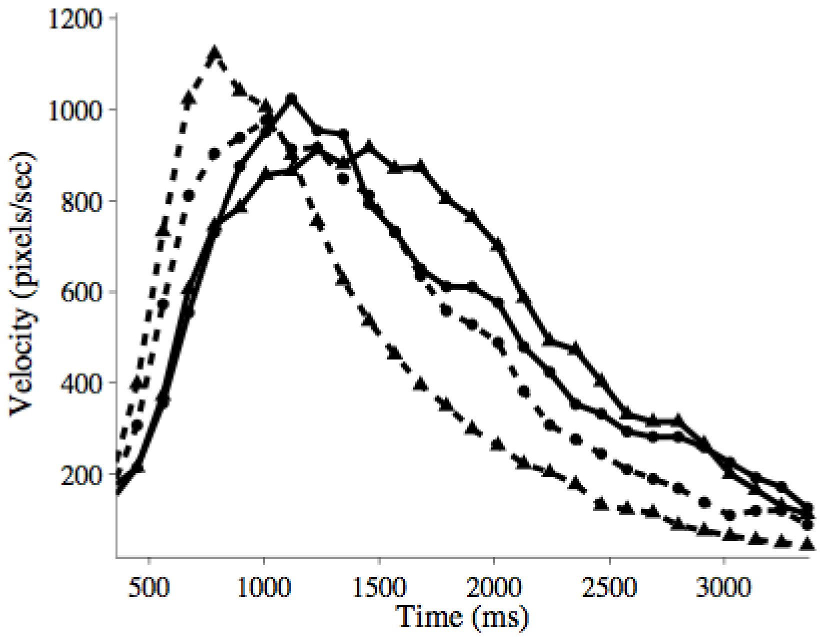

Figure 2. The velocity profiles plotted as a function of total time for true “no”, (dashed line with circles), true “yes” (dashed line with triangles), false “no” (solid line with circles), and false “yes” (solid line with triangles) responses. The plotted range of 376 to 3400 ms covers 90% of all completed trajectories and eliminates extremely early and late movements that have near-zero velocities. (NB: Average velocity profiles are defined over a broad range around the mean total time because these profiles are partly based on slower trajectories that go beyond the mean total time, that is, trajectories with a total time that is slower than the mean total time.)

Trajectory Properties. In this final analysis we computed eight properties that characterize temporal and trajectory behavior along continuous scales of measurement (Dale et al., 2007). The variables are listed and summarized below:

(a) *Total Time*: the amount of time elapsed between the initiation of the prompt and making a “yes” or “no” response;

(b) *Latency*: the amount of time the mouse cursor stays in a "latency region", with region defined as a 100-pixel radius that surrounds the mouse cursor position that initiated the response prompt (Dale et al., 2008);

(c) *Distance*: the Euclidean distance traveled by the trajectory after leaving the latency region and making a "yes" or "no" response;

(d) *Motion Time*: the amount of time elapsed while moving between the latency region and completing a "yes" or "no" response;

(e) *High x-value*: a measure of how close (in coordinate position) each trajectory curves toward the "no" response box, with the "no" response box at the maximum x-value position;

(f) *Low x-value:* a measure of how close (in coordinate position) each trajectory curves toward the "yes" response box, with the "yes" response box at the minimum x-value position;

(g) *x-flips in Latency*: the number of times a trajectory moves back and forth on the x-axis within the latency region; and

(h) *x-flips in Motion*: the number of times a trajectory moves back and forth on the x-axis while in motion to a "yes" or "no" response.

The *total time* and *latency* variables are primarily temporal measures, whereas the remaining variables capture dynamical processes that occur along the trajectory of motion. For example, *x-flips in latency* and *x-flips in motion* provide an intuitive measure of response instability, and *high x-value* and *low x-value* variables are indicators of competing attractor strengths that occur while reacting either to "no" falsely (moving leftward on the x-axis toward the competing "YES" response region) or to "yes" falsely (e.g., moving rightward on the x-axis toward the competing "NO" response region).

A 2 (prompt type: false vs. true) x 2 (response type: yes vs. no) repeated-measures ANOVA was conducted for each of the eight dependent variables. Each variable, as well as their mean values and SEs for each condition, are provided in Table 1. The results of the repeated-measures ANOVA are provided in Table 2.

Table 1. *Means and SEs for the Wiimote trajectory variables by prompt and response type.*

	Yes				*No*			
	False		*Truth*		*False*		*Truth*	
Variable	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Total time (ms)	2806.00	83.00	1996.00	86.00	2802.00	115.00	2408.00	111.00
Latency (ms)	1247.00	73.00	999.00	59.00	1246.00	67.00	1111.00	57.00
Distance (pixels)	1423.00	102.00	996.00	98.00	1426.00	145.00	1273.00	114.00
Motion time (ms)	1558.00	103.00	997.00	65.00	1556.00	122.00	1297.00	107.00
High x-value	229.00	24.00	69.00	12.00	515.00	4.00	515.00	4.00
Low x-value	-505.00	17.00	-509.00	22.00	-190.00	25.00	-139.00	2.00
x-flips in latency	8.67	0.66	6.87	0.50	9.00	0.78	8.32	0.79
x-flips in motion	3.87	0.36	2.20	0.24	4.22	0.48	3.42	0.39

Table 2. *F scores for the repeated measures analysis using Wiimote movements.*

Variable	*Yes vs. no response*	*Truth vs. false prompt*	*Prompt x response*
Total time	9.92**	80.67**	11.85**
Latency	4.68**	25.82**	--
Distance	6.05*	16.03**	4.70*
Motion time	8.04**	44.12**	6.90*
High x-value	693.63**	68.38**	72.72**
Low x-value	288.49**	7.18*	6.77*
x-flips in latency	5.09*	12.83**	--
x-flips in motion	9.41**	38.50**	--

Note: * indicates statistical significance at $p < .05$; ** indicates statistical significance at $p < .001$; the degrees of freedom for all analysis are 1, 24.

To ensure that the results for trajectory properties were not unduly influenced by possible confounds, we also performed a Hierarchical Linear Model (HLM) analysis with random factors that controlled for variance due to practice effects (by nesting trial number in participants), for NO/YES response position, and for items. A noted advantage of a HLM analysis is that it is also quite robust against unequal *N* in conditions, as is the case with our data. In all analysis, the results were consistent with the repeated-measures ANOVA, except the interaction terms for *latency* and *x-flips in latency* were no longer significant.

Discussion

The current study provides the first investigation into the action dynamics of deceptive behavior. The movements of the arm revealed distinct signatures of cognitive activity as participants made false and true responses to autobiographical questions. During false responses, the dynamics were slower and more disorderly than true responses, and were also curved towards a competitor "truth" region that was visually co-present with the target response region. This curvature suggests the presence of a truth-bias attractor that pulls processing "off-course" during the production of a false response (Gilbert et al., 1993; McKinstry et al., 2008). For truthful responses, there was no equivalent pull in the direction of a competitor "false" region.

The competition effects for false responding are similar to decision tasks that involve competition between featurally similar exemplars (Dale et al., 2007), ambiguous syntactic completions (Farmer et al., 2007), and phonological competitors (Spivey et al., 2005). Similar to these studies, the competition exhibited in false responding is the result of cognitive components that evolve smoothly over the response movement itself, suggesting overlapping processes in overcoming an initial belief and generating a false response. This view of deception naturally extends response time measures by incorporating the fine-grained changes that occur during the response. By doing so, a clearer distinction between deceptive and truthful behavior is possible.

One notable distinction is the greater trajectory curvature and slower responses for false "yes" responses compared to false "no" responses. This differentiation within false responses was not found for truth responses. Interestingly, the greater difficulty of falsely saying "yes" is at odds with an earlier finding that normal "yes" responses elicit faster and smoother trajectories than "no" responses, or a "yes-bias" (McKinstry et al., 2008). Indeed, based on Gilbert and colleagues findings (Gilbert, 1991; Gilbert et al., 1993), automatic acceptance of propositions should give "yes" responses a facilitative advantage. However, when deception is involved, this "yes" automaticity conflicts with the more deliberative goal to respond falsely.

To lay the initial foundation in quantifying deceptive response movements, we chose a "guided lie" paradigm that permits a straightforward contrast between false and true response behaviors. Unfortunately, this distinction is not always realized in "real-world" scenarios, where elements of truth are intermingled with the motives and content of a lie. There is also a limitation in our experimental paradigm that concerns the *intention* to deceive. Ekman (1997) argues that deception is an act of conscious volition that requires the deceiver to know what is accurate, and then to purposefully violate that knowledge with false information. Clearly, deception requires a certain degree of motivation that is absent from a guided lie paradigm. Nevertheless, this type of responding is still closely aligned with deceptive behavior, and is widely used and accepted in the deception literature (DePaulo et al., 2003; Vrij, 2001). Of course consensus does not negate further investigation, and future work will allow participants greater choice in the scope of their deception, both in when they deceive and under what circumstances they do it (e.g., personal or social gain, avoiding embarrassment).

It is clear that deception is a complex behavior that garners both theoretical and applied attention. To concoct a false reality requires one to maintain a mental representation of the truth, and then to violate this representation with all appearances of sincerity. Not only is this behavior cognitively challenging, but it also interacts with a host of social, motivational, and emotional factors. Notwithstanding this complexity, researchers have devised a multitude of techniques to identify cues of deception. This study provides the first steps towards applying an action dynamics framework to the exploration of false response behavior. The results suggest that dynamic measures capture deceptive processes that are unavailable to response time measurements alone. If so, these measures could improve existing prediction models that have been touted in recent years (Gregg, 2006; Sartori et al., 2008; Walczyk, Mahoney, Doverspike & Griffith-Ross, 2009), as well as supplement techniques for detecting online deception (Monrose & Rubin, 2000). While there is much more work ahead, we admonish deceivers everywhere: your arm might just reveal when you are lying.

Study 1b: PROMPT-FINAL/QUESTION-AUTO

NOTE: The following studies, Study 1b, 2a, and 2b, are novel extensions of Study 1a that constitute unpublished data. This study, Study 1b, is a computer-mouse replication of Study 1a.

Participants

Fifty-three Mechanical Turk workers were paid 40 cents to participate in this study. Two participants did not reveal their gender or age. Of the 51 participants who provided demographic data, 28 were female and the median age was 29.5 (with a minimum age of 18 and a maximum age of 69 years old). Due to the online nature of data collection, it could not be verified whether participants were predominantly right or left-handed.

Procedure

To replicate the Wiimote experiment (Study 1a), the current procedure is mostly consistent, albeit with a few necessary changes for online implementation. Instead of 80 questions, participants only answered 40 questions. These questions were randomly selected from the original 120 question set. This selection process was repeated ten times to produce ten lists, and of these ten lists, one was randomly presented to each participant. It should also be noted that presentation of the "YES" and "NO" response boxes, on either the right or left sides of the screen, was randomly placed for each participant (and stayed consistent throughout all trials). Also, in using Adobe Flash to present and collect data, the sampling rate of trajectory movements was reduced to 40 Hz (instead of 80 Hz as used in the original Python implementation). And lastly, new colors were selected for the response prompts. The color ORANGE is now associated with a false response instead of the color RED, and the color BLUE is now associated with the true response instead of the color GREEN. The reason for this change is to remove any possible confound of inhibition/facilitation that might be elicited by color type. The color RED is often associated with STOP (e.g., traffic lights and stop signs), and slower responses in false responding might be attributed to this association.

Question Stimuli

The same questions as Study 1a were used.

Results

Based on the responses given in the follow-up verification task, four participants were removed for failing to answer the false response prompts properly. Also, 2.61% of trials were removed for exceeding 3 standard deviations above the mean (e.g., 5633 ms). No participant had more than a single trial that exceeded this upper threshold. In total, there were 536 trials in the “false/no” condition, 590 trials in the “true/yes” condition, 389 trials in the “false/yes” condition, and 373 trials in the “true/no” condition.

Trajectory Shape. The trajectories across each condition, for each participant, are time-normalized to 101 time steps. These trajectories are combined to form four composite vectors, representing each condition (see Figure 3). To determine if false trajectories diverge from a truth baseline, paired t-tests were computed at each of the time-steps, moving up the sequence from 1 to 101 steps to produce 101 t-tests. To show evidence for divergence, there must be a sequence of 8 or more sequential t-tests that are statistically significant (based on bootstrapping methods). For the “no” responses, there was no evidence for divergence between the true and false responses. However, for the “yes” responses, the true and false trajectories diverged between the 48 and 87 time steps, with greater curvature toward a response competitor exhibited in false trajectories.

The failure to find any divergence for false “no” responses is likely the result of using true “no” trajectories as a baseline. A visual inspection of Figure 3 suggests that the true “no” responses are pulled toward a “yes bias” attractor to the same extent that the false “no” responses are pulled toward a “truth” attractor. Given the curvature in true “no” responses, it is difficult to make any direct inferences about false “no” behavior with true “no” trajectories acting as a baseline. A better analysis for examining false response behavior is to compare the false trajectories against each other. Because false “yes” trajectories clearly demonstrate the effect of a “truth” bias (i.e., movement toward the “truth” response competitor), these trajectories will serve as an improved baseline. If no differences are found, then this suggests that false “no” are also influenced by a “truth” bias, at least to the same extent as false “yes” responses.

For the new comparison, false “no” trajectories were mirror-reversed to lie on the same coordinate plane as the false “yes” trajectories. Using the t-test method, there was some evidence that false “no” trajectories diverged toward a truth attractor (from the 17 to 21 time steps), but because this occurred only over 5 time steps, it is not considered statistically significant. Thus, it is safe to conclude that the false “no” and “yes” responses are similar in difficulty, according to a measure of deviation along the x-axis.

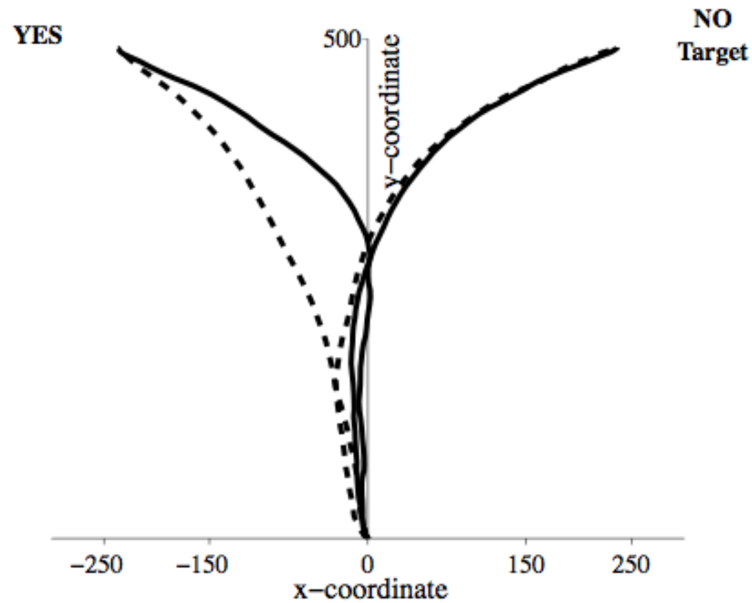

Figure 3. Shape of computer-mouse trajectories for prompt-final presentation, using autobiographical-based questions. False answers (solid lines) display a greater arc towards the competing response option than true answers (dashed lines). However, for "No target" responses, the greater curvature in true "no" trajectories suggest interference from a "yes bias."

Trajectory Location. In this analysis, the x,y coordinates of the "true/yes," "false/yes," "true/no," and "false/no" trajectories were extracted to capture a 50 ms range around three stages in processing, at an initial stage around 500 ms (x,y coordinates between 450 and 550 ms), at a mid-range stage around 1100 ms (x,y coordinates between 1050 and 1150 ms), and a later stage in processing, around 1700 ms (x,y coordinates between 1650 and 1750 ms). These stages and range of times are also used in all subsequent studies. For this study, the average position of each of these time ranges, for each condition across all participants, is plotted in Figure 4.

To evaluate the divergence toward a “truth bias” attractor, separate mixed effects model were conducted at each time range for displacement in x-coordinate position, with subject entered as a random effect. The analysis was conducted using the lmer package in the R statistical software. In this package, *p*-values are computed with 10,000 Monte Carlo Markov Chain simulations, using lmer’s pvals.fnc function (see Baayen, Davidson, & Bates, 2008). I report these *p* values, as well as the SEs from each model.

Starting with a comparison of “no” responses between true and false trajectories, the analysis now shows differentiation at particular time ranges. There were statistically significant displacements of the x-coordinate around 500 ms, $b = -3.26$, $p < .001$, and around 1700 ms, $b = 16.95$, $p = .002$, but not around 1100 ms (Figure 4a). For the “yes” responses, there was evidence of displacement at all time ranges, around 500 ms, $b = -3.15$, $p < .001$; 1100 ms, $b = -74.07$, $p < .001$; and around 1700 ms, $b = -77.08$, $p < .001$. For every time range where a displacement was observed, it was the false response that was most affected by the competitor response option, with the notable exception of the true “no” trajectories around 500 ms (an explanation for this and other patterns provided below in *Discussion*) (Figure 4b).

I also tested for the existence of a general “yes” bias. To do so, the true “no” trajectories (which are susceptible to a “yes” bias) were compared against the true “yes” trajectories. At each time point, the true “no” responses deviated from the true “yes” responses and toward the competitor response option: around 500 ms, $b = 8.72$, $p = .002$; around 1100 ms, $b = 78.27$, $p < .001$; and around 1700 ms, $b = 46.35$, $p = .019$ (Figure 4c). Next, comparing just the false “no” and “yes” responses against each other, there is evidence for an additive effect of a “truth” bias and a general “yes” bias for the false “no” responses. Around the 500 ms mark, the false "no" responses, compared to the false "yes" responses, are shifted 2.5 x-coordinate positions toward the attractor region, $p < .001$. However, at all other time points, there are no statistically significant differences (Figure 4d).

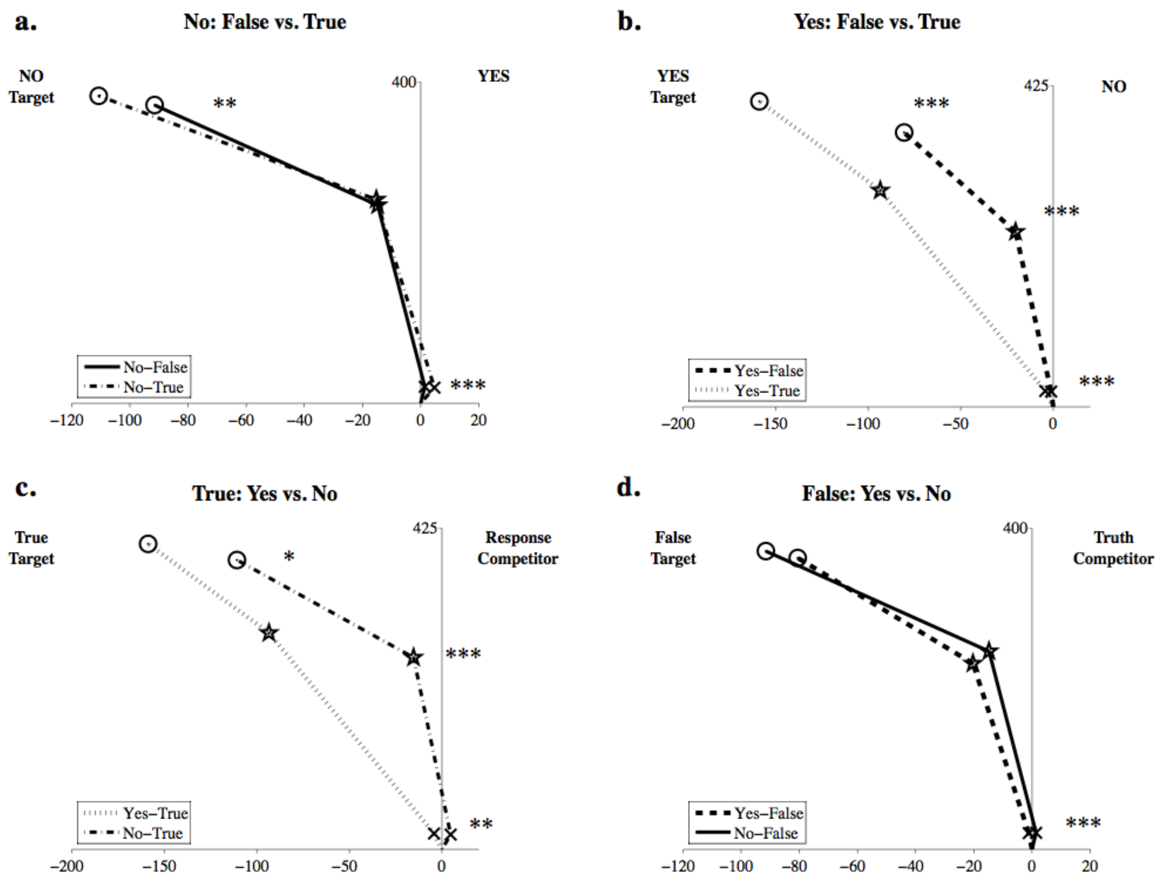

Figure 4. Location of computer-mouse trajectories for prompt-final presentation, using autobiographical-based questions. The *x*-coordinate position is plotted around 500 ms (cross), 1100 ms (star), and 1700 ms (circle). (a) False "no" vs. true "no" trajectories, (b) False "yes" vs. true "yes" trajectories, (c) True "yes" vs. true "no" trajectories, and (d) False "yes" vs. false "no" trajectories.

Temporal and Trajectory Properties. For this set of analysis, I keep the same trajectory variables as Study 1a: (a) *Total Time*, (b) *Latency*, (c) *Distance*, (d) *Motion Time*, (e) *High x-value*, (f) *Low x-value,* (g) *x-flips in Latency*, (h) *x-flips in Motion,* (i) *Time to Reach Peak Velocity*, and (j) *Magnitude of Peak Velocity*. The latter two variables were also used in Study 1a in their own section, but here they are combined with the larger *temporal and trajectory properties* analysis. I also include an additional variable, (k) *Area Under the Curve*, that measures the area between each response trajectory and a hypothetical straight line drawn from the starting and ending positions of each trajectory. This measure has been included for further validation of attractor strength, capturing degree of deviation toward a visually present attractor region.

For each measure, a complete mixed-effects model was conducted comparing all four conditions, “true/yes,” “false/yes,” “true/no,” and “false/no,” controlling for random variation between subjects, items, and trial number embedded in subject (to specifically control for practice effects). The means and SEs for each measure are reported in Table 3, and the outcome of the mixed effects models are reported in Table 4.

To evaluate measures with a statistically significant interaction, planned comparisons were performed between trajectories for false and true “no” responses, and a separate analysis for false and true “yes” responses. For the “no” responses, none of the trajectory measures, which capture moment-to-moment changes in trajectory movement, differed between the true and false response movements. Again, because true “no” trajectories (which act as a baseline for false “no” responses) are affected by competition from a “yes bias”, the failure to find a difference is likely driven by increased processing difficulty for true “no” responses. However, this processing difficulty for true responses is not so great that it supersedes the difficulty associated with false “no” responses. Together, the temporal variables *Total Time*, *Motion Time,* and *Time to Reach Peak Velocity* show that true “no” responses are still faster than those of false “no” responses.

Turning now to the comparison of false “yes” and “no” responses, there were statistically significant differences for the majority of measures. Of the 10 variables evaluated, 8 were significant, including 4 temporal variables: *Total Time*, *Latency, In Motion Time,* and *Time to Reach Peak Velocity,* and 4 trajectory property variables: *Distance, High x-value, x-flips in Motion,* and *Time to Reach Peak Velocity.* For the temporal variables, the difference in means across all variables indicates that the false “yes” trajectories are slower than true “yes.” And similarly, for the trajectory property variables, the false “yes” trajectories travel a longer distance, travel closer to the attractor region, show greater instability, and take longer to reach maximum velocity.

It should also be noted that *Magnitude of Peak Velocity* did not interact with “yes” or “no” responses, and was generally least for “false” responses. This finding indicates that false responses have a much lower maximum velocity than true responses, and highlights the supposition that false responses are slower, and thus can be considered more cognitively difficult to execute.

Table 3. *Means and SEs for Study 1b (prompt-final, computer-mouse) trajectory variables by prompt and response type.*

	Yes				*No*			
	False		*True*		*False*		*True*	
Variable	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Total time (ms)	2567.00	46.00	2077.00	30.00	2571.00	43.00	2346.00	44.00
Latency (ms)	1081.00	29.00	969.00	18.00	1079.00	25.00	1029.00	25.00
Distance (pixels)	876.00	19.00	732.00	12.00	854.00	13.00	872.00	22.00
Motion time (ms)	1486.00	44.00	1108.00	24.00	1492.00	37.00	1317.00	38.00
High x-value	97.00	6.16	48.00	3.96	251.00	1.74	251.00	2.43
Low x-value	250.00	2.17	251.00	1.58	100.00	4.68	99.00	5.89
x-flips in latency	0.27	0.03	0.27	0.03	0.32	0.03	0.29	0.04
x-flips in motion	1.82	0.78	1.39	0.05	1.70	0.07	1.57	0.07
Vel max mag (pix/sec)	2924.00	32.00	2974.00	40.00	2923.00	38.00	3027.00	56.00
Vel max time (ms)	1120.00	31.00	900.00	19.00	1120.00	28.00	1040.00	30.00
Area under	0.12	0.01	0.09	0.01	0.13	0.01	0.14	0.01

Note: Vel max mag = Magnitude of the Maximum Velocity; Vel Max time = Time to Reach Maximum Velocity

Table 4. *b value (estimates) for the LMER analysis for Study 1b (prompt-final, computer-mouse) movements.*

Variable	*Yes vs. no response*	*Truth vs. false prompt*	*Prompt x response*
Total time	-149.00***	366.00***	205.00**
Latency	-40.00*	-82.00***	-68.00*
Distance	-69.00***	-68.00***	163.00***
Motion time	-114.00***	-285.00***	-172.00**
High x-value	-178.00***	-25.00***	-47.00***
Low x-value	-148.00***	--	--
x-flips in latency	--	--	--
x-flips in motion	--	-0.29***	-0.31*
Vel max mag	--	80.26*	--
Vel max time	-4.57***	-7.99***	-6.49**
Area under	-0.02***	-0.01**	-0.03**

Note: * indicates statistical significance at $p < .05$; ** indicates statistical significance at $p < .01$; *** indicates statistical significance at $p < .001$

Discussion

In this study, the goal was to replicate the action dynamics of false responding that were found in Duran et al. (2010), and to do so with computer mouse movements that were collected outside the laboratory. The evaluation of the movements followed the same analytical approach as Duran et al., and in general, provided similar results (also see Appendix D for an analysis based on mouse movements collected in the laboratory; results are again comparable). Overall, I was able to distinguish “false” from “true” responses in terms of a “truth” attractor. This attractor, which was visually co-present with the target response option, acted as a competitor, and quite literally pulled the arm toward it. This effect was most clearly demonstrated in the false “yes” responses, where a strong competitor attractor was found for false trajectories, both in the *Trajectory Shape* and *Trajectory Location* analysis. For false “no” trajectories, there was little evidence for a competitor attractor – but only when comparing false “no” trajectories to true “no” trajectories. In contrast, the comparison of false “no” to false “yes” trajectories showed near equivalence, for all 11 *Temporal and Trajectory Property* variables and with the *Trajectory Location* analysis (except at the early stage of processing, which will be interpreted below). The reason that the signal for false “no” responses appeared weak when compared to true “no” responses is because of a strong “yes bias” present in the true “no” trajectories.

However, a “yes bias” is likely a factor in more than just true “no” responses, and might also be present at the earliest moments of processing for false “no” responses. At the initial processing stage (around 500 ms), false "no" trajectories were more likely to be pulled toward a competitor region than false “yes” trajectories (Figure 4d). One explanation for this result is that at the earliest moments of processing, there is competition not only from a "truth bias," but also from a "yes bias," where apparently the "yes bias" is strongest (as evidenced by the true “no” trajectories with greater deviation toward the competitor region than false “no” trajectories; see Figure 4a, at 500 ms). Interesting, when comparing the strengths of a “yes bias” (based on the deviation of true “no” trajectories) versus a “truth bias” (based on the deviation of false “no” trajectories), both biases are nearly equal at 1100 ms, until the “yes bias” becomes significantly weaker, compared to the “truth bias” at 1700 ms (Figure 4a). This series of comparisons reveal the shifting attractor dynamics of different response modes as processing evolves over a relatively short timescale, from 500 to 1700 ms. One important insight from these analyses is that clearest signal of deception, where there is the maximal influence of a "truth bias" with minimal influence of a "yes" bias, is in the later stages of response execution.

Study 2a: PROMPT-INITIAL

Procedure and Question Stimuli

The procedure was identical as to the previous Study 1b, but with the critical change of now presenting the response prompt at the beginning of each question, with sustained presentation throughout each question trial. Thus, for the question, "Have you ever met Elvis?," a true or false prompt would appear at the word "Have" and stay visible until YES or NO was selected. The questions were also identical as the earlier study, although with new stimulation lists that were randomly generated (i.e., 40 questions from a 120 question set).

Participants

Fifty-two participants were recruited from Amazon Mechanical Turk and paid the same amount as in Study 1b (40 cents). In this sample, there were 33 females and 17 males, with a median age of 32 (maximum age: 77 and minimum age: 18).

Results

Forty-nine participants were retained after removing three participants who failed to follow directions correctly. 1.80% of the data was removed for exceeding 3 standard deviations above the mean (e.g., 6092 ms). No participant had more than a single trial that exceeded this upper threshold. In total, there were 516 trials in the "false/no" condition, 569 trials in the "true/yes" condition, 386 trials in the "false/yes" condition, and 374 trials in the "true/no" condition.

Trajectory Shape. The false and true “no” time-normalized trajectories were evaluated by running t-tests at each of the 101 positions (Figure 5). Like Study 2b, the “no” responses did not display any signature of deviation. However, for the false and true “yes” time-normalized trajectories, there was consistent separation for 40 time steps, between the 40th and 80th time steps.

Comparing just “true” responses, there was evidence for a “yes bias” for true “no” responses, with separation of true “no” responses from true “yes” responses for 50 time steps, between the 22nd and 72nd time steps. Next, comparing just “false” responses, there was no separation, indicating that the attraction strength for the false “no” responses (where a “yes bias” might be present) was just as strong for false “yes” responses.

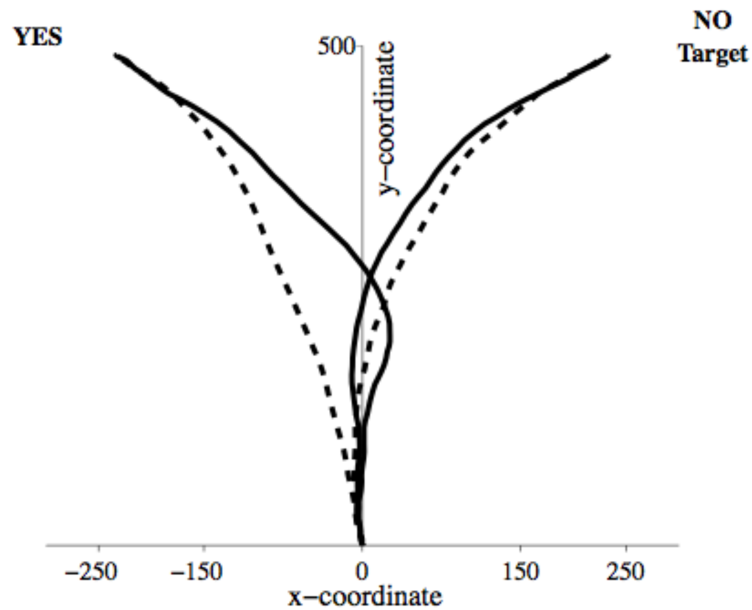

Figure 5. Shape of computer-mouse trajectories for prompt-initial presentation, using autobiographical-based questions. False answers (solid lines) display a greater arc towards the competing response option than true answers (dashed lines). However, for "NO target" responses, the greater curvature in true "no" trajectories suggest interference from a "yes bias."

Trajectory Location. Mixed effects models were used to evaluate the x-coordinate position of false and true "no" response movements at 500 ms, 1100 ms, and 1700 ms (Figure 6a). There was statistically significant displacement of the x-axis around 500 ms, $b = -2.39$, $p < .001$, around 1100 ms, $b = 23.46$, $p = .04$, and around 1700 ms, $b = 54.41$, $p = .001$. For the displacement around 1100 and 1700 ms, it is the false "no" responses, compared to the true "no" responses, that are most influenced by the attractor region, i.e., pulled toward the competitor response option. However, this pattern reverses around 500 ms, where the "yes bias" is most potent, such that it pulls true "no" trajectories away from the false "no" trajectories and towards the competitor response option.

The comparison between false and true "yes" response movements was much more straightforward (Figure 6b). At every position, the false "yes" responses deviated from the true "yes" responses and towards the competitor response option: around 500 ms, $b = -5.75$, $p < .001$, around 1100 ms, $b = -79.42$, $p < .001$, and around 1700 ms, $b = -70.86$, $p < 001$. However, as discussed for Study 1b, the more "straightforward" nature of the false "yes" responses does not necessarily mean that they are influenced more by a "truth bias" attractor. Indeed, when comparing just false trajectories (Figure 6d), there were no statistically significant differences at any time range (although false "yes" responses appear to be more difficult based on overall temporal measures, as discussed in the *Temporal and Trajectory Properties* section).

And finally, the assumed "yes bias" was indeed present. When comparing just true responses, the true "no" responses (i.e., those that would be susceptible to a "yes bias") exhibited the greatest effect of response competition (Figure 6c). Accordingly, all mixed effects models comparing x-coordinate position of true "no" trajectories versus true "yes" trajectories revealed greater competition for "no" responses, around 500 ms, $b = 5.64$, $p < .001$, around 1100 ms, $b = 56.63$, $p < .001$, and around 1700 ms, $b = 17.51$, $p = .003$.

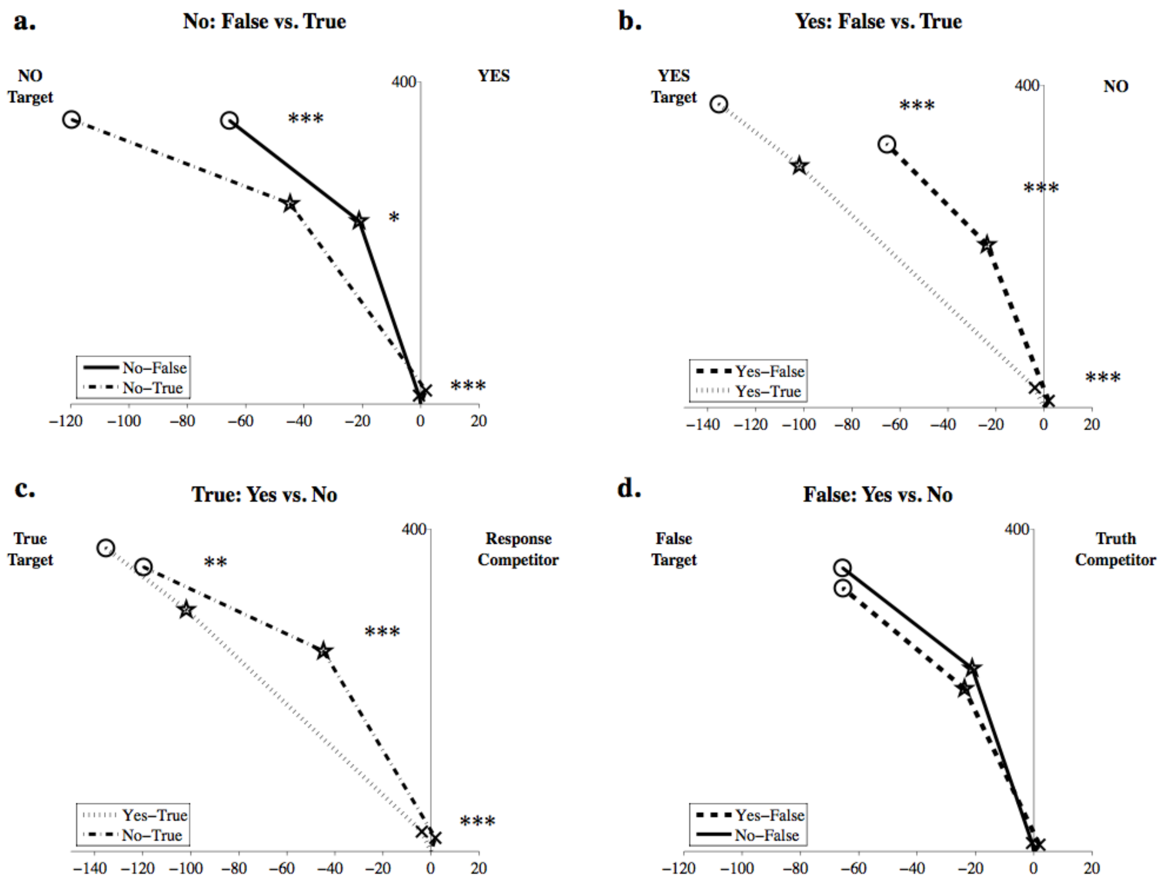

Figure 6. Location of computer-mouse trajectories for prompt-initial presentation, using autobiographical-based questions. The *x*-coordinate position is plotted around 500 ms (cross), 1100 ms (star), and 1700 ms (circle). (a) False "no" vs. true "no" trajectories, (b) False "yes" vs. true "yes" trajectories, (c) True "yes" vs. true "no" trajectories, and (d) False "yes" vs. false "no" trajectories.

Temporal and Trajectory Properties. The same 11 variables used in the previous study were used here. The means and SEs for each variable are reported across the four response conditions (see Table 5). Mixed effects models were also run for each variable (controlling for subjects, items, and practice effects). The *b* values and significance values for the main effects of answer (“yes” vs. “no”), prompt (“false” vs. “true”), and their interaction are reported in Table 6.

Before investigating the interactions in greater detail, it should be mentioned that two measures, *Distance Traveled* and *x-flips in Motion,* were driven solely by the *prompt* factor; that is, there was no main effect or interaction with the *answer* factor (“yes” vs. “no”). Thus, false responses, in general, traverse a greater distance and have increased instability.

Of the interactions that were statistically significant, starting with “no” responses, planned comparisons revealed increases in *Total Time*, *Total Time in Motion*, and *Time to Reach Maximum Velocity* for false responses, suggesting increased cognitive difficulty overall. For “yes” responses, the same temporal variables were significant, with the addition of the *Latency* variable that showed slower initiation time for false responses. There were also differences with the trajectory property variables *High x-value and Area under the Curve*, indicating that false “yes” responses, compared to true “yes” responses, traveled closer to the response competitor.

For the next analysis, true “no” and “yes” trajectories were evaluated with each other. The true “no” responses were slower and more unstable than true “yes” responses, with the true “no” responses taking longer overall (b = -219.95, p < .001), taking longer to initiate (b = -67.41, p = .009), spending more time in motion (b = -182.98, p < .001), taking longer to reach maximum velocity (b = -3.40, p = .021), and curving more, overall, toward a response competitor (b = -.015, p = .05). Again, these results suggest the presence of a “yes” bias influencing “no” target responses.

Finally, false trajectories were compared independent of the "true" trajectories. In doing so, it appears that the false "yes" responses are slower and more susceptible to a response competitor, i.e., "truth bias" than false "no" responses, specifically, by taking longer to initiate (b = 130.39, p < .001), taking longer to reach maximum velocity (b = 5.44, p = .006), and bending more, overall, toward a response competitor (b = .02, p = .008).

Table 5. *Means and SEs for Study 2a (prompt-initial, computer-mouse) trajectory variables by prompt and response type.*

	Yes				*No*			
	False		*True*		*False*		*True*	
Variable	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Total time (ms)	2794.00	54.00	2099.00	39.00	2645.00	45.00	2448.00	52.00
Latency (ms)	1367.00	42.00	1066.00	26.00	1223.00	30.00	1190.00	33.00
Distance (pixels)	879.00	25.00	718.00	13.00	951.00	95.00	798.00	18.00
Motion time (ms)	1427.00	42.00	1033.00	24.00	1423.00	38.00	1257.00	39.00
High x-value	96.00	6.17	41.00	3.72	255.00	2.82	249.00	3.14
Low x-value	-248.00	2.76	-248.00	1.85	-122.00	32.00	-70.00	5.21
x-flips in latency	0.26	0.03	0.20	0.02	0.22	0.02	0.22	0.03
x-flips in motion	1.64	0.08	1.25	0.06	1.69	0.07	1.47	0.07
Vel max mag (pix/sec)	3163.00	74.00	3224.00	45.00	3533.00	364.00	3089.00	56.00
Vel max time (ms)	1300.00	37.00	940.00	25.00	1180.00	29.00	1080.00	32.00
Area under	0.13	0.01	0.10	0.01	0.11	0.01	0.12	0.01

Note: Vel max mag = Magnitude of the Maximum Velocity; Vel Max time = Time to Reach Maximum Velocity

Table 6. *b values (estimates) for the LMER analysis for Study 2a (prompt-initial, computer-mouse) movements.*

Variable	*Yes vs. no response*	*Truth vs. false prompt*	*Prompt x response*
Total time	-93.00**	-425.00***	-322.00**
Latency	--	-151.00***	-173.00**
Distance	--	-163.00**	--
Motion time	-118.00***	-281.00***	-186.00**
High x-value	-181.00***	-32.00***	-46.00***
Low x-value	-153.00***	--	--
x-flips in latency	--	--	--
x-flips in motion	-0.13*	-0.32***	--
Vel max mag	--	--	--
Vel max time	--	-11.00***	-8.93**
Area under	--	-0.01*	-0.04**

Note: * indicates statistical significance at $p < .05$; ** indicates statistical significance at $p < .01$; *** indicates statistical significance at $p < .001$.

Discussion

This study involved a simple task-based constraint of cueing the false or true prompt early in the trial, thus allowing the participant to anticipate how to respond (either truthfully or falsely) while the question was being presented. Many of the same effects as the previous study were found. Focusing on the *Trajectory Location* analysis, where the results are clearest, the comparison between the false "yes" and true "yes" trajectories revealed a strong competitor "attractor" for the false trajectories, where the effect persisted at every time point, from initial (450-550 ms), to mid (1050-1150 ms), to later processing (1650-1750 ms) (see Figure 6b). To determine if the strength of the competitor attractor for false "yes" responses was equivalent to that of the false "no" responses, these two response types were compared, and trajectories were shown to be equivalent at every time range (see Figure 6d). Thus, false responses, regardless of whether the response required a “yes” or a “no,” were influenced equally by a “truth bias” attractor. But there was a curious result when comparing false “no” responses around 500 ms to true “no” responses. As already shown, true “no” responses are likely influenced by a “yes bias,” whereby there is a tendency to respond “yes” when en route to a “no” response. This deviation toward the “YES” response option, around 500 ms, was strongest of the four response conditions, including false “no” responses that also deviate toward “YES” (see Figure 6a). Thus, there is evidence for a very early and strong "yes bias" that is stronger than any “truth bias.” However, as processing continues, the strength of the "yes bias" for true "no" responses became much weaker than the "truth bias" for false "no" responses, with significant differences around 1100 ms and 1700 ms. This pattern also occurred in the previous PROMPT-FINAL study, but it was only around the 1700 ms mark that the “truth bias” overcame the “yes bias.”

This difference in a “yes bias/truth bias” trade-off between the PROMPT-INITIAL and PROMPT-FINAL studies suggests that the “yes bias,” although still present, was weaker in this PROMPT-INITIAL study. Thus, any influence that the “yes bias” had on the false responses might be somewhat diminished, and suggests that the competition for false responses is better attributed to a truth bias alone, rather than a truth bias plus a “yes bias.” However, in the current study, it is not clear to what extent these biases can be separated. What is more certain is that for true responses, a “yes bias” was indeed weaker in the PROMPT-INITIAL study, and weaker still when compared to false response trajectories.

Comparing Late vs. Early Prompts

Although the results presented in Study 2a (PROMPT-INITIAL) had considerable overlap with Study 1b (PROMPT-FINAL), critical differences between the studies show how simple processing cues can change the dynamical landscape of true and false responding. I begin with the *Trajectory Location* analysis, which is, as mentioned previously, well adapted for demonstrating attractor strength modulation at various points along a response movement.

One example of such modulation is the apparent weakening of a “yes bias” (embodied in true “no” trajectories) that occurred at different rates between the PROMPT-FINAL and PROMPT-INITIAL studies. Using mixed-effects ANOVAs, with subject as a random factor, the comparison between STUDY for true “no” responses revealed statistically significant effects around 500 ms ($b = 2.90$, $p = .005$) and 1100 ms ($b = 28.77$, $p < .001$). At both times, the influence from the “yes” response competitor was much less for the PROMPT-INITIAL study (see Figure 7a). It should also be noted that even though there is modulation, the “yes bias” is still stronger than any “no bias” (suggesting a “no bias” might not exist at all for true responses, or is at a floor level).

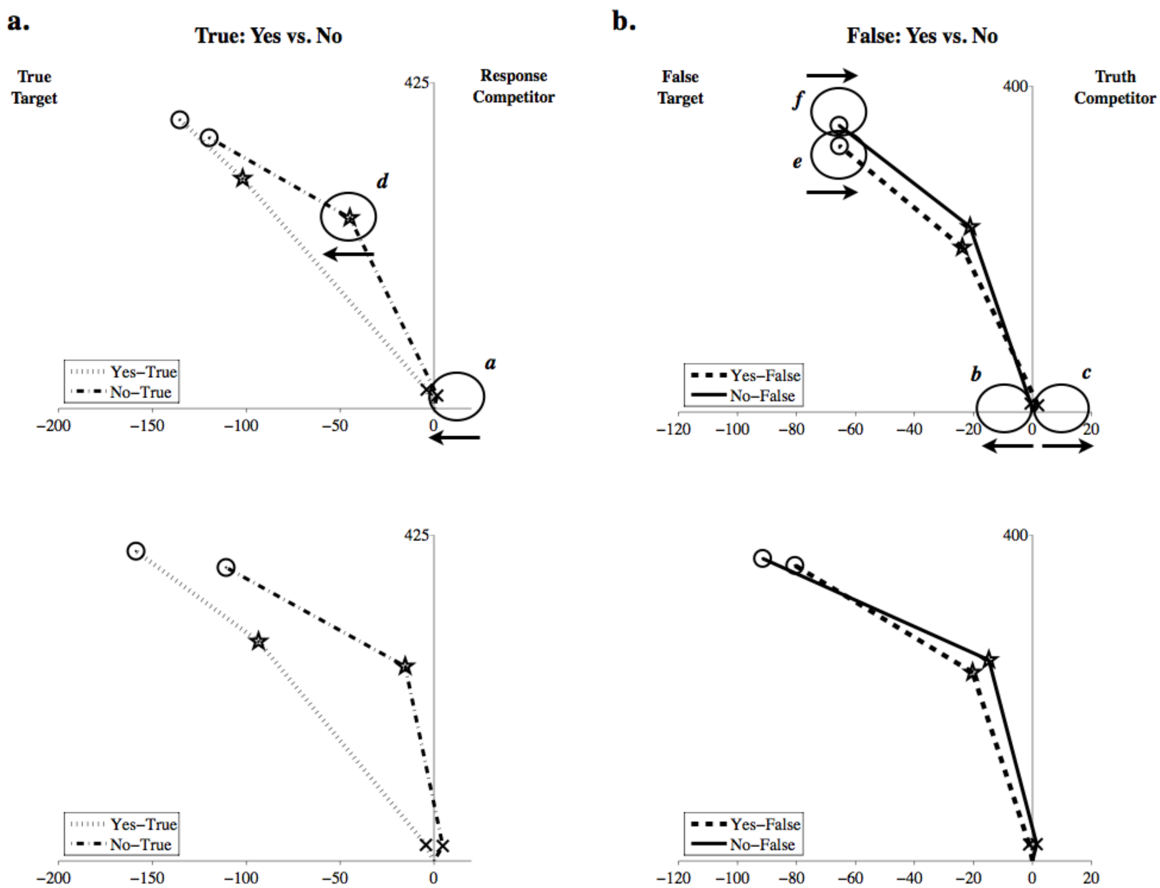

Figure 7. Modulation of attractor dynamics of trajectory movements around 500 (star), 1100 (star), and 1700 ms (circle). Using Study 1b (PROMPT-FINAL) as a reference (the figures in the bottom right and left panels), circled regions indicate whether trajectories in Study 2a (PROMPT-INITIAL) were shifted towards or away from the critical competitor region. All circled regions show statistically significant differences. In (a, top panel), "yes bias" is decreased at 500 (*region a*) and 1100 ms (*region d*). In (b, top panel), greater tendency to say "no" in false "yes" responses at 500 (*region c*) and 1700 ms (*region e*). For false "no" responses, facilitation to say "no" at 500 ms (*region b*), but interference with the truth competitor at 1700 ms (*region f*).

Turning now to responses where participants were prompted to respond falsely, the competitor response is now the "truth". And as reported earlier, this competitor is stronger than any "yes bias" or "no bias." However, competitor strength is also modulated by whether the response is a "yes" or a "no," and whether participants are given the false prompt at the beginning at each question (PROMPT-INITIAL) or at the end of each question (PROMPT-FINAL). In particular, false "yes" responses are most influenced by a "truth bias" in the PROMPT-INITIAL study; being pulled more toward the "no" truth competitor around 500 ms (b = -2.99, p = .003) and around 1700 ms (b = -17.82, p = .003) (see Figure 7b). A similar, but somewhat more complex pattern emerges for the false "no" responses. Here, the "yes" truth competitor is weakened early on, around 500 ms in the PROMPT-INITIAL study (b = 2.02, p < .001), but by 1700 ms, the "yes" truth competitor is much stronger in PROMPT-INITIAL (b = -25.90, p < .001).

Taken together, the differences between PROMPT-FINAL and PROMPT INITIAL suggest that early prompt presentation acts as a prime that influences the activation of "truth values." It is also likely that this activation occurs at different rates depending on whether one is primed to respond truthfully or falsely. Beginning with the "truth", when there is no prime present, there is an early and strong tendency to confirm responses, in other words, as tendency to say "yes." Thus, when a target answer is "yes," this response is aligned with the "yes bias," and will be faster, more stable, and less influenced by competitor response options than other response types. Indeed, these response dynamics were present for target "yes" responses when compared with the "no" responses – a response type which was misaligned with the "yes bias." However, with a "true" prime (PROMPT INITIAL study), the immediate bias to say "yes" was diminished for “no” responses, such that the truth-value, even if it requires denial (as with true “no” responses), was more readily employed during processing. Thus, there is support for the hypothesis that “preparation” for how to respond will diminish response competition.

However, this hypothesis was not supported when primed with a false prompt. Rather, a false prime tended to increase the attraction of the response competitor, particularly when the competitor was “no.” A false prime tends to increase a tendency to deny information, that is, to say “no.” When a false response required one to say “yes,” these participants were immediately pulled toward the “no” region, and continued to be influenced by the “no” attractor through the later stage of processing. Thus, when there is competition between a bias to deny and a target answer of “yes,” interference in the response movement will be greatest. But what about when a bias to deny is matched with a target answer of “no?” At the initial moments of processing, it does appear that there is automatic facilitation; responses do indeed move toward the false “no” response. But downstream in processing, at the later stage (around 1700 ms), there is a reversal, such that the true “yes” competitor becomes more active. This effect suggests that the false prime activates a denial response that has a facilitatory role early on, but is soon overcome when the “truth value” of the proposition is also activated, and interferes with the goal of responding falsely. Thus, there is little support for the hypothesis that a false prime will make it easier to respond falsely. This hypothesis, commensurate with a discrete-account of cognitive processing, is based on the notion that a false prime should remove a cognitive “step” of “deciding to falsify” (see Walczyk et al., 2003; 2005). Instead, “preparation” or “foreknowledge” that one must respond falsely appears to activate a bias to say "no," as well as increases the strength of a "truth bias," making the task of false responding that much harder.

The modulation of response competition between response types suggests the presence of three unique (but overlapping) attractors that are influential at different stages in processing: a confirmation attractor (i.e., "yes" bias), a denial attractor, and a truth attractor. The confirmation attractor was evident with the true "no" trajectories, which appeared to create the most interference around 500 ms to 1100 ms (early to mid-range stages of processing) (see Figure 7a and 7c). The denial attractor was most evident with the false "no" trajectories, which had the greatest facilitatory role around 500 ms (early stage of processing) (see Figure 7b-*b*). The denial attractor was also likely interfering with the false "yes" trajectories around 500 ms, and quite possibly around 1700 ms (see Figure 7c-*c*/*e*). But both false trajectories were also likely influenced by a truth attractor that interfered with processing. This truth attractor was most clearly demonstrated with the false "no" trajectories around 1700 ms (see Figure 7b-*f*). At this point in time, the truth attractor interfered with processing more than any facilitation that might have emerged with the denial attractor. Thus, the patterns here suggest that confirmation and denial attractors (corresponding to a true and false prime, respectively) were activated early, but the truth attractor (triggered by a false prime) occurred later in processing.

There is also one last comparison that supports the attractor dynamics described above. According to the attractors at play, the false "yes" responses are likely most difficult in the context of a "false" prime. To say "yes" falsely, there is interference from a tendency to deny information (say "no"), and interference from the truth. Using *the Temporal and Trajectory Properties* analysis to compare these false "yes" trajectories with false "no" trajectories, it was found that the false "yes" trajectories are slower overall ($b = 196.47$, $p = .003$), slower to initiate ($b = 261.19$, $p < .001$), and take more time to reach maximum velocity ($b = 7.42$, $p = .002$). Thus, these results suggest that the best way to catch a liar is to forewarn the person that a lie may be necessary, and then ask a question that must be confirmed in order to reveal the deception. For example, someone suspected (and is guilty) of reckless driving might find the question: "Were you driving with caution?" challenging, particularly if the person had been forewarned or asked to intentionally lie on previous questions.

Study 2b: QUESTION-SEMANTIC

Procedure and Question Stimuli

The procedure used here was similar to *Study 1b: PROMPT-FINAL*, whereby the prompt occurred in conjunction with the presentation of the final word. The critical change in this study is the nature of the questions. The questions in the previous studies probed autobiographical information. Here, the questions involve semantic information that tap simple trivia-like knowledge (questions are listed in Appendix C). Semantic questions were chosen to represent information that people generally do not explicitly think about on a daily basis. Previous research has found that information that is frequently retrieved also becomes easier to recall (Danker & Anderson, 2010). Such questions also match the autobiographical questions, which generally dealt with material that people do not frequently think about. There are two lists of 30 questions each. A question in one list is identical to a question in the second list except for the final word. For example, a question in the first list might read: "Is Moscow the capital of Russia?" and the matching question in the second list would be: "Is Moscow the capital of

Germany." Thus, the questions in one list are always truthfully answered with a "yes," and the questions in the second list are always truthfully answered with a "no." To create stimuli lists, 14 questions from each list were randomly selected, ensuring that the two versions of the same question never appeared together. Each question was then randomly paired with a false or true response prompt. In this way, the four response conditions could be equally represented (e.g., true "no" responses", true "yes" responses", false "no" responses, and false "yes" responses). A total of 10 stimuli lists were created, and participants were randomly assigned to one of these stimuli lists.

Participants

Fifty-eight participants were recruited from Amazon Mechanical Turk and paid the same amount as in Study 1b and 2a (40 cents). Six participants did not voluntarily provide demographic data, but of the 52 that did, there were 31 females and 21 males, with a median age of 27 (maximum age: 60 and minimum age: 18).

Results

Fifty participants were retained after removing eight participants who did not follow directions correctly. Also, 1.78% of the data was removed for exceeding 3 standard deviations above the mean (e.g., 6614 ms). No participant had more than a single trial that exceeded this upper threshold. In total, there were 284 trials in the "false/no" condition, 291 trials in the "true/yes" condition, 290 trials in the "false/yes" condition, and 292 trials in the "true/no" condition.

Trajectory Shape. The time-normalized trajectories did not diverge between false "no" and true "no" trajectories; however, the trajectories for false "yes" and true "yes" trajectories did diverge between the 56 and 81 time steps (for 43 steps; see Figure 8).

Comparing just true responses, there is divergence between the “yes” and “no” trajectories, between the 56 and 81 time steps, with the true “no” responses being more susceptive to the pull of an alternative response attractor (embodied in the visually co-present “YES” response target). Turning now to false responses, there is also divergence between “yes” and “no” responses, between the 41 and 57 time steps, with the false “yes” responses being much more influenced by a “truth” attractor.

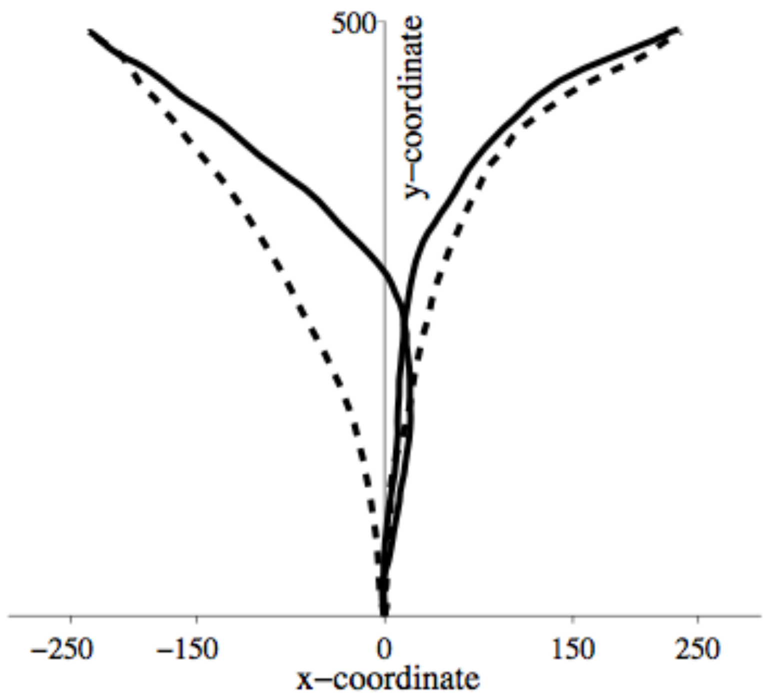

Figure 8. Shape of computer-mouse trajectories for prompt-final presentation, using semantic-based questions. False answers (solid lines) display a greater arc towards the competing response option than true answers (dashed lines). However, for “NO target” responses, the greater curvature in true “no” trajectories suggest interference from a “yes bias.”

Trajectory Location. The comparison between false "no" and true "no" responses revealed no statistically significant differences between x-coordinate position around the 500 and 1100 ms time points. However, around 1700 ms, there was significant divergence (b = -16.11, p = .016), with the true "no' responses most influenced by a competitor region (see Figure 9a). On the other hand, for false "yes" and true "yes" responses, there was separation around every time point, with the false "yes" responses shifted toward the attractor region more so than the true "yes" responses: around 500 ms (b = -1.62, p .025), around 1100 ms mark (b = -94.51, p < .001), and around 1700 ms mark (b = -65.67, p < .001) (see Figure 9b). Next, comparing just true responses, there is separation between "yes" and "no" trajectories also around every time point: 500 ms (b = 1.21, p = .047), 1100 ms (b = 63.12, p < .001, and 1700 ms (b = 52.39, p < .001), with the greatest influence of an attractor on the true "no" trajectories (see Figure 9c). And for just false responses, there was a statistically significant effect at the 1100 ms mark (b = -37, p < .001) and at the 1700 ms mark (b = 25.51, p < .001) with false "yes" trajectories showing greater deviation (see Figure 9d).

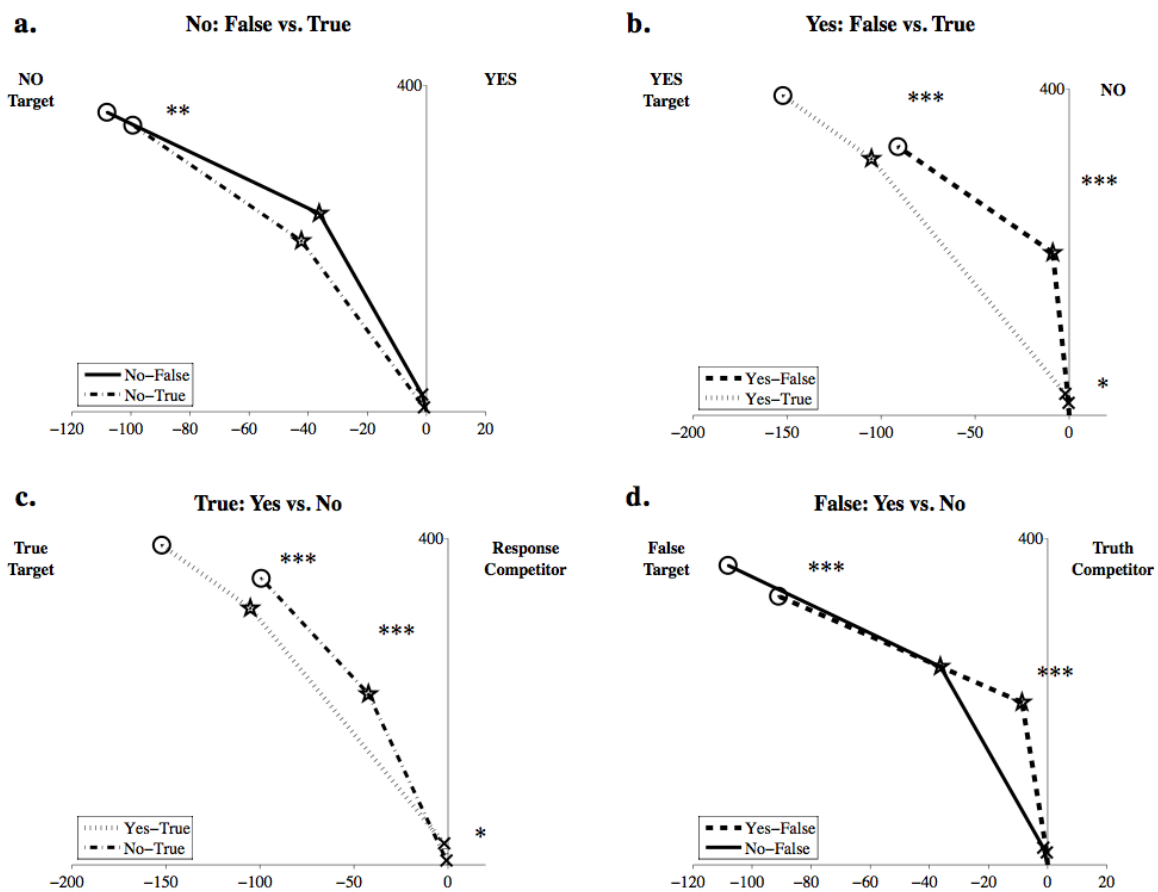

Figure 9. Location of computer-mouse trajectories for prompt-final presentation, using semantic-based questions. The *x*-coordinate position is plotted around 500 ms (cross), 1100 ms (star), and 1700 ms (circle). (a) False “no” vs. true “no” trajectories, (b) False “yes” vs. true “yes” trajectories, (c) True “yes” vs. true “no” trajectories, and (d) False “yes” vs. false “no” trajectories.

Temporal and Trajectory Properties. The means and SEs for all property variables are reported in Table 7. The *b* and *p-values* for the mixed effects models are reported in Table 8. Planned comparisons of the interactions between *answer* and *prompt* type show that there were no variables that distinguished the false “no” responses from the true “no” responses. For the false “yes” and true “yes” responses, every variable showed increased difficulty and instability for the false “yes” responses. Furthermore, the measure *Area Under the Curve* did not interact with *answer* and *prompt type*, but did show a significant main effect for prompt type, with greater deviation exhibited in false responses compare to true responses.

Comparing just “false” response types, there were few differences. However, there was evidence that false “yes” responses took longer to initiate (b = 103.02, p = .019), but false “no” responses exhibited more x-flips while in motion (b = -0.25, p = .03). Next, in comparing true “no” and “yes” responses, true “no” responses were much more difficult and unstable than true “yes” responses, with the true “no” responses taking longer overall (b = -481, p < .001), taking longer to initiate (b = -153.68, p < .001), traveling a greater distance (b = -167, p < .001), spending more time in motion (b = -296.29, p < .001), exhibiting more x-flips while in motion (b = -0.61, p < .001), and taking longer to reach maximum velocity (b = -12.57, p < .001). These latter results of greater difficulty in true “no” responses supports the general conclusion of a “yes bias” that interferes with true “no” processing.

Table 7. *Means and SEs for Study 2b (question-semantic, computer-mouse) trajectory variables by prompt and response type.*

	Yes				*No*			
	False		*True*		*False*		*True*	
Variable	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Total time (ms)	2710.00	58.00	2164.00	55.00	2602.00	60.00	2670.00	55.00
Latency (ms)	1301.00	41.00	1057.00	33.00	1176.00	38.00	1263.00	33.00
Distance (pixels)	779.00	16.00	669.00	11.00	820.00	21.00	820.00	22.00
Motion time (ms)	1409.00	43.00	1107.00	34.00	1426.00	44.00	1407.00	41.00
High x-value	80.00	5.70	31.00	4.05	245.00	2.89	247.00	2.55
Low x-value	-244.00	2.18	-241.00	1.99	-76.00	5.72	-72.00	5.77
x-flips in latency	0.15	0.03	0.12	0.03	0.12	0.02	0.15	0.03
x-flips in motion	1.31	0.07	0.93	0.07	1.54	0.08	1.49	0.08
Vel max mag (pix/sec)	2828.00	57.00	3090.00	53.00	2942.00	67.00	2951.00	65.00
Vel max time (ms)	1200.00	35.00	900.00	30.00	1120.00	38.00	1180.00	33.00
Area under	0.12	0.01	0.11	0.01	0.13	0.01	0.12	0.01

Note: Vel max mag = Magnitude of the Maximum Velocity; Vel Max time = Time to Reach Maximum Velocity

Table 8. *b values (estimates) for the LMER analysis for Study 2b (question-semantic, computer-mouse) movements.*

Variable	*Yes vs. no response*	*Truth vs. false prompt*	*Prompt x response*
Total time	-170.00***	-243.00***	-536.00***
Latency	--	-84.00**	-243.00**
Distance	-102.00***	-65.00***	-142.00***
Motion time	-160.00***	-148.00***	-252.00**
High x-value	-191.00***	-24.00***	-52.00***
Low x-value	-167.00***	--	--
x-flips in latency	--	--	--
x-flips in motion	-0.41***	-0.21**	-0.37*
Vel max mag	--	126.00*	228.00*
Vel max time	-4.44**	-5.82***	-14.00**
Area under	--	0.02*	--

Note: * indicates statistical significance at $p < .05$; ** indicates statistical significance at $p < .01$; *** indicates statistical significance at $p < .001$

Discussion

This study used the same “PROMPT-FINAL” set-up as in Study 1b, but differed according to the type of questions asked. Here, the questions probed knowledge of simple facts, requiring one to verify or falsify information retrieved from semantic knowledge. First, examining the *Trajectory Shape* analysis, there appears to be a “yes bias” that is clearly demonstrated in the true “no” trajectories (Figure 8). This is also validated in the *Trajectory Location* analysis, where the true “no” trajectories diverged from the true “yes” trajectories throughout the entire course of processing (from 500 to 1100 to 1700 ms) (Figure 9a). Given that the true “no” trajectories were influenced by a “yes bias,” these trajectories make a poor baseline in which to compare false “no” responses. By doing so, no separation was found between true "no" and false "no" responses with *Trajectory Shape*, and where differences were found (based on the *Trajectory Location* analysis), there was a stronger "yes bias" than a "truth bias" around 1700 ms (Figure 9a). Thus, the false "no" responses appear less influenced by a "truth bias." To examine this finding further, I compared false "no" and false "yes" trajectories to each other. The false "yes" trajectories are a good baseline because they clearly deviate from their true "yes" counterparts (at every time range), which suggests a strong "truth bias" (Figure 9b). The results show that, compared to false "yes" trajectories, the false "no" responses were less likely to be influenced by a truth attractor, mostly around mid and later stages of processing (around 1100 and 1700 ms) (Figure 9d). Overall, when accessing from semantic knowledge, the greater interference from the truth is when falsely confirming information (with a "yes" response) than when falsely denying information (with a "no" response). Interestingly, this is the opposite pattern of when responding truthfully. During these responses, the least interference is when truthfully confirming information compared to truthfully denying information (Figure 9c). This overall pattern was also generally found for both Studies 1b and 2a.

Comparing Autobiographic vs. Semantic (Late Prompts)

Although the general response pattern between studies is similar, changes in simple parameters can also systematically alter response dynamics. These changes make particular responses more or less easy, and are important for detection purposes, as well as for understanding the cognitive basis of deception. As was done for Study 2a, I compare the current study's (Study 2b) trajectory properties with Study 1b. Both studies employ a "PROMPT-FINAL" presentation, but differ in terms of question type. For this section, Study 1b will be referred to as QUESTION-AUTO because the questions used in that study draw on autobiographical memory. Likewise, the current study (Study 2b) will be referred to as QUESTION-SEMANTIC because the questions used here draw on semantic memory.

Figure 10 shows the differences between studies, using the circle and letter nomenclature employed in the earlier analysis. Beginning with "true" responses (Figure 10a), the attraction toward a response competitor (and therefore the interference from the competitor), is decreased significantly for true "no" trajectories in QUESTION-SEMANTIC, beginning from the early stage to mid-range stage of processing (500 to 1100 ms) (Figure 10a-*a*/c). Thus, the "yes bias" is greatly diminished for true "no" trajectories. The true "yes" trajectories are also less swayed by the attractor region (i.e., less interference) in QUESTION-SEMANTIC, with the largest decrease around 1100 ms into processing (Figure 10a-*d*). Taken together, these results suggest that it is easier to truthfully respond to questions that require one to confirm or deny semantic facts, and more difficult to confirm or deny autobiographical facts.

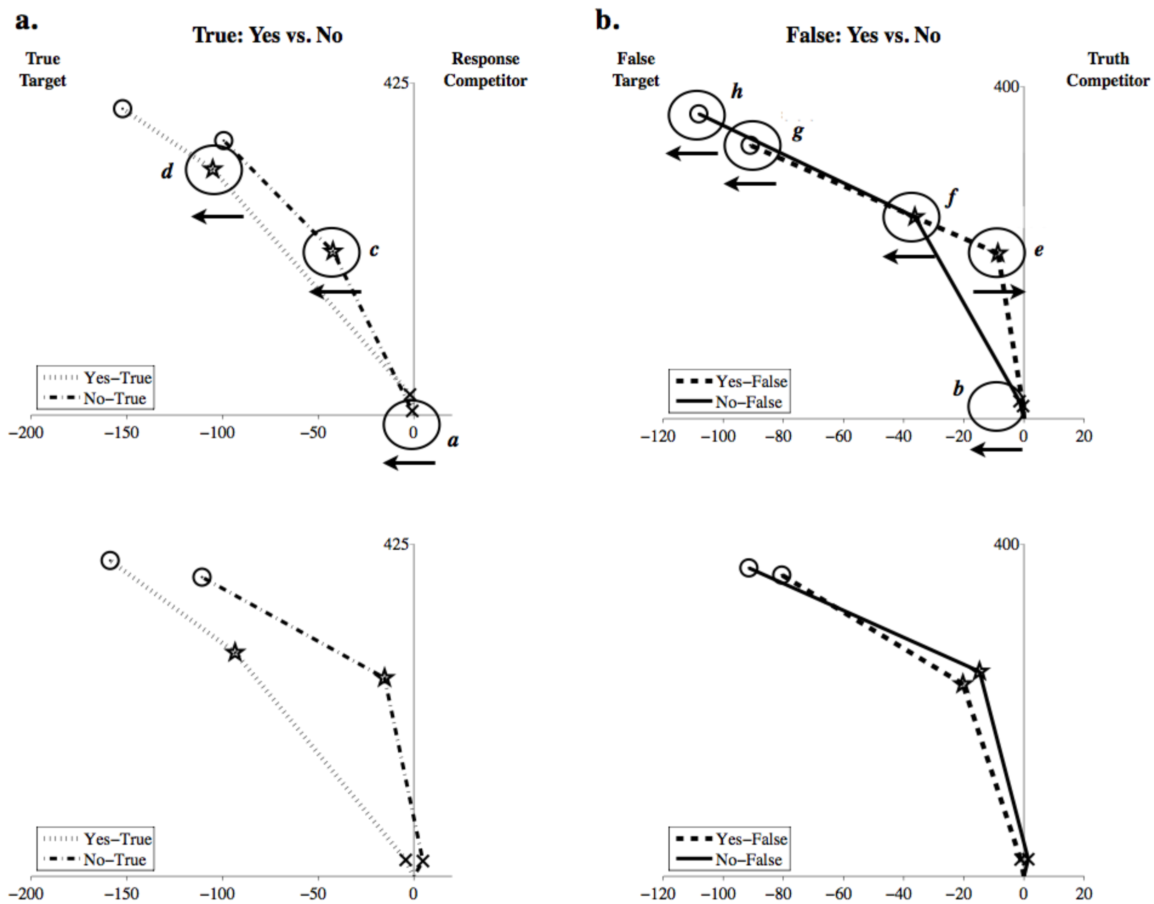

Figure 10. Modulation of attractor dynamics of trajectory movements around 500 (star), 1100 (star), and 1700 ms (circle). Using Study 1b (QUESTION-AUTO) as reference (the figures in the bottom right and left panels), circled regions indicate whether trajectories in Study 2b (QUESTION-SEMANTIC) were shifted towards or away from the critical competitor region. All circled regions show statistically significant differences. In (a), for true "no," the "yes bias" is decreased around 500 (*region a*) and 1100 ms (*region c*); and for true "yes," there is facilitation for "yes" response around 1100 ms (*region d*). In (b), greater facilitation to say "no" for false "no" responses around 500 (*region b*), 1100 (*region f*) and 1700 ms (*region h*); and for false "yes" responses, greatest interference from truth around 1100 ms (*region e*) but decreases greatly by 1700 ms (*region g*).

Similarly, turning now to "false" responses (Figure 10b), the false "no" responses also exhibited less interference from the competitor response, i.e., the "truth attractor," in the QUESTION-SEMANTIC study. Interference was minimized at every time range, from 500 to 1100 to 1700 ms (Figure 10b-*b*/*f*/*h*). Thus, it appears easier to falsely deny information when it is semantic in nature versus autobiographical. For the false "yes" responses, there was also less competition in QUESTION-SEMANTIC later in processing, around 1700 ms (Figure 10b-*g*). Here, it was easier to falsely confirm information when it is semantic versus autobiographical. However, this easier processing was not present throughout the response. At the mid-range stage of processing, the false "yes" responses showed a greater "pull" toward the competitor response option, in this case, the "NO" response (Figure 10b-*e*). With a false prompt, even when it is presented at the final word of a question, it still primes a denial response (to say "no."). This might explain why the false "no" responses were fastest, and why there is interference with the false "yes" responses early on. To interpret this finding in relationship to the QUESTION-AUTO study, it appears much more difficult to falsely deny autobiographic information than semantic information. Thus, to catch a would-be liar, signatures of increased processing difficulty should be best exhibited by questions that require a suspect to access information from autobiographic memory. For example, questions like, "What were you doing on the morning of February 12th?" might be more revealing than semantic episodic facts, like "What color was the getaway car?." However, further work needs to examine whether accessing semantic "trivia-like" knowledge is similar to semantic episodic knowledge.

Chapter 3: Discourse Analysis and Deception

Study 3a: Automated Feature Extraction with Coh-Metrix

NOTE: The following study was originally published in Applied Psycholinguistics; Duran, McCarthy, Hall, & McNamara, 2010.

Of the spinmeisters, fibbers, or equivocators among us, their success often hinges on the ability to conceal a lie with well-chosen words. However, truth's traces may still lurk amidst their verbal eloquence, as subtle linguistic features of language have been shown to reveal inner states of thought and feeling. These features go beyond the literal meaning of words and focus instead on how words are arranged and structured in discourse (Pennebaker et al., 2003). By pursuing these features, some progress has been made in uncovering the linguistic correlates of deception. The gains, though, are not without their unique challenges. Deception is a behavior designed to defeat detection, and thus identifying salient linguistic features of deception may be difficult even for the trained researcher (Vrij, Edward, Roberts, & Bull, 2000). Indeed, attempts by human judges to detect deception are fraught with problems of reliability and depth of analysis. One approach to this problem has been to turn to *Natural Language Processing* (NLP) algorithms that incorporate advances in technology and linguistic theory. At the forefront of these technologies is an application called Coh-Metrix[4] (Graesser et al., 2004). Coh-Metrix is the largest NLP tool of its kind, with over 700 indices of computed language characteristics that have been validated across a variety of psychological domains (Crossley, Louwerse, McCarthy, & McNamara, 2007; Hempelmann, Rus, Graesser, & McNamara, 2006; McCarthy, et al., 2008; McNamara, Louwerse, McCarthy, & Graesser, in press).

[4] http://cohmetrix.memphis.edu

In the current study, we take the first steps in using Coh-Metrix to identify features of deception. By doing so, we also address another challenge in the linguistic analysis of deception research. Deception occurs in a variety of settings and for a variety of purposes. Accordingly, the linguistic features relevant to one context do not necessarily hold in another context (Zhou et al., 2004). Moreover, in research conducted thus far, the linguistic features that have been identified for a particular context have not been corroborated, or even extended, with multiple NLP tools. Given the nebulous nature of deception, there is an impetus for researchers to clearly specify the context of the targeted deception, and to use convergent NLP approaches to evaluate the various types of linguistic features. Therefore, to meet these challenges, we build from prior research to compare and establish conceptual validity between NLP tools.

Specifically, we turn to the work of Hancock, Curry, Goorha and Woodworth (2008), who use a NLP tool called *Linguistic Inquiry and Word Count* (LIWC; Pennebaker et al., 2001). In their study, Hancock et al. (2008) collected transcripts of deceptive and truthful conversations that occurred within an instant-messaging (IM) environment. To evaluate their data, those transcripts were submitted to LIWC, a tool that evaluates over 70 dimensions of language. LIWC has gained a tried and trusted reputation for tracking linguistic features that are indicative of social and psychological phenomena, including personality traits (Pennebaker & King, 1999), emotional expression (Kahn, Tobin, Massey, & Anderson, 2007), and mental health (Pennebaker, Mayne, & Francis, 1997). Much like Coh-Metrix, the success of LIWC is aided by its automated and easy-to-use interface. The two NLP tools also share similarities in their ability to analyze a large number of linguistic features, preeminence in their respective literatures, and accessibility for a general audience. Moreover, the tools have a number of conceptually similar indices (i.e., computational instantiations of linguistic features) that allow for an evaluation of algorithmic validity. By comparing Coh-Metrix with LIWC, we can offer a unique, but complementary analysis that strengthens our investigation into the nature of deceptive language.

To conduct our study, we use the conversational transcripts and LIWC results from Hancock et al. (2008). We do so to provide a basis for comparison within one specific context of deception. According to Zhou et al. (2005), the features of deceptive language vary from context to context, with particular contrast within communication channels (e.g., face-to-face, telephone, email). Therefore, it becomes necessary to focus on a single communication channel to control for any changes in language use. In Hancock and colleagues' study, the context for deceptive language is expressed as computer-mediated communication (CMC) using instant messaging. These conversations occur as synchronous exchanges between two (or more) interactive participants. In recent years, this CMC channel has received greater attention because of its increased use in business and industrial settings (Andersen, 2005; Quan-Haase, Cothrel, & Wellman, 2005). Because of this increase, it is appropriate to investigate deception in CMC, which is as common, if not more so, than lies told during face-to-face conversations (Carlson, George, Burgoon, Adkins, & White, 2004). As in face-to-face conversations, deceivers using instant messaging can monitor the interaction as it occurs, but are not burdened by paralinguistic cues that might otherwise be incriminating. Although this CMC context is growing in popularity and is open to feature deception, there are few studies that explicitly address this communication channel. As such, the Hancock and colleagues transcripts offer an opportunity to further explore a promising CMC context for deceptive cues.

Another reason to revisit the Hancock et al. (2008) conversational transcripts is to place greater emphasis on the dynamics of deception in real-time conversation. Hancock and colleagues were largely motivated by the research of Burgoon and Buller (Buller & Burgoon, 1996; Burgoon, Buller, Floyd, & Grandpre, 1996; Burgoon, Buller, & Floyd, 2001), who argue that deception is concomitant to maintaining plausibility in social interaction. Deception often occurs in a dialogue between interlocutors, and as such, the linguistic features that identify deceptive competence emerge from the joint contribution of both conversational partners (*sender* and *receiver* of deceptive exchanges). Indeed, many researchers claim the mutual influence between conversational partners creates an inter-dependent relationship in language use (Clark, 1996; Pickering & Garrod, 2004). Hancock and colleagues were particularly interested in whether the receiver engaged in what Niederhoffer and Pennebaker (2002) refer to as *linguistic style matching*, whereby the receiver takes on the linguistic features of the deceptive sender.

The dynamic maintenance of conversational deception also has unique cognitive and social 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, O'Hair, Cody, & Yen, 1985); they must actively monitor the receiver's understanding in order to establish and maintain conceptual common ground (Clark & Schaefer, 1987); and senders must adjust pragmatic strategies on the fly when discussing different topics. Hancock et al.

(2008) and others (e.g. Zuckerman, 1981; Johnson, Barnhardt, & Zhu, 2004) have 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.

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 selected sets of Coh-Metrix measures that are operationalized to capture the cognitive and social influences of conversational deception.

In the section that follows, we first review the method Hancock et al. (2008) used for collecting the transcripts of deceptive and truthful conversations. We then present and provide a theoretical rationale for the Coh-Metrix measures chosen for this study. We then compare the data of Hancock and colleagues alongside our expanded approach.

Hancock et al.'s (2008) Conversational Transcripts

Participants

The original cohort of participants from the Hancock et al. (2008) study 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 who were unacquainted with each other prior to their participation in the study.

All participants were recruited under the pretense of studying *how unacquainted individuals communicate about various conversation topics*. As such, participants were not aware that deception would be required in the study. Participants' social interaction was also limited by placing each member of a pair in a separate room upon arrival at the laboratory.

Procedures

The experiment was conducted within a text-based, computer-mediated communication environment (CMC). CMC is simply using a computer interface to send message in text, video, or audio format via a computer interface. Participants were led to separate rooms and seated in front of a computer console. The instant-messaging software, *Netmeeting*, was used to collect the written communication of participants. This software allows messages to be sent instantaneously between computers, via an internet connection. Both the sender and receiver of a message enter text into a large interface text window that can be viewed easily. All messages were recorded automatically and stored anonymously.

Still in their separate rooms, participants were randomly assigned the role of *receiver* (the "deceived") or *sender* (the "deceiver") for each dyad. The sender's role was to initiate and maintain a conversation using four simple, icebreaker topics provided by the experimenter. 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 four topics were presented to the sender and receiver on a sheet of paper along with the practice topic: *When I am in a large group, I…* The practice topic allowed participants to become comfortable with one another in the experimental setting. Along with initiating the 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 topics pre-selected by the researchers, and to tell the truth on the other two topics. Specifically, they were asked *to NOT tell 'the truth, the whole truth, and nothing but the truth'*. This broad conceptualization of deception was considered to be the most naturalistic, thus giving senders some flexibility in how they chose to lie. On the sheet of paper with the experimental topics, the two topics that involved NOT telling the 'truth, the whole truth, and nothing but the truth' (i.e., to be deceptive) were signaled to the sender with an asterisk. The receivers, blind to the sender's deception, were merely instructed to stay engaged and responsive to the ongoing conversation. The receiver's sheet of paper outlining topic order had no asterisk markers. The presentation of topics, as well as the order of deception, was counterbalanced across all participant pairs.

The online interactions were automatically stored and monitored by the experimenter on a separate, third console. The experimenter's role during the conversational phase was to initiate the conversation and mediate the interaction with the practice topic. Prior to initiation, participants were allowed 5 minutes to reflect upon the topics, thus allowing senders (i.e., the deceivers) time to prepare the gist of their fabricated responses. There was no time limit to the subsequent conversation and participants were instructed to stop only when both conversational partners felt they had exhausted the topic matter. After completing all four topics, participants were introduced to each other in person and then fully debriefed.

For preparation of the data, 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: four transcripts of the *sender* dialogue and four transcripts of the *receiver* dialogue. Because two of the four topics discussed were considered *deceptive*, there were four transcripts labeled *deceptive*, two from the sender and two from the receiver (recall, however, that the receiver was not aware that the sender was being deceptive). The remaining four transcripts were labeled *truthful*.

Linguistic Features of Deception

For dependent variables, Hancock et al. (2008) used eight LIWC based linguistic indices. With LIWC, 72 different word characteristics can be tracked per written response. For each of the 72 word characteristics, LIWC provides the percentage of words that adhere to that particular characteristic. Computational algorithms in LIWC compare the content of each transcript to over 2300 words that have been coded for a variety of psychological and linguistic characteristics; including part-of-speech, emotional saliency, and cognitive complexity.

In our current study, we used the same transcripts as Hancock et al. (2008) but analyzed them with the Coh-Metrix software. Coh-Metrix was initially developed to explore cognitive constructs of cohesion in written text. Cohesion here refers to the linguistic features that explicitly link words, propositions, and events in a text; that in turn, facilitate a reader's coherent mental representation of a text. To construct a profile of cohesion, Coh-Metrix tracks word-level features that are similar to LIWC, but also incorporates modules and algorithms that assess collocations of words. 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. As such, there are over 700 linguistic indices available in Coh-Metrix. Combinations of these indices have been applied to a wide-range of domains, including the validations of coherence breaks in academic texts (Duran, Bellissens, Taylor, & McNamara, 2007; Ozuru, Best, & McNamara, 2004); discriminating low- and high-cohesion versions of academic texts (McNamara, Ozuru, Graesser, & Louwerse, 2006; McNamara, Louwerse, McCarthy, & Graesser, in press); identifying shifts in writing style between professional writers, even shifts that occurred during the careers of each respective writer (McCarthy, Lewis, Dufty, & McNamara, 2006); and evaluating the pedagogical importance of authentic and simplified texts for SLA education (Crossley, McCarthy, et al., 2007).

The current analysis is the first attempt to use Coh-Metrix to characterize linguistic patterns of conversational deception. However, using over 700 linguistic indices presents two major theoretical problems. One problem is that spurious distinctions are likely to arise when there is an excess of variables. Too many variables can result in a statistical “over-fitting,” such that small and largely irrelevant differences between deceptive and truthful conditions may be exaggerated. The second problem of using the full set of linguistic indices is the overwhelming task of establishing each index’s explanatory power. Before a specific index is used, it should be justified by a general framework of deception; however, no such framework exists (that we are aware of) because deceptive linguistic behavior is highly flexible with different external (e.g., social) and internal (e.g., cognitive) influences (DePaulo et al., 2003). As such, it becomes necessary to first consider the conversational context in which the deception is embedded and only then select linguistic indices that are most relevant to that particular context. For example, it is reasonable to assume that deceptive behavior in a casual conversation will be very different from deceptive behavior in a criminal interrogation. Accordingly, our selection of Coh-Metrix indices was guided by many of the principles of deception established in Hancock et al. (2008) and elsewhere in the deception and communication literature (Burgoon, Buller, Floyd, et al., 1996; Zhou et al., 2005). These principles are based on the cognitive and social influences that are hypothesized to arise during deceptive behavior. Ultimately, we operationalized the linguistic indices in six categorical constructs that will be explained in further detail later in this article: (a) Quantity, (b) Immediacy, (c) Specificity, (d) Accessibility, (e) Complexity, and (f) Redundancy.

Each category above is represented by 2 to 3 Coh-Metrix indices that were chosen to provide converging validity, one of the explicit goals of our research. At least one of these indices was selected to be conceptually similar to a LIWC index. These similar indices may seem trivially redundant; however, they provide a basis for comparison with Hancock et al. (2008) and for establishing simple measurement reliability. Unfortunately, several categories do not have a representative and/or a conceptually similar LIWC index. These omissions are addressed in turn.

We proceed by briefly explaining the theoretical motivation for each of our 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). As in Hancock and colleagues' work, the Coh-Metrix data are analyzed in a 2 (message type: deceptive vs. truthful) x 2 (speaker type: sender vs. receiver) repeated measures type General Linear Model (GLM) procedure. We additionally provide partial eta squared values to assess the strength of any significant effects.

This analytic method not only allows us to examine the differences between deceptive and truthful conversations but also allows us to examine the differences between sender and receiver. As mentioned earlier, the receiver might exhibit a pattern of linguistic style matching with the sender. Alternatively, the sender's behavior may elicit a subtle, but unique pattern of linguistic behavior in the receiver. For these reasons, it is theoretically important to consider the linguistic profiles of both conversational partners in deceptive exchanges.

Table 9. *Categories of deceptive behavior based on linguistic features operationalized by Coh-Metrix.*

Classification	Definition
Quantity	
Total word count†	Total words in text (based on Charniak parser)
Words per conversational turn†	Mean words per sentence
Immediacy	
Tentative statements††	Modal verbs (e.g., *should*, *might*, *may*)
Personal pronouns†	e.g., *I*, *me*, *he*, *they*
Specificity	
Spatial††	Locational prepositions (e.g., *here*)
Temporal††	Ratio of temporal elements
Questions†	Incidence of wh- adverbs (e.g., *why, what*)
Accessibility	
Familiarity of words††	Word rating from MRC database
Meaningfulness of words††	Word rating from MRC database
Concreteness of words††	Word rating from MRC database
Complexity	
Negation†	Negation connectives (e.g., *did not*, *except*, *but*)
Sentential complexity††	Mean words before main verb of main clause

†: Linguistic cue is an approximate replication of Hancock et al. (2008)
††: Linguistic cue is novel to current study

Table 9. *Categories of deceptive behavior based on linguistic features operationalized by Coh-Metrix. (continued)*

Classification	Definition
Redundancy	
Given information††	LSA given/new value
Referential overlap††	Argument word overlap, adjacent sentences

†: Linguistic cue is an approximate replication of Hancock et al. (2008)
††: Linguistic cue is novel to current study

Coh-Metrix Results and LIWC Comparison

Quantity

In both Hancock et al. (2008) and the current study, the *total word count* and *number of words per conversational turn* were computed and compared between deceptive and truthful conversation transcripts. These indices are 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, & Memon, 2002). As such, senders' total word count and number of words per conversation turn should be significantly lower 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.

In the current Coh-Metrix analysis, a significant main effect of message type (deceptive vs. truthful) was observed for total word count, $F(1, 33) = 8.87$, $p = .005$, partial eta squared[5] = .21. More words were produced during deceptive conversation (*M* = 159.38, *SE*[6] = 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. In neither study was there an effect for speaker type (sender vs. receiver), nor did the total word count between message types differ across speakers (i.e., there was no interaction between message type and speaker type).

The second quantity analysis was on the mean number of words per conversational turn. Using Coh-Metrix, a significant main effect for message type was observed, $F(1, 33) = 3.50$, $p = .05$, partial eta squared = .10. Fewer words were produced per conversational turn in the deceptive conversations (*M* = 7.73, *SE* = 0.27) than per truthful turn (*M* = 8.37, *SE* = 0.36). Senders produced fewer words per conversational turn when deceptive (*M* = 7.98, *SE* = 0.42) compared to telling the truth (*M* = 8.19, *SE* = 0.55), and receivers produced fewer words per conversational turn in the deceptive conversations (*M* = 7.48, *SE* = 0.35) than per truthful turn (*M* = 8.55, *SE* = 0.55). Taking this result in conjunction with the previous total word count results, the Coh-Metrix analysis demonstrates that senders and receivers in deceptive conversations use more

---

[5] To interpret the partial eta squared values, Stevens (2002) suggests the following: .01 is considered a small effect, .06 is considered a medium effect, and .14 is considered a large effect. However, the reader is reminded that such interpretations are merely "guides", and the importance of any effect size is always relative to the task at hand.

[6] We report standard errors in this study to be consistent with Hancock et al's. (2008) results section

words overall, but fewer words per conversational turn. However, this conclusion does not hold for Hancock et al. (2008). In their analysis, they did not find an equivalent decrease in words per conversational turn for senders and receivers in deceptive conversations. Rather, Hancock and colleagues report a marginally significant interaction (two-tailed, $p = .06$) indicating that only receivers used fewer words per conversational turn in deceptive conversations.

The incongruent conclusions between the two computational tools may have resulted from implemented differences of what LIWC and Coh-Metrix consider a word. LIWC simply computes as a word any sequence of alphanumeric characters that is separated by a white space from another sequence. Coh-Metrix, however, computes words on the basis of the Charniak syntactic parser and corresponding word tags. As a result, the definition of a word is more precise. Contractions, for example, are counted in the expanded form (e.g., *don't* > *do not*, *they're* > *they are*). A more relevant difference is that an ellipsis is counted as a distinct pause filler. LIWC would treat *so…* as one word, whereas Coh-Metrix would output two words by distinguishing *so* and the ellipsis. This specificity is important for the current study where pause fillers are believed to hold semantic content.

Based on overall word counts between deceptive and truthful conditions, Coh-Metrix counts 2.6 more words on average per deceptive conversational turn and 0.42 more words on average for truthful conversational turn when compared to LIWC's counts. This comparison suggests that Coh-Metrix distinguishes more word types, and that this precision particularly affects the interpretation of the utterance length measurement.

Immediacy

Introducing deception into a conversation always carries the risk of detection. Although the consequences might be no more than slight embarrassment, deceivers may take cautionary measures to distance themselves from their lies, even while engaged in the act of lying. Wiener and Mehrabia (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* (e.g., *I*, *me*, *ours*; Newman et al., 2003). 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 the deceiver.

Another distancing strategy is an increased use of *tentative constructions* with words and phrases like *might*, *would*, *I guess*, *it seems to me*. These are often referred to as *hedges*. Tentative constructions imply a noncommittal to the content of the lie, thereby mitigating negative judgment of personal character or attributions of blame (Vrij & Heaven, 1999).

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. The researchers found a statistically significant main effect for speaker type (sender vs. receiver) for third person pronouns, as well as an interaction between message type and speaker type for third person pronouns. The main effect provides evidence that senders use more third person pronouns than receivers; but more importantly, the interaction reveals that it is only the deceptive senders who are more likely to discuss others in the third person.

We also used Coh-Metrix to assess pronoun use as a distancing strategy. To do so, we simply computed the percentage of the different pronouns in each conversational transcript. Much like Hancock et al. (2008), we did not find any statistically significant effects for first and second person pronoun use. However, our results for third person pronouns differed from Hancock and colleagues' results. We found a main effect for speaker type and the interaction, with marginally significant values, $F(1, 33) = 3.84$, $p = .06$, partial eta squared = .10 and $F(1, 33) = 3.20$, $p = .08$, partial eta squared = .09, respectively. Nonetheless we found the same the trend, $F(1, 33) = 5.73$, $p = .02$, partial eta squared = .15, showing senders using more third persons pronouns during deception ($M = 2.93$, $SE = 0.32$) than the truth ($M = 1.94$, $SE = 0.22$). The statistical differences here are most likely explained by differences in word count when computing percentages.

In a second immediacy analysis, we used Coh-Metrix to evaluate the distancing strategy of increased tentative construction phrases. There is no equivalent analysis in Hancock et al. (2008). The current approach underscores the advantages of using a syntactic parser and part-of-speech tagger. With these additional modules, an incidence score (out of 1000 words) for modal verbs (e.g., *should*, *might*, *may*) can be computed. Despite these noted advantages, the Coh-Metrix index of tentative constructions via modal use was not statistically significant. The Coh-Metrix index may have been too general to make subtle distinctions. Coh-Metrix does not distinguish among different uses of modals. Consequently, all modals were included in the computation - even modals that are non-tentative in nature. For example, the root use of *may* and *must* produces a non-tentative use in statements like *You must go now* or *You may not*. Taken together, the non-specific modal index was too general to support the immediacy category.

Specificity

Language has many linguistic features that allow speakers to reconstruct events from memory with certain temporal and spatial characteristics. The reconstructed events are often isomorphic mappings to perceived external events or, as is the case with deception, fabrications generated from internal cognitive processes of imagination and reasoning. As such, the mental representation of each event differs in terms of origin; the event can be initially encoded as a perceptual experience or as a simulation of an imagined experience. According to Reality Monitoring theory (Johnson & 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. To continue with our goal of automatically cataloging the linguistic patterns of deceptive and truthful speech in conversation, we chose two Coh-Metrix indices that capture the linguistic features of temporal and spatial characteristics. The temporal features index tracks words that have a high probability of being embedded in temporal expressions. These words include specifiers (e.g., *next*, *following*), deictics (e.g., *yesterday*, *now*), absolutes (e.g., *1997*, *Monday*), time of day (e.g., *12:00 AM*, *noon*), and time periods (e.g., *summer*, *week*). The 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*) is 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-words (e.g., *why*, *what*) to assess possible changes in receivers question asking behavior.

The first specificity analysis using Coh-Metrix indices of temporal and spatial specificity was not statistically significant. However, for the Coh-Metrix index of general specificity, there was a significant interaction between message type and speaker type for number of wh-adverbs used, $F(1, 33) = 6.83$, $p = .01$, partial eta squared = .17. 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 ($M = 9.04$, $SE = 1.09$), $F(1, 33) = 4.19$, $p = .05$, partial eta squared = .11; conversely, receivers used marginally more wh-adverbs when being deceived ($M = 10.34$, $SE = 1.23$) than when told the truth ($M = 7.33$, $SE = 1.02$), $F(1, 33) = 3.30$, $p = .08$, partial eta squared = .09. These patterns of results suggest that receivers ask more questions when being deceived, while senders ask fewer questions when being deceptive. In Hancock et al. (2008), they too found the same effect for the receiver, but failed to find a similar effect for the sender. Again, the incongruence might be attributed to differences in the computational approach for operationalizing question use (i.e., proportion of wh-adverbs vs. percentage of question marks).

Accessibility

We hypothesized that deceivers would select vocabulary that is easier to retrieve from memory. Based on the seminal work of Paivio (1965) and Underwood and Schulz (1960), word retrieval accessibility is modulated by experiential influences of word meaningfulness, familiarity, and concreteness. 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 and is typically assessed on a Likert-type scale from 1 - 7. More familiar words are more likely to 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. For word meaningfulness and familiarity, Coh-Metrix provides an average score based on human ratings of over 150,000 words compiled in the MRC database (Coltheart, 1981). For word concreteness, Coh-Metrix computes abstractness and ambiguity scores by incorporating a module based upon WordNet (Miller, 1995). WordNet is an online lexicon tool that groups words into sets of synonyms that are connected by semantic relations. One such relationship, the hypernym value, refers to the number of levels a word has above it in a conceptual, taxonomic hierarchy. A high hypernym value is a proxy for word concreteness because the word has many distinctive features.

All indices for the accessibility category are computed as incidence scores in Coh-Metrix. There are no equivalent indices for accessibility in Hancock et al. (2008). There was a statistically significant main effect of message type for word meaningfulness in conversations, $F(1, 33) = 7.88$, $p = .008$, partial eta squared = .19. 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, $F(1, 33) = 5.42$, $p = .02$, partial eta squared = .14. 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), $F(1, 33) = 3.25$, $p = .05$, partial eta squared = .09. 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.

In summary, senders and receivers used more meaningful words when being deceptive, with the deceptive sender specifically using words that are more concrete. As we suggested earlier, these word characteristics facilitate the activation and retrieval of semantic meaning from memory. A consequence of this facilitation is that meaningful and concrete words are more likely to be used if cognitive resources are directed elsewhere (e.g., in concocting a deceptive message during conversation). Thus, the increased use of meaningful and concrete words by deceptive speakers supports our earlier hypothesis that deception places greater demands on cognitive resources.

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 index of sentence complexity is the number of words before the main verb of the main clause. It is assumed that as the number of words increases, so does the demand on the speaker's working memory (see Graesser, Zhiqiang, Louwerse, & Daniel, 2006). Assuming that the process of lying would tax a deceiver's memory resources, we can expect a decrease in words before the main verb (i.e., lower complexity) compared to the truth-telling condition.

Alternatively, we could also hypothesize that an increased number of words before the main verb are to be expected in conversational contexts where deceptive messages are created on the fly. An increase in words before the main verb would reveal a *stalling* strategy used to formulate a lie while still staying engaged in the conversation.

Coh-Metrix computes the main verb of each sentence by first automatically parsing each sentence using the Charniak parser (1997, 2000). Each parse generates a syntactic tree that represents the underlying formal grammar. From this formal representation, the main verb of the main clause is identified and preceding words are tallied. The sentential complexity for deception and truth-telling is then assessed by collapsing the sentences of each conversational transcript into a mean score. There is not an equivalent index in Hancock et al. (2008).

A significant main effect of message type was observed for this complexity measure, $F(1, 33) = 5.63$, $p = .02$, partial eta squared = .15. More words were used before the main verb in deceptive conversations ($M = 7.14$, $SE = .46$) than in truthful conversations ($M = 5.79$, $SE = .37$). Specifically, senders use more words before the main verb when deceptive ($M = 6.79$, $SE = .71$) than when telling the truth ($M = 6.16$, $SE = .61$. Likewise, 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.

These results suggest that senders and receivers use more syntactically complex sentences in deceptive conversations. Increased sentence complexity does not support the hypothesis that complexity results from working memory demands, but rather supports the alternative hypothesis that generating deception in spontaneous conversation requires a stalling strategy. For working memory to be the prevailing factor, senders must know exactly what they want to say before they say it. It is only under these circumstances that a sender will intentionally minimize the use of words before the main verb. This active strategy of advanced planning is unlikely in the current conversational context.

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 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. Although negative connectives help clarify event depictions, the speaker must also recall additional detail from memory. Deceptive speakers must conjure that detail up at that moment. As such, deceptive speakers may 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.

The Coh-Metrix index of negation connectives is a proportion value computed from the Charniak syntactic parser and part-of-speech taggers. The LIWC index uses the pre-defined word list and computes the value as a percentage.

This measure of complexity is computed similarly for LIWC and Coh-Metrix and is also assumed to reflect demands on working memory. Our results agree with those reported by Hancock et al. (2008) that there are no statistically significant effects for negation connectives.

Redundancy

In both deceptive and truthful conversations, an important component of event narration is the coherence of statements and ideas. Coherence is a psychological interpretation of comprehension. The greater the coherence, the easier the narration will be to understand (Graesser, McNamara, & Louwerse, 2003). Coherence is modulated by various factors, but a crucial factor is cohesion - the explicit language used to connect information and provide conceptual consistency. Most cohesion research suggests that text cohesion influences text comprehension, particularly with texts that consist of formal written monologues (Beck, McKeown, Sinatra, & Loxterman, 1991; McNamara, Kintsch, Songer, & Kintsch, 1996), but little work has been conducted on the relationship between cohesion and coherence in informal spoken dialogue. 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.

It could be possible, however, that incoherent mental representations are not mirrored in speech, but instead, the difficulty of remembering and structuring spontaneous deception may promote simpler and more cohesive speech. Characteristics of such language include conceptual redundancy and more accessible words (Duran et al., 2007). We have already demonstrated in this study that deceivers tend to use more accessible words (e.g., high concreteness; high meaningfulness). It may be the case that deceivers also capitalize on conceptual redundancy for greater cohesion.

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 (McNamara et al., 2006) and LSA given/new values (Hempelmann, 2005). 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. The LSA given/new values also compute overlap between sentences, but it requires more explanation to understand how it works.

The LSA given/new value is based on Latent Semantic Analysis (LSA; Landauer, McNamara, Dennis, & Kintsch, 2007). This measure compares adjacent sentences to determine if the meaning in a target sentence is new (different) or given (redundant) to preceding sentences. Sentence meaning is first computed by representing each word in the sentences as a distributional pattern of frequency occurrences within a large corpus of texts (representation is in vector format). Words that have similar patterns of occurrences are considered similar in meaning. Word similarity vectors are then combined linearly into a composite meaning vector. The target vector is projected into a hyperplane constructed from all preceding composite meaning vectors and based on the target sentences relationship to the hyperplane and a *new* or *given* value is generated (see McCarthy et al., in press, for more information). High values on both the argument overlap and LSA given/new values suggest high cohesion between-sentences. These measures are unique to Coh-Metrix; there is no equivalent in Hancock et al. (2008).

For the first 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, $F(1, 33) = 9.32$, $p = .004$, partial eta squared = .22. 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 ($M = .22$, $SE = .01$).

These results provide evidence that deceptive conversations contain more given information relative to preceding context. This result should be expected if we consider an important goal for deceivers is to minimize opportunities for self-incrimination. A strategy to avoid self-incrimination may be to reiterate particular topics or themes in the conversation. Deceivers do not reiterate by explicit repetition, as evidenced by the null finding with referential overlap, but by an implicit focus on a few semantic focal points. However, there may be no conscious decision to avoid self-incrimination. Instead, the high LSA given/new values support a hypothesis that redundancy strategies are triggered by differences between memory representations of deceptive and truthful narratives. For example, it is possible that the details of truthful events are more extensively linked in memory than the fabricated details of a lie. As a truthful account unfolds, the activation and recall of remembered details are likely to activate other details in a distributed and global manner, thus a greater variety of information is available for use. Conversely, the details in a deceptive account are often constructed and cued from the local and developing context. As such, there is less new information activated from memory and deceivers may default to more redundant language.

Brief Summary of Coh-Metrix Analysis

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 (due to a *stalling* hypothesis), and (d) less unique information introduced during the course of the conversation.

The effects we have discussed so far changed in the same direction for both sender and receiver. However, other changes in linguistic behavior were specific to either the sender or receiver. For example, personal pronouns and word concreteness increased only for senders while they were being deceptive. Demonstrating another pattern, receivers asked marginally more questions in deceptive conversations than the senders who asked fewer questions.

Brief Summary of Coh-Metrix and LIWC Comparison

Table 10 provides a side-by-side comparison of the results using the indices that were similar in Coh-Metrix and LIWC. While these indices are not exact replications due to differences in algorithmic operationalization, they are quantifications of the same linguistic features. Overall, five indices were comparable, and of these five, *total word count*, *negation,* and *personal pronouns* had the same result. This convergence confirms that more words are used in deceptive conversations, that there are no differences in the use of negation, and that deceptive senders use more third person pronouns. 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; the difference in the indices most likely results from the different definitions of a word used by the two tools. 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. LIWC showed only that the receivers asked fewer questions during deception. In general, for both of the non-converging indices, the Coh-Metrix analysis found a statistically significant effect that was not found in the LIWC analysis.

Table 10. *Comparison of similar Coh-Metrix and LIWC index results in conversational transcripts.*

	Coh-Metrix	LIWC
Total word count	More words overall in deceptive conversations	More words overall in deceptive conversation
Words per conversational turn	Fewer words per conversational turn in deceptive conversations	Receivers used *marginally* fewer words per conversational turn than senders in deceptive conversation
Personal pronouns	Senders used *marginally* more 3^{rd}-person pronouns when deceptive as compared to when telling the truth	Senders used more 3^{rd}-person pronouns when deceptive as compared to when telling the truth
Questions	Receivers ask more questions during deceptive conversations, senders fewer	Receivers ask more questions during deceptive conversations
Negation	None	None

General Discussion

Both this study and Hancock et al. (2008) demonstrate that at least one type of deception is detectable through natural language processing (NLP) tools. For our analysis, we compared the Coh-Metrix and LIWC tools on the original corpus of deceptive conversations used by Hancock and colleagues. Using this 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. In this discussion, we first turn to the expanded analysis and the identification of eight Coh-Metrix indices that distinguish deceptive conversations from truthful conversations. Using this winnowed set of indices, we provide new insights into the cognitive and social constraints that are hypothesized to influence deceptive behavior – in both the deceiver and their naïve conversational partner. Turning next to the comparison with Hancock and colleagues, we discuss complimentary insights provided by the LIWC analysis. In particular, we consider the findings of Coh-Metrix and LIWC within the context of computer-mediated communication (CMC). Throughout this discussion, we address the limitations of our current research and end with suggestions for future work.

There is a well-established conversational Maxim of Quality that a speaker should avoid saying what the speaker knows to be false (Grice, 1975). When lying to a friend, colleague, or foe, a speaker often violates this maxim, and as a consequence, new goals and task demands are introduced into the conversation. The deceiver must now maintain representations of both the truth and a falsified version of that truth. In doing so, the deceiver must also appear convincing while avoiding unintentional "slips" of the truth. The cumulative effect is that deception requires increased cognitive control in the presence of social scrutiny. Previous research suggests that even with the best attempts to maintain control, the inner states brought on by deception are manifested in subtle changes of language use (Pennebaker et al., 2003).

For the Coh-Metrix analysis, we created six theoretically guided categories to represent these changes in language. Each category is composed of two or three Coh-

Metrix indices. The categories include a) the amount of information in the conversation (i.e., *quantity*), b) the readiness to identify with message content (i.e., *immediacy*), c) the breadth of detail used to describe a narrative (i.e., *specificity*), d) the change in semantic memory retrieval (i.e., *accessibility*), e) the change in grammatical phrasing (i.e., *complexity*), and f) the repetition of given information (i.e., *redundancy*). We then compared truthful and deceptive conversations for changes in the six categorical dimensions. We found statistically significant results for all categories.

Several of our findings provide novel contributions to the relationships between deception and language. A key discovery is that *quantity* of word use changes for the level of analysis. For example, in deceptive conversations, fewer words were used at the level of conversational turn. Based on this finding alone, we might conclude that deceivers use fewer words to minimize opportunities for incrimination; however, in the same conversations, there are also more conversational turns and more words used overall. This result challenges the original conclusion and suggests that the deceivers are attempting to establish rapport with their conversational partner. Because our results also show that receivers ask more questions of deceptive senders, an alternate interpretation might be that the deceivers do, indeed, use fewer words per conversation turn to minimize opportunities for incrimination; however, the 'paucity' of information in these restricted turns require the receivers to ask for additional information or clarification which then generates more overall turns and words. Unfortunately, there is not enough information to make a conclusive interpretation either way. However, the results do highlight the rich interplay between the often-conflicting goals of cautiously limiting information and the appearance of affability.

Another new discovery is that the words used in deceptive conversations are more meaningful than those used in truthful conversations; for the sender in particular, the words are also more concrete. The *accessibility* of meaningful and concrete words from semantic memory indicates that the deceiver is using an unconscious strategy to decrease burdens on cognitive processing. Because meaningful and concrete words are highly associated to other words in semantic memory, these words are easier to retrieve, and in turn, allow cognitive resources to be redirected to the more difficult task of maintaining deception in conversation.

Related to increased difficulty, we also found evidence for *redundancy* in deceptive conversations. The redundancy is the repetition of content from contiguous utterances. Previous research that has investigated linguistic features of redundancy has failed to find significant effects because it applied a strict lexical overlap criterion (e.g., Zhou et al., 2005). Instead, redundancy in deception appears to be more subtle. In our analysis, we used an algorithm that compares the semantic similarity of two words based on their likelihood to appear in similar contexts. This algorithm, called the LSA given/new value, revealed that words similar in meaning are used more often in deceptive conversations than in truthful conversations. This redundancy in meaning suggests the deceiver may find it simpler to focus on consistent themes. Part of the reason for such focus is that the deceiver may have difficulty in using the same interconnected memory representations that are formed with real experiences. Instead, deceivers rely more on local conversational cues for what information can or cannot be reasonably fabricated. This orientation towards local context decreases the likelihood of using novel information and increases the chances of repeating what has already been stated.

Deceptive conversations are also characterized by a change in the *complexity* of grammatical constructions. A complex sentence is defined in Coh-Metrix as having more words before the main verb of the main clause. In deceptive conversations, we found that this type of sentence complexity increases. It is important to note that complex grammatical constructions identified by Coh-Metrix are not necessarily more difficult to produce, and in fact, may be preferred when attempting to generate a spontaneous lie. For instance, consider a lie about what you did yesterday. If you were telling the lie in conversation, it might take some time to think of a false response, such as *I watched TV at my house.* While constructing the response, it would be useful to buy some time with a stalling strategy that provides genuine information. Thus, you begin with *It was really cold outside…* and continue with the lie *...so, I'd thought I'd stay in and watch TV.* By doing so, there would be a higher occurrence of words before the main verb, and as such, greater evidence for our hypothesized stalling strategy.

Finally, for *specificity* of the deceptive narrative, we found that deceptive conversations were marked by the receiver asking the sender more questions. This result implies that the sender lacked specificity and that the receiver was requesting greater clarification. During these exchanges, the deceptive sender also asked fewer questions compared to when they were telling the truth. Again, these findings can be interpreted in multiple ways. The receiver may ask more questions because of an unconscious suspicion of the sender's deception. Likewise, the deceptive sender may ask fewer questions to defend against the suspicion. It may also be that because deceivers use fewer words in each conversational turn, receivers need to ask for more clarification. The receivers may be responding to a perceived violation of the Maxim of Quantity rather than the Maxim of Quality that deals with truthfulness (Grice, 1975). In either interpretation, the important finding is that receiver linguistic behavior systematically varies from that of the sender in terms of specificity.

For most of the analyses, receiver behavior was mostly aligned with the deceptive sender, with the exception of wh-adverbs, concreteness, and third person pronouns. The alignment of linguistic features is not uncommon between conversational partners. There is extensive research that shows implicit alignment can occur and cut across lexical, syntactic, and conceptual levels (Garrod & Anderson, 1987; Pickering & Garrod, 2004). The underlying mechanism is due to *priming*, whereby the linguistic features used by one partner elicit a similar representation in the other. In this way, coordination of form and meaning is automatically generated and maintained. In the current study, we find evidence for alignment in the number of words used, the meaningfulness of words, the repetition of similar words and concepts, and the complexity of grammatical constructions. A possible limitation in the alignment is not knowing whether the sender or the receiver is predominantly priming or being primed. The limitation is a particular concern because we are interested in the linguistic features generated by the deceptive sender. As such, we assume that it is the deceptive sender's linguistic behavior that is most influential. We base this assumption on two factors. First, the design of the experiment gives the sender more control by allowing the sender to introduce new topics (total = 4) into the conversation. Second, being deceptive may invoke a greater desire for the sender to be convincing, where an equivalent desire is not present in the receiver. As a result, this unique desire may translate into greater linguistic influence.

In detecting linguistic features of deception, the problem of who influences whom is heightened in conversational interactions. Unlike monologues or scripted interviews, there are cognitive and social constraints that present additional and novel challenges. Moreover, our use of a CMC corpus of deceptive and truthful conversations adds to these challenges. Despite the increasing difficulty, the CMC conversational context is an ecologically important domain that is gaining in popularity and use. However, we must be careful in generalizing our findings from the CMC context to other domains with their own constraints. Face-to-face conversations, for example, are not the same thing as instant-messaging conversations, and thus the linguistic features characterizing each conversation may substantially differ. For these reasons, we felt justified in using and extending the Hancock et al. (2008) study. Importantly, their data provided a common context to compare and contrast Coh-Metrix with LIWC.

Our first step in the comparison was simply to assess the degree to which the systems differed in their analysis of deception. Our results suggest that Coh-Metrix was largely able to reproduce LIWC results (e.g., in areas of *quantity* and *immediacy*) and offer many areas of deception detection in addition to LIWC (e.g. *accessibility*, *complexity*, and *redundancy*). For these reproduced results, the replication occurred despite two different computational approaches for operationalization. The alignment gives greater credence to the original findings in the Hancock et al. (2008) study, specifically their findings that more words are used in deceptive conversations and that deceptive senders project the focus of conversation onto others (as evidenced by the greater use of third person pronouns).

Our study also showed where LIWC and Coh-Metrix were not able to reproduce the same results on similar indices, namely, *words per conversational turn* and *questions*. For Coh-Metrix, deceptive conversations were marked by fewer words from both the sender and receiver, as well as more questions from the receiver and fewer questions from the sender. In contrast, LIWC did not find a difference of word use for senders and found only a marginal difference for the receiver. In addition, no difference in question use was found for senders. This inability to reproduce the same results using the identical corpus might suggest that one NLP tool is superior to the other. However, we take an alternative perspective. The algorithmic operationalization for each tool is a matter of preference that should be chosen to best address a research question. In other words, the operationalization does not capture a "truer" representation of reality. More than anything, the operationalization is a manifestation of computational expediency. For example, LIWC uses computationally inexpensive algorithms to process texts. During processing, words are identified by surrounding white space and matched to an internal set of words that are coded for linguistic and psychological features. In contrast, Coh-Metrix goes beyond a predefined set of words and incorporates sophisticated algorithms to maximize the scope of analysis. By including syntactic parsers and psycholinguistic databases, linguistic features can be distinguished at the word, sentence, and discourse levels. To understand this in practice, we consider the operationalization of word count. Words are not separated by white spaces alone (as in LIWC), but are expanded from contraction form (e.g., *don't* > *do not*) and distinguished from a trailing ellipsis to create unique entries for ellipsis occurrence. Furthermore, for the operationalization of questions, instead of counting the number of question marks (as in LIWC), detailed part of speech information, like wh-adverbs (e.g., *where*, *what*), can be used as an index of question use. Based on these differences in operationalization and given the current task, the Coh-Metrix analysis may have an advantage over LIWC. Because the data are typed conversations, there are a large number of ellipsis occurrences that might be an important linguistic feature of deception (e.g., pauses, incomplete thoughts). In addition, participants tend to use multiple question marks at the end of a sentence. By just counting question marks, there is a risk for over exaggerating the number of questions. This miscount is not a problem with wh-adverbs.

For the current task, LIWC does have an important advantage over Coh-Metrix. LIWC codes words for psychological dimensions, such as sensory information like “see”, “touch”, and “listen”, that may be related to a deceiver’s goals of convincing story-telling. Indeed, in Hancock et al. (2008), the deceptive conversations were reported as having a greater degree of these sensory words. Future work will require adapting Coh-Metrix to include linguistic features that have been successful in detecting deception in multiple contexts. Other candidates include positive and negative word connotations, as well as content word diversity measures (Zhou, et al., 2005). Additional work also needs to be conducted in predicting the likelihood that a narrative is deceptive or truthful. Given our established set of deceptive linguistic features, we can include these features as variables into statistical prediction models (e.g., logistic regression, discriminant function analysis). By doing so, we can also evaluate how well our linguistic features collected in a CMC context explain the variance in other interactions, such as business negotiations or even police interrogations.

Finally, it would be naïve for us to argue that a straight and easy road lies between identifying linguistic features of deception and using them in real-world practice. There are many individual differences to account for, as well as the consideration of ethical and legal concerns. Nevertheless, we begin the journey with the current study. We have shown that deception is a feature of language that is identifiable through many variables, established that Coh-Metrix is a computational system that can identify deception, and revealed that there is insight to gain by comparing computational NLP tools.

Study 3b: Automated Phrasal Analysis with Gramulator

NOTE: The following study was prepared for journal publication; Duran & McCarthy (2011).

Of all human behaviors that are considered to breach conventions of social and communicative interaction, deception is one of the most pervasive and by far the most elusive. Deception is a violation of what is known to be true for the purpose of providing misleading, but seemingly trustworthy information (Ekman, 1997). To succeed, the deception must be covert and is thus designed to thwart detection. Yet, despite the potential social risks involved (e.g., *face loss*: Brown, 1977), deception is surprisingly common in everyday interactions (DePaulo et al., 1996). But why would a speaker be so brazen as to use deception? Personal gain notwithstanding, the risks of deception are mitigated by the simple fact that deception usually goes undetected or is excused as hyperbole. From the “tall tales” that people choose to tell to their fishing buddies, to the excuses they use to get out of work, deceivers are generally believed without reproach. Even when the potential risks are increased, or when the lie strongly deviates from the truth, detection rates are still little better than chance (Feeley & deTurck, 1995; Vrij et al., 2000). One of the reasons for poor detection is that humans come equipped with a truth-bias, whereby all statements are initially assumed to be true (Gilbert, 1991; Levine & McCornack, 1991). Researchers have attempted to overcome this truth-bias by explicitly training people to look for "leakage" cues that are expressed in a deceiver's actions, such as facial movements and body posture (Vrij, 2001); or in their language output, such as in the vividness of spatiotemporal descriptions or number of verbal hedges (Johnson & Raye, 1981; Sporer, 2001). However, even when people are trained in these various techniques, their performance is still too inconsistent for real-world applicability (Bond & DePaulo, 2006; Vrij et al, 2000). Although this poor performance might be attributed to techniques that are theoretically misguided, a more likely account is that the grain-size of leakage cues is outside the normal processing abilities of trained and novice judges. For this reason, many researchers interested in detecting deception have turned to computational techniques that are unbiased and better suited to detect hidden linguistic patterns.

The current studies build from the published work of Hancock et al. (2008) and Duran et al. (2010) and their respective computational textual analyses of the linguistic features characterizing the deceptive and truthful conversations of native English speakers. We also build from Newman et al. (2003) and their linguistic analysis of arguments (both truthful and lying) on personal beliefs about abortion. In these aforementioned studies, as is true in other computational work (e.g., Zhou et al., 2004), the focus is on the stylistic organization of language; that is, the abstract linguistic properties that exist at the word, sentence, and discourse level. These properties can include the number of words of a certain grammatical category (e.g., *percentage of articles*), or they can measure the number of words that repeat across locally distributed sentences (e.g., *proportion of referential overlap*). One of the advantages of a computational approach is that the analysis does not depend on variations in topical or thematic content because the computation is neutral to context. Thus, the output is generated more rapidly than a qualitative or narrative analysis that requires subjective interpretation (Reissman, 1993).

Although computational research has certainly provided a great deal of insight into deception, many content-analytic questions remain unaddressed. These questions include: what are the topics that people tend to lie about? How does a particular phrasing reflect the cognitive, social, and motivational biases involved in deception? Such questions are important because they promise to expose content that taps the psychological processes of lie-telling, from recurring story elements that offset the processing difficulty associated with deception, to tell-tale themes that arise when the contents of long-term memory are fabricated.

To address these issues within an automated natural language processing approach, we introduce a computational tool called the *Gramulator* (McCarthy, Watanabe, & Lamkin, in press). This tool provides a numeric representation of relevant *qualitative* content – content that consist of short sequences of text (up to four words) that are more probable in one corpus compared to another. With these features in hand, we can go back to the text in which the features were found and draw conclusions based on how they were used in context. By doing so, a richly detailed characterization of deceptive language can be offered.

In the sections that follow, we present our attempt to develop an automated analysis that respects content-analytic concerns of transforming qualitative content to quantitative output. We first discuss two methodological perspectives that are common in evaluating deceptive discourse, namely, *abstract feature extraction* and *phrasal analysis*. We then introduce our method of automated phrasal analysis, describing the operational underpinnings of the Gramulator. We then turn to our primary objective of applying the Gramulator to deceptive scenarios involving everyday conversations and persuasive argumentation. In doing so, we show how certain phrases and word choices are uniquely tailored to each scenario, and how overarching psychological themes can emerge from a diverse set of narrative and rhetorical styles. Finally, we speculate on future developments and needs of the current approach.

An Inductive Approach to Automated Phrasal Analysis

In general, content-analytic research is defined as the attempt to extract meaningful representations from large sets of qualitative material, where these representations are derived from objective methods that can be easily reproduced, and that can be interpreted to yield new insights on how people might differ (Holsti, 1969; Smith, 2001; Stone, Dunphy, Smith, & Ogilvie, 1966). In the current studies, we evaluate discourse generated during open-ended verbal communication. We analyze two contexts of such communication, one involving casual, typed conversations, and the other involving expository monologue (spoken and written). We are interested in how people's language might change during deception, particularly language consisting of short phrases of two to three contiguous word sequences, also called n-grams. We hypothesize that these sequences are important for capturing salient narrative themes (e.g., types of characters, locations, events, feelings; Mandler & Johnson, 1977; Reissman, 1993) or pragmatic elements (e.g., dialogue acts, disfluencies, editing expressions; Clark, 1996; Schober & Brennan, 2003) that best characterize deceptive language. However, this approach is *inductive* insofar that the organization of these textual units into psychologically interesting constructs is not known *a priori*, but must be interpreted based on compatibility with existing theories. As described below in the section, *The Gramulator*, we go to great lengths to ensure that the extracted textual units are statistically more probable in deceptive texts than in non-deceptive texts (and vice versa), and that these textual units are interpreted within the local sentential context in which they originally occurred (by using a specially adapted concordancer tool).

For deception, homing in on the specific phrases and unique wording can have potentially important consequences in detecting deception. In studies where people are asked to record what they lie about during the course of a day, the bulk of deception tends to dwell on feelings and opinions, as well as personal preferences, achievements, and failures (DePaulo et al., 1996; DePaulo & Kashy, 1998). Thus, knowing what people tend to lie about can signal when a would-be detector should be particularly vigilant. The thematic content of deceptive speech is also relevant for understanding information management strategies that accompany deceptive intent (McCornack, 1992). These are strategies that deceivers use to control the content of a message by obfuscating the truth and thwarting perceptions of guilt. Such control is showcased by Burgoon et al. (1996) who evaluated what was said by deceivers in structured interviews. The researchers concluded that deceivers tend to provide impoverished details, downplay personal involvement, and provide less relevant information. Again, such information is crucial for enhancing the goals of deception detection.

To evaluate language, content-analytic researchers often rely on human raters to code theoretically interesting features, where high agreement between raters is a priority. However, this process can be extremely time-consuming, particularly when there are many texts and multiple features to code. Furthermore, human raters can easily overlook subtle semantic patterns that are embedded in more salient content.

Of course, the limitations of human raters are easily contrasted with the processing speed and pattern extraction abilities of computational approaches. As in the studies of Duran et al. (2010), Hancock et al. (2008), and Newman et al. (2003), natural language processing tools have been used to process hundreds of linguistic features in a matter of seconds. Many of these features are also likely impossible for human raters to identify. For example, Duran et al. (2010) used the *given-new index* available in the Coh-Metrix natural language processing tool to capture a construct of *information novelty* (Graesser et al., 2004; McCarthy, Dufty, Hempelman, Graesser, Graesser, & McNamara, in press). This index functions by computing the co-occurrence patterns of content words across contiguous sentences in a text. The algorithm is not dependent on any *a priori* notion of what raters might agree to be typical, or even what raters would recognize as being typical. Rather, the algorithm is designed to mindlessly (quite literally) evaluate hundreds of texts in terms of the amount of new information present in each text.

As is generally the case in this and other computational approaches, the data are interpreted without any direct reference to the specific words (or extended text). That is, the analysis is based on a composite measure of the abstract properties of words and relationships between words. When encountering the phrase “this is a chair,” computational algorithms similar to *given-new* might track information like: there are four words, there is one verb phrase, there is a noun that is a hypernym of furniture, et cetera. Thus, the analysis, by design, transforms the semantic content into higher-level, abstract properties. Two texts could be about very different topics, but potentially have the same Coh-Metrix values. However, as noted earlier, there are notable advantages in evaluating short, semantic phrases of discourse. For example, although a *given-new* evaluation might show that deceivers tend to be more redundant, this conclusion could be strengthened by also determining what people are more likely to talk about when telling the truth, or avoid talking about when telling a lie. In this way, researchers can begin asking why certain themes are avoided and others are not, and ultimately use this information to improve our understanding of the psychological underpinnings of deception, as well as the development of techniques for enhanced detection.

The Gramulator

Natural language processing tools have been tremendously successful at offering insight into language register differences (Duran, McCarthy, Graesser, & McNamara, 2007; Kahn, Tobin, Massey, & Anderson, 2007; McCarthy, Myers, Briner, Graesser, & McNamara, 2009; McNamara et al., 2010; Pennebaker & King, 1999). However, in all such analyses, the emphasis is on converting content to abstract representation. The fact that a corpus contains the words “happy”, “grateful”, or “confident” is only important insofar that these words count towards a pre-defined measure, such as *percentage of positive emotion words*, or that these words can be categorized by statistical techniques (e.g., factor analysis) into high-level conceptual categories. Given that such output is removed from the actual context in which these words originally appeared, researchers might overlook changes in meaning that are context-dependent. What is needed then is a computational tool that can complement existing techniques by revealing context-embedded features that occur within a text. In carrying out this goal, it is also of great benefit to have a tool that is computationally inexpensive (i.e., a tool that requires few resources, is easy to install and run, and does not need special coding for each analysis). And unlike other approaches that purport to integrate qualitative and quantitative analysis (e.g., DocuScope; Kaufer, Ishizaki, Ishizaki, & Butler, 2000), the Gramulator's extracted text features are not pre-defined in a dictionary-like structure. As such, the features themselves require are not restricted to any particular linguistic or psychological theory. Consequently, the Gramulator provides a straightforward means to quantitatively explore the qualitative nature of language.

The *Gramulator* is a freely available computational textual analysis tool.[7] It is designed to identify the differential linguistic features of correlative text types. That is, it primarily functions by identifying the key linguistic features of two related, yet theoretically distinguishable, sets of data. The Gramulator performs this function by first extracting what is *typical* to each individual set of data, and then eliminating what is common to those two sets of data (*reciprocal*). The elements that are left (*differentials*) are indicative of one set of data while being antithetical to the other set of data. The Gramulator also provides many associate modules for the further analysis of the revealed data. These modules include a concordancer, parser, lemmatizer, and an array of textual

---

[7] The Gramulator can be downloaded at: https://umdrive.memphis.edu/pmmccrth/public/index.html

measures such as frequency counts, genre categorization, lexical diversity measures, semantic assessment and so forth.

Central to the Gramulator's identification of relevant linguistic features is the extraction of statistically relevant distributions of n-grams. N-grams are adjacently positioned lexical items in a text that can be composed of any number of words, although here (as is typical) we focus on two-words (i.e., bi-grams) and three-words (i.e., tri-grams). For example, the initial four bi-grams of the first sentence of this paragraph are *Central to*, *to the*, *the Gramulator*, and *Gramulator is*; whereas the initial four tri-grams are *Central to the*, *to the Gramulator*, *the Gramulator is*, and *Gramulator is the*.

N-grams are useful textual analysis units because they capture examples of language that may be unique or rare to a given text type (Jurafsky & Martin, 2009). They are also widely used in statistical models of language to generate predicted linguistic structures (Christiansen & Chater, 2002), and have been instrumental in designing speech recognition and information retrieval systems (Manning & Schütze, 2000). For this study, we extend the use of n-gram analysis by incorporating n-gram features with *contrastive corpus analysis* (*CCA*) (McCarthy et al. in press; Min & McCarthy, 2010). CCA refers to the approach of characterizing one corpus by contrasting it with a related corpus (Cobb, 2003; Damerau, 1993; Granger, 1998). In doing so, CCA redirects the question a discourse analyst might ask, from: *what do the most frequent features in a text type tell us about the text type*?, to instead: *what are the most frequent features of one text type relative to another, and how do these features distinguish either text type?* This latter question, unlike the former, can only be answered by simultaneously considering *what is commonplace* and *what is not commonplace* across two corpora.

The Gramulator's operation can be understood as a multi-step process. These steps include a) collecting two candidate corpora for analysis, b) identifying and tallying the n-grams that appear in each corpus, c) retaining only the n-grams that appear with above average frequency in each corpus (i.e., the *Typicals*), and d) comparing the Typicals of each corpus and removing those Typicals that overlap. In this latter step, the goal is to identify Typicals that are not shared (the *Differentials*), and therefore are indicative of their respective corpus relative to the corpus against which they have been compared (see Figure 11).

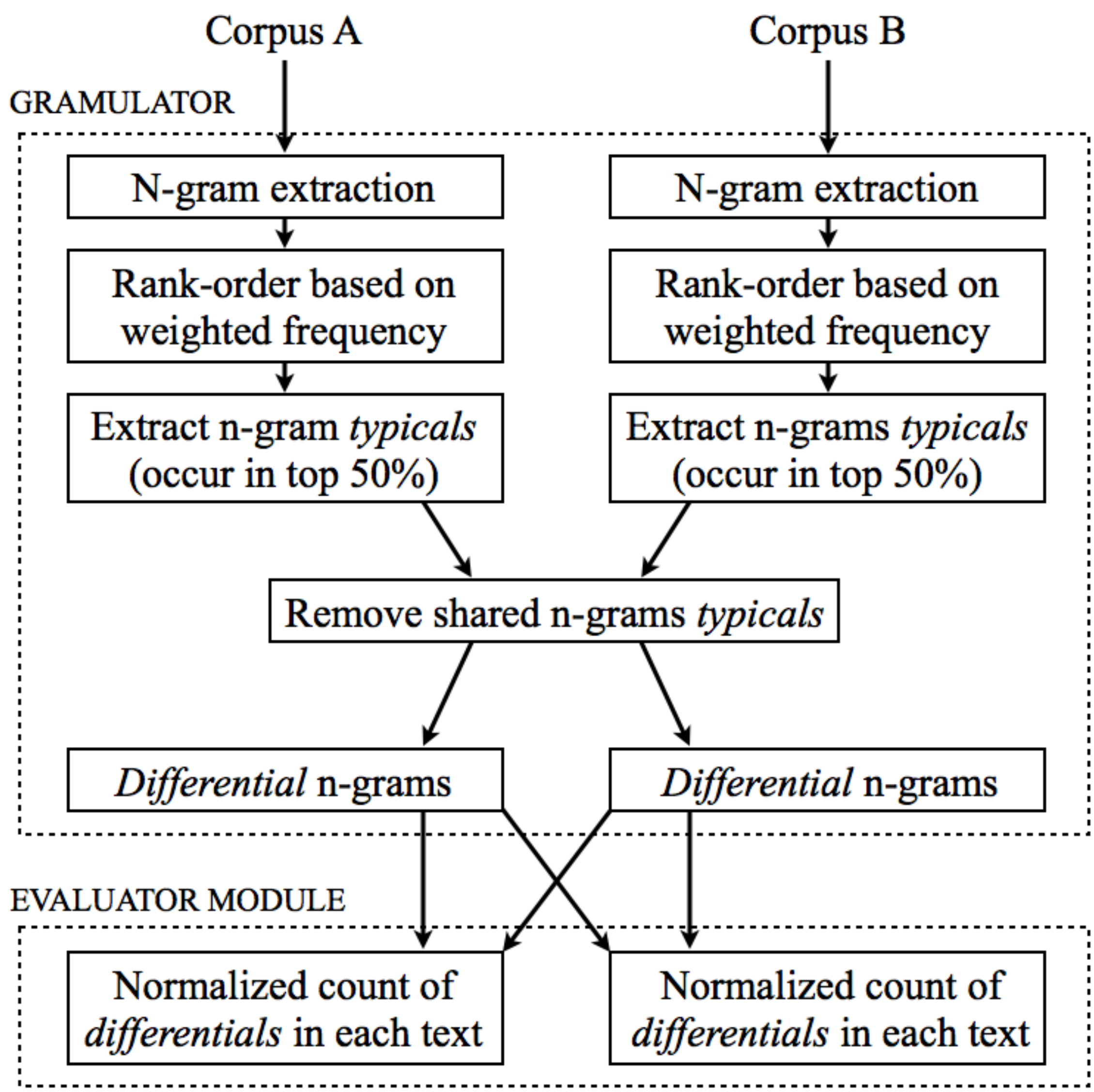

Figure 11. A schematic flow chart for the operations involved in deriving and computing differential

The first step of corpus collection is to select sister corpora. These are corpora where systematic differences are minimized apart from a single experimental manipulation. This process is exemplified with the current studies where experimental factors were held constant except for the manipulation of deceptive versus honest response instructions. The next step is to separately compute the *typical n-grams* for each of the two sister corpora. These are n-gram frequency counts that have been weighted to ensure that any n-gram that appears in multiple texts is considered more *typical* than an n-gram that occurs multiple times in just a few texts. The Gramulator's weighting approach is based on a modified procedure common in information retrieval and text mining procedures (Spärck Jones, 1972). As an example, let *x* and *y* be two n-grams that have a raw frequency count of 100 occurrences in one of the sister corpora. N-gram *x* occurs in just 10% of the corpus' texts, whereas *y* occurs in 50% of the texts. The Gramulator's default weighting function considers *y* to be more typical than *x* because *y(weighted count)* = 100*.50 = 50 and *x(weighted count)* = 100*.10 = 10. Next, in the third step, we simply rank order all n-grams in each corpus by their weighted counts and only select those that are among the top half of frequent n-grams.

At this point in the analysis, there are two *typical* n-grams sets corresponding to each of the sister corpora. Of course, among these typical sets there will be *shared n-grams*, that is, n-grams that are highly frequent in each corpus (e.g., *of the*). In the fourth and final step, the Gramulator removes all shared instances because any n-gram that is typical to both corpora is diagnostic of neither. This process of removing high frequency shared n-grams to leave only high frequency non-shared n-grams is a form of Machine Differential Diagnostics, a technique commonly used in medical diagnostic software (Graber, Tompkins, & Holland, 2009; Rahati & Kabanza, 2010). Consequently, the n-grams that are left after the removal process are referred to as "differentials." These differentials are diagnostic of each corpus because they are typical of one corpus while being atypical of the corresponding sister corpus.

The Indicative Language of Deception

For the following studies, we use the corpora from the published work of Hancock et al. (2008) and Newman et al. (2003). These corpora have been collected under rigorous standards of experimental control and are thus ideal for present purposes. Moreover, these corpora have been subjected to extensive linguistic computational analyses that have produced linguistic characteristics unique to deceptive and truthful language. As such, these corpora provide an important point of departure for the current approach. By using the corpora here, we will reveal additional meaningful linguistic characteristics that have remained hidden, and in doing so, highlight the importance of an automated n-gram analysis for gaining new theoretical insights.

Study 1: Everyday Conversation

Within almost any everyday conversation, there is a good chance that one or both conversational partners will tell at least one lie (DePaulo et al., 1996). Although such "fibs" are generally harmless, introducing deception to conversation creates unique challenges for the conversational participants. One such challenge arises from the violation of the well-established conversational *Maxim of Quality*: that speakers should avoid saying what they know to be false (Grice, 1975). By violating this maxim, the speaker might sabotage the willingness of the conversational partner to contribute meaningfully to the ongoing discourse. Thus, it is important that deceivers stay committed to preserving their partners' presumption of truth. In doing so, deceptive speakers must take special care to monitor their partners' speech and respond in such a way as to maintain the appearance of common ground and mutual knowledge (Clark, 1996). Furthermore, deceptive speakers must maintain representations of both truthful and non-truthful versions of that truth, and convincingly present the non-truthful version while avoiding unintentional “slips” of the potent truth. Given the increased cognitive control required, as well at the need to maintain conversational norms, it is reasonable to suspect differences in language use.

In recent years, there has been growing interest in how these cognitive and social factors shape conversational deception. An important strand of this research has been a focus on computer-mediated communication (e.g., Carlson et al., 2004; Hancock et al., 2004; Zhou et al., 2004). As the technology becomes more advanced, the use of CMC is quickly becoming a daily fixture in people’s lives. Although CMC varies greatly with the conversational affordances it shares with face-to-face communication, CMC does offer face-to-face attributes of synchronous and cotemporaneous message exchange (Brennan & Lockridge, 2006). Of these CMC approaches, instant messaging is a primary example.

Not surprisingly, there has been a strong move toward evaluating deception in an instant-messaging communication medium. One such evaluation is the research by Hancock et al. (2008). In that study, Hancock and colleagues asked unacquainted same-sex pairs to discuss four “ice-breaker” topics over instant-messaging software. The topics included 1) *Discuss the most significant person in your life*; 2) *Talk about a mistake you made recently*; 3) *Describe the most unpleasant job you have ever had to do*; and 4) *Talk about responsibility*. Prior to the conversation, the researcher gave one participant (the *sender*) all four topics and told the sender to lie about two of them and tell the truth about the other two. The assignment of lie or truth to a topic was counterbalanced, such that each topic was equally likely to be the basis for deception or truth-telling. The conversations were allowed to proceed as long as was needed, and were always initiated by the sender. Based on their analysis, which used the LIWC computational tool (Pennebaker et al., 2001), Hancock and colleagues found that deceptive conversations compared to truthful conversations used more overall words, but fewer words per utterance. Deceptive speakers also used more third person pronouns (e.g., *he*, *she*, *they*) and sense words (e.g., *see*, *touch*, *listen*), while more question marks were used by the receivers of the lie.

For the current study, we used the same corpus of deceptive conversations employed in Hancock et al. (2008). However, rather than identifying counts of linguistic properties, we were interested in the indicative themes that emerge in the senders' deceptive speech relative to their truthful speech. And, while Hancock et al. (also see Duran et al., 2010) looked at both the sender and receiver of deception, we focus here on just the language of the deception. The reason for doing so is that although receivers may also lie (just like the senders), the receivers were not instructed to do so, and so it is more difficult to assess their contributions as truthful or deceptive. In total, the study includes 130 transcripts (66 true and 66 deceptive) from 33 unique participants, with each transcript containing an average of 140 words (or *tokens*).

Data Extraction and Validation

As previously discussed, our goal is to merge the strengths of automated phrasal analysis with that of a contrastive corpus analysis approach. To this end, the Gramulator was used to capture qualitative content in the form of statistically relevant n-gram examples. We begin with the quantitative data extraction.

The Hancock et al. (2008) texts were processed using the Gramulator, which automatically stores the *typicals* (i.e., the weighted most frequent n-grams for each corpus regardless of the frequencies in the sister corpus) along with the *differential n-grams* (i.e., the weighted most frequent n-grams for each corpus relative to the sister corpus). For present purposes we are most interested in the differentials, which are indicative of the TRUE and LIE corpus. Again, a "true" differential are those n-grams that are among the most frequent in the true corpus *and* are *not* among the most frequent in the lie corpus (noting that "not among the most frequent" does not necessarily mean absent). And for a "lie" differential, these are the n-grams that are among the most frequent in the lie corpus *and* are *not* among the most frequent in the true corpus. The reader is also asked to keep in mind that frequency here has been weighted to ensure that n-grams are widely distributed across texts within each corpus (see *The Gramulator* section above). In the end, the Gramulator produces a set of *true differential bigrams* and a set of *lie differential bigrams* (hereafter referred to as *dT-grams* and *dL-grams*, respectively).[8]

To assess the validity of these differential sets, we used the Gramulator's default normalization function to count the occurrence of dT- and dL-grams in individual texts while taking into account text length. This step is important because it ensures that subsequent analyses are not unduly influenced by differential sets that have more instances, or by longer texts that are more likely to contain differentials. This normalization process is streamlined by one of the Gramulator's built-in modules, *the Evaluator*. The output of this module consists of two variables for each of the texts under evaluation, consisting of a normalized frequency count of dT- and dL-grams. Again,

[8] An analysis was also run using tri-grams. The results largely overlap with the bi-gram analysis; however, a few examples of tri-gram differentials are reported in the "Content-based Analysis" section for the purpose of providing additional support to the bi-gram identified trends.

normalization is based on the total number of words in the text[9], and by the total overall number of dT- and dL-grams being processed. The final value is raised by a factor of 100 for ease of reading. Thus, if Text 1 features 180 words, the number of dT-grams found in the text is 10, and there are 100 dT-grams overall, then the normalized dT-gram value for Text 1 is 10/100/180 * 10000 = 5.55.

Cross-validation. In our first validation test, we are interested in the generalizability of the extracted dL-grams across different groups of participants that are lying or telling the truth. Specifically, if the dL-grams are indicative of deceptive language, then these dL-grams are likely to generalize to an independent group of deceptive participants that are tested under similar conditions. Moreover, when compared to an independent group of truthful participants, also similarly matched, the dL-grams are likely to be used to a lesser extent.

To evaluate generalizability, we randomly divided the TRUTH and LIE corpora (66 texts in each corpus) into training and test sets. The training sets contained two thirds of all texts, or 43 texts, and the test sets contained the remaining one third, or 23 texts. Starting with the extraction of dL-grams, we used the Gramulator to evaluate the LIE and TRUTH training sets. This analysis produced 88 dL-grams. With these 88 dL-grams, we simply counted their normalized value, per text, in the LIE training set, the LIE test set, and the TRUTH test set. As a reminder, the dL-grams used here were derived from participants contributing to the LIE training set alone, and not from those participants contributing to the two test sets.

[9] The Gramulator can produce values based on *type* count instead of *token* count; however, the present default setting is token count, which we use throughout the study.

Evidence for generalizability will be provided if the dL-gram count in the LIE *training* set (in which the dL-grams were derived) is similar in number to the LIE *test* set, and importantly, compared to the LIE *test* set, the dL-gram count in the TRUTH *test* set is much lower. In other words, the language of deception from one group of participants should generalize to another group of deceivers, but not to a group of truth-tellers – even when the general topics being discussed are identical.

Our results confirmed these general hypotheses. There were no statistically significant differences between the LIE training set data and the LIE test set data. However, between the two test sets (LIE and TRUTH), a t-test revealed a higher dL-gram occurrence for lie texts ($M = 6.80$, $SD = 2.98$) than for true texts ($M = 2.61$, $SD = 1.95$), $t(42) = 14.96$, $p < .001$. Both findings validate the claim that the Gramulator is extracting content that is indicative of deception.

Embedded lies and truth. Having established cross-validation, another validation question is whether there is more "truth" in LIE texts or more "lies" in TRUTH texts. In Gramulator terms, we ask whether there is a greater occurrence of dT-grams in the LIE texts compared to the occurrence of dL-grams in TRUE texts. To find these embeddings, it is important to remember that differential n-grams are derived by comparing the top half of frequent n-grams across two corpora. However, the differential n-grams for one corpus might still occur among the bottom half of n-grams in the contrasting corpus. It is the occurrence of dL-grams or dT-grams in these bottom halves that is most relevant for this analysis.

We expect that the occurrence of dT-grams in LIE texts will outweigh the occurrence of dL-grams in TRUTH texts. As noted by Spence (2004), there is rarely a clear distinction between truth and lie in the “real-world.” To lie successfully, it is often best to start with the truth and strategically add elements of deception. And while a large part of a lie must be true, no part of a truth need be a lie. This approach to deception is well-aligned with Gricean expectations that the truth is the default in communication (Grice, 1989). Even when communicators violate a conversational pact of truthfulness, the bulk of what is said will be grounded in truth. By doing so, appearances of cooperativity and quality are maintained.

To conduct this analysis, we first evaluated all 132 texts (66 TRUE, 66 LIE) in the complete TRUE and LIE corpora to produce 51 dT-grams and 87 dL-grams. We then used a linear mixed-effects ANOVA to compare the normalized dL-gram count in TRUE texts with the normalized dT-gram count in LIE texts, while simultaneously controlling for random effects due to each participant’s contribution of multiple text types. The results show that there are more dT-grams in lie texts ($M = 2.67$, $SD = 2.15$) than dL-grams in true texts ($M = 1.98$, $SD = 2.12$), $F(1,130) = 3.68$, $p < .05$. As hypothesized, these results indicate that the composition of deception and truth contain elements of each other, with more of the truth found in the language of deception. These results also hint at one of the reasons deception is so difficult to detect; it is shrouded in elements of everyday or normal language.

Overall, in both validation studies, the quantitative analysis here suggests that the Gramulator extracts thematic content that is most characteristic of one corpus, relative to the other. The analysis also suggests that the indicative language of lies may be sufficient to distinguish LIE texts from TRUTH texts. And finally, we provide evidence that the indicative language of truth is likely to be found in lies, at least to a greater extent than the language of lies can be found in truth.

Inductive Interpretation

Turning now to the question of what is actually said, we present a breakdown of the top-ranked differentials that were extracted in the previous analysis. Although we are interpreting data that is qualitative in nature, we bring a rigorous experimental and computational perspective that deviates somewhat from traditional qualitative procedures. We are interested in patterns that are generalizable over many randomly selected participants, rather than the "exceptions" that are the focus of some qualitative and ethnographic case studies (Miles & Huberman, 1994). The patterns themselves are extracted not by subjective observations, but with machine learning algorithms. It should also be recognized that the corpora under analysis also come from carefully controlled experimental studies, and that various quantitative analyses of these corpora have been published elsewhere in well-established, peer-reviewed journals.

To continue with our treatment of the data, we have developed strict criteria for guiding our interpretations. These criteria are based on the simple notion that the differentials with the highest frequencies of weighted-occurrence are likely to give the strongest and clearest signal (i.e., diagnostic). Examination of such differentials generally leads to contextual clues as to where differences between the corpora lie. From these differences, the objective is to identify *thematic trends* and *pragmatic elements* that capture the underlying sentiment of the data. For these trends to be plausible, they need to be based on theory and to include statistically distinct differentials. Each of these criteria are expanded upon in the subsequent section, but first we give special consideration to the notion of “statistical distinctiveness.”

For a differential to be a candidate for interpretation, it must statistically occur more often in one corpus than in another. To ensure this is the case, theoretically interesting differentials are evaluated using a Fisher’s Exact test. The process can be exemplified as follows: a corpus (e.g., LIE) contains the target n-gram of “going to” for 9 instances across 8 of a total of 66 texts. The sister corpus (here, TRUTH) contains the target n-gram of “going to” for 4 instances across 2 of a total of 66 texts. Using Fisher’s Exact, the differences between these two counts is deemed not significant (p =.096) and would not be used for further interpretation. In reporting these tests throughout the paper, we have translated the operations into the format (9, 8:66; 4, 2:66, p = .096), where the first three numbers correspond to the corpus of interest (LIE), and the second three to the comparative corpus (TRUTH).

Distancing by focus on the “other.” In previous research examining the occurrence of personal pronouns as a marker of deception, it is assumed that deceivers use personal pronouns to avoid personal ownership of the deceptive content (Chung & Pennebaker, 2007). With more 2nd and 3rd person personal pronouns, the greater the chance that co-conspirators are being introduced into the deceptive narrative, thereby mitigating the deceivers own personal culpability. Likewise, by minimizing the use of 1st person personal pronouns, the deceiver can weaken any direct association to the deceptive content (Anolli, Balconi, & Ciceri, 2003). While such hypotheses are plausible, without further exploration of the larger context, any explanation remains highly speculative. What is needed is grounded interpretation that relates personal pronouns to their context of use.

Here, the differentials *are you, you feel* and *you can* were featured more regularly in deceptive language compared to the language of truth. One or more of these differentials occurred across 27% of the lie texts, but only within 12% of the truth texts (26, 18:66; 8, 8:66, p = .048). In Sentences 1-12 below, we highlight representative contexts in which *you*-based differentials occurred.[10] A quick glance at these sentence fragments reveals a distinct pattern: 2nd person personal pronouns are primarily used to elicit information from the conversational partner. Thus, the use of personal pronouns, when evaluated in context, co-occurs with a unique type of "other-oriented" question-asking behavior that is not seen in truthful exchanges. For example, in deceptive texts, 21 of 26 instances of *you*-based differentials were involved in getting the partner to talk more about themselves, but only 3 of 8 instances in truthful texts were used for the same purposes. And it is not that deceivers are merely avoiding opportunities to communicate, as would be the case if they were minimizing their odds of self-incrimination. In a previous analysis on this corpus, Duran et al. (2010) found that participants use more words overall when they are told to lie. What might be happening is that deceivers are attempting to establish a sense of rapport with their partner, and by doing so, attempting to appear more believable – a hypothesis aligned with the Interpersonal Deception Theory of Buller and Burgoon (1996).

[10] Although the contexts here are representative of the data as a whole, there were also a few examples that did not clearly fit with the overall interpretation. These are treated as outliers for present purposes.

1. if they are meant to be, you can always get back together in the end, right?
2. you and your mom are really close? do you feel comfortable about telling her everything?
3. why do you feel "not being concise" was a big mistake?
4. so are you fluent in Thai?
5. why? where are you going there? a major city?
6. where are you from? wow, never talked to a miner before.
7. like, are you more responsible?
8. what are you going to do differently? next time you talk...
9. are you in communications? did you hurt yourself?
10. are you responsible for any organizations here at...
11. where, are you ok?
12. wow. are you a biology major or chemistry major?

Biases in the narrative elements of setting and subject. The conversations in this current study were open-ended, allowing for a great deal of spontaneity. The narratives generated, both in the deceptive and truthful conditions, represented personal anecdotes that ranged widely in where they were set and who they were about. Indeed, for truthful narratives, there was no particular preference for the *where* and *who* being discussed, as might be expected given the freedom of the topics. However, for the deceptive narratives, there were indeed observable preferences for the where and who.

In terms of a location bias, there was a statistically significant tendency to situate the n-gram *high school* in deceptive texts. This differential was found in 15% of lie texts, compared to just 3% in true texts (14, 10:66; 2, 2:66, p = .030; see example sentences 1-5 below). Given that the majority of participants were a few years removed from high school, it appears that when being deceptive, participants return to familiar territory. Here, this recent event is better represented in memory than a more temporally distant event, and might be distorted with greater control. However, does this mean that deceptive content should always be based on the most recent events? This is problematic because the most recent events might also include a shared temporospatial setting with the conversational partner, thus making deception more easily detected. It seems that lies situated in high school strike a middle ground. These lies are removed from any possible shared experience with the conversational partner, but are also not too temporally distant that the truth is forgotten.

1. ...was just a very very messed up situation! <u>high school</u> years! oh yes, we all regret some stuff from...
2. ...like there were a lot of petty fights in <u>high school</u> over nothing at all!
3. ...impact on my life, it would probably be my <u>high school</u> hockey coach...
4. ...i've had to do was to break up with my <u>high school</u> boyfriend.
5. did you ever forge letters in <u>high school</u> so you could skip a class? haha.

In addition to locational biases in situating deception, there also appears to be a bias in *who* is mentioned in deceptive narratives. When performing deception, the prominent subject-oriented differentials that appear include *my boyfriend* (9, 5:66; 0, 0:66, p = .058; see example sentences 1-5 below) and *my friend/s* (14, 11:66; 5, 3:66, p = .045; see example sentence 6-12 below). One or more of these differentials occurred in 24.2% of the lie texts, and just in 4.5% of the truth texts (23, 16:66; 5, 3:66, p = .002; see example sentences 1-5 below).

1. yeah. broke up with my boyfriend of three years. yikes.
2. i live in the same building as my boyfriend. it's a great feeling.
3. i'd have to say my boyfriend is also the most significant person in my...
4. but i spend most of my time with my boyfriend.
5. that's good. my boyfriend and i are in entirely different majors.
6. well my friend back home also has a drug problem.
7. cool. my friend went there over christmas break.
8. ...took off on a weekend trip with my friend and basically the mice...
9. ...and thought i put my friends up to it, which is stupid, so he fired...
10. ...the wrong house looking for my friend's party. it was embarrassing
11. ...mistake was forgetting to call my friend on her birthday.
12. because my friend worked there so that made it a little better.

The greater use of the differential *my boyfriend* was not found for the reciprocal relationship, *my girlfriend*, despite there being no statistically significant differences in the number of male and female participants. These results are suggestive that females may be more comfortable lying about boyfriends than males lying about girlfriends. Clearly, the evidence here cannot put forward such a strong claim, but the direction of results certainly called for further investigation.

There was also a greater use of the differential *friend/s* in deception texts, but notably no difference between the truth and deceptive narratives with mentions of *best friend/s* (7, 5:66; 5, 4:66, p = 1.00). Furthermore, when *best friend/s* is used in the deceptive texts, it is in the context of *high school* (see example sentences 13-15 below), the location where deception might be best situated in order to avoid detection.

13. ...was to break up with my best friend's boyfriend for her while in *high school*...

14. made recently is that i forgot about my best friend from *high school's*

15. birthday. like my best friend in the entire world!

16. ...a friend who was suicidal. probably my best friend in *high school*.

It seems that mentions of *friend* are favored in deceptive texts over mentions of close personal relationships (e.g., *best friend*). Furthermore, when close personal relationships were part of the conversation; for example, referring to *my mother/my mom,* or *my dad/my father,* there was a trend (although not statistically significant) for these terms to occur more frequently in truthful texts (25.8% overall) compared to lie texts (16.7%). What might be likely, is that when formulating a lie, deceivers distance the lie from close friends and family, and opt to use more general, and perhaps less emotionally meaningful relationships (in which boyfriends are seemingly included). Similar behavior is evident in other work that has documented who lies are told to, with fewer lies being told to those who one feels closest (DePaulo & Kashy, 1998).

Narrative alteration by negation. Phrases of uncertainty were another set of frequently used differentials that occurred more often in deceptive texts. Differentials such as *i didn't, don't think, and don't know,* were present at least once in 24% of the lie texts, but in only 7% of the true texts (23, 16:66; 5, 5:66, p = .016). In previous research, the use of "thinking" terms (e.g., "think," "know"), has been hypothesized to occur more in deceptive language, and might be a byproduct of internally generated fabrications (Johnson & Raye, 1981; Vrij et al., 2000). Conversely, the use of negation (e.g., specifying what is and what is not) is expected to occur less frequently because it is more cognitively challenging to generate and maintain in dialogue (Hancock et al., 2008). However, an examination of context, via an n-gram analysis, shows that thinking and negation terms often occur together. This relationship would not have been identified if our analysis was limited to single words. Thus, we are able to capture a relationship between thinking and negation terms that points to a tendency for deceivers to avoid commitment to the deceptive narrative (see example sentences 1-4 below).

1. hmm. <u>don't think</u> i have. or, it doesn't stick out in my mind.
2. yeah. i <u>don't know</u> my personal stance on this---but i feel...
3. i <u>don't know</u> what to do with it. i don't make many...
4. see, i <u>don't know</u>. i would like to think talking things out...

Moreover, more negation in deceptive texts, as evidenced here with the greater use of *i didn't* (see example sentences 5-8 below), contradicts the hypothesis that negation will be avoided because of increased cognitive difficulty. Another account is needed as to why negation is employed more frequently. For example, it might be the case that negation makes the task of lying easier by simply allowing one to negate one aspect of a true event to make it false, allowing other details from the true event to be used. In Sentence 5 below, the fact that the deceiver did not want to live with "her" could be an example of negated truth, where the deceiver indeed did live with "her." By changing just one aspect of the narrative, true details about the person can be communicated. Although much follow-up work is required to confirm this interpretation, the more fundamental point is that different forms of negation can occur more or less frequently, and that by doing linguistic analysis that lumps all forms together (as in having a composite measure for "number of negated terms"), any interpretation of the results is likely to have meaningful exceptions that are not considered.

5. the hard thing was telling her that <u>i didn't</u> want to live with her. she took it ok,
6. <u>i didn't</u> tell them about the incidents. i just told them...
7. ...felt the same way. <u>i didn't</u> really speak to her afterwards.
8. ...so <u>i didn't</u> make it, but it was really embarrassing.

Temporal and causal sequencing. During more elaborate deceptive narratives, involving multiple event sequences, there is evidence that deceivers organize events in linear fashion, each unfolding in chronological order. According to Criteria-Based Content Analysis, a technique often used in forensics to evaluate verbal testimony, adherence to linear structure is easier to generate and maintain in memory, particularly when the lie is prepared in advance (Vrij, 2005). Related to this temporal sequencing is a tendency for deceivers to directly state the causal sequence for event and belief states, particularly those that directly involve the critical falsification. Deceivers do so to infuse confidence into their story, providing enough detail for plausibility, but also structured in such a way to be easily encoded and remembered by the deceiver.

In the current automated analysis, we found evidence for both these organizational tendencies. The deceptive narratives were more likely to contain the differential *and then* (example sentences 1-5 below), a transitional phrase that marks a chronological continuation of an event (Moens & Steedman, 1988) and the differentials *because i/my/we/the* (example sentences 6-10 below), causal phrases that clearly specify the reasons for events. Overall, these critical differentials occurred in 41% of the lie texts and only in 22% of the truth texts, a difference that was statistically significant (40, 27:66; 21, 15:66, p = .039).

1. ...business it was soooooooo boring. i hated it. <u>and then</u> i had another job, a camp counselor.
2. ...people burn popcorn <u>and then</u> we all have to leave the building.
3. but i would watch kids all day <u>and then</u> work till closing <u>at the</u> ice cream place.
4. ...it just completely slipped my mind. <u>and then</u> after the prelim my girls...
5. ...and she was not cute. <u>and then</u> my mom asked me about it.
6. turned out to be a huge mistake <u>because the</u> grad students were following...
7. and he's lucky <u>because the</u> missionaries there are taking care of
8. mine sucked <u>because my</u> boss told me i had to make sure each day...

9. i had a pretty high position because i had been working there for a few summers.

10. i couldn't do much for him, because he was gay and in the closet.

Brief Synopsis

The deceptive narratives were ones where certain locations and types of people were likely to be mentioned, and done so with chronological sequencing of events and definitive causal statements. There was also some hesitation in committing to personal involvement, and a greater tendency to redirect the focus of the conversation to the conversational partner.

Study 2: Persuasive Argumentation

The number of possible types of deception is virtually limitless. From warping the truth to omitting critical details, each type of deception has an associated difficulty of execution. But perhaps the lies that are most difficult to tell are those that require the rejection of personal beliefs and values. As anyone that has ever been in a heated political debate can testify, personal beliefs are often intertwined with notions of personal identity, and are strongly resistant to change (Noonan, 2003). And for this reason, the “Devil’s Advocate” approach has been recommended as one of the best ways for exposing hidden beliefs (Leal, Vrij, Mann, & Fisher, 2010). It is in this scenario that human judges achieve their most reliable and stable detection rates. When people must argue for a belief they do not truly believe, it is particularly difficult to maintain the same type of language as they might use when telling the truth.

In our second study, we extend the Gramulator to this more sensitive domain of deception. Our interest here is to test the robustness of the Gramulator and to provide further insight into the language of truthful and non-truthful representations of beliefs. We again reanalyze and expand upon an existing corpus by extracting *differential bigrams*, this time using essays and transcripts on abortion attitudes, published in Newman et al. (2003). In that study, participants were asked to argue for a *pro-choice* and *pro-life* position on abortion, either in a written or oral form. Prior to providing these arguments, the participants first indicated which of the two positions they favored. In this way, the researchers were better able to distinguish between TRUE and LIE contributions. To ensure that the participants were as persuasive as possible, they were told that their arguments would be read by naïve judges that would attempt to guess their true views. Participants were given 5 minutes to provide a response. The corpus consists of 352 texts from 176 unique participants (81 men, 95 women). Of these 176 participants, 50 identified themselves as pro-life (24 men, 26 women), and 126 identified themselves as pro-choice (57 men, 69 women). Texts were an average of 195 words in length. Although the Newman et al. (2003) deceptive context is clearly different from the earlier conversational corpus, the use of the Gramulator and method of evaluation are similar. As with our analysis of the Hancock et al. corpus in Study 1, we first compare TRUE pro-choice and pro-life arguments with LIE pro-choice and pro-life arguments. Then, for purposes of a cross-validation and inductive interpretation, we hone in on the language of TRUE pro-lifers and LIE pro-choicers.[11] In these corpora, the topic (pro-life) is held

[11] The focus on "pro-life" as a topic was to provide a parsimonious demonstration of the general differences that exists between deception and truth, and to do so in a domain that sufficiently varies from Study 1. Although the inclusion of "pro-choice" as a topic would have also met these general goals, it could also be considered redundant. Moreover, for analytical purposes, there are more unique participants deceptively arguing for a pro-life position than a pro-choice position. Thus, there is greater variation in the deceptive language, thus making the results of pro-life more likely to be generalizable.

constant between the truth and deception, ensuring that possible language differences are the result of the deception manipulation, and not confounded with topic shifts.

Data Extraction and Validation

By contrasting the TRUTH and LIE corpora using the Gramulator, we are able to extract bigrams that are indicative of one corpus but not of the sister corpus (e.g., dL-grams and dT-grams). Following the same series of analyses as Study 1, we first test whether the dL-grams used by one group of deceptive participants also occurs in a similarly matched group of participants. And in the second analysis, we isolate the language that is indicative of each of the two sister corpora (but, by definition, not typical of both), to determine how much of the lie information is embedded in truth texts, and how much of the true information is embedded in lie texts.

Cross-validation. In this study, we identify the dL-grams used by pro-choicers deceptively arguing for a pro-life position (i.e., “fake pro-lifers”) and evaluate whether these dL-grams also occur in a separate group of fake pro-lifers. In addition, because the dataset includes pro-life arguments from real pro-lifers, we can assess how convincing the fakers are at lying. Presumably, to be convincing, a good liar should use the same language as someone telling the truth. Thus, for the deception to be successful (i.e., undetectable), the language of the fake pro-lifer should also be found in the language of a real pro-lifer. If they are not similar, then we have strong support that fake pro-lifers are unconvincing, and importantly, the Gramulator is detecting qualitative information that distinguishes deception from truth.

To begin, we randomly selected two-thirds of fake pro-lifers deceptive essays/transcripts to form a LIE *training* set of 83 texts (corresponding to 83 participants) and the remaining third as a LIE *test* set. Using the Gramulator, the LIE training set was contrasted against their honest (pro-choice) arguments. At the end of analysis, 60 dL-grams were produced. Next, using the Gramulator's Evaluator module, we determined how many of the 60 dL-grams occurred in a) the original LIE training set, b) the LIE test set of 43 fake pro-lifers, and c) a randomly selected TRUTH test set of 43 real pro-lifers.

In the first analysis, comparing the LIE training and test sets, the occurrence of dL-grams did not differ for deceptive writers. As expected, the thematic content of one group of deceivers generalized to a separate group of deceivers. Next, in the comparison between the LIE and TRUTH test sets (i.e., fake and real pro-life arguments), a t-test revealed a higher dL-gram count in lie texts (M = 6.79, SD = 4.27) than in true texts (M = 4.19, SD = 2.62), $t(68) = 3.36$, $p < .01$. This result suggests that the test sets contained different qualitative information, even though both test sets were apparent arguments for the same pro-life position. The dL-grams, or what we are calling the indicative language of deception, appear to differentiate lies from truth.

Embedded lies and truth. Next, we evaluate how much of the indicative language of truth occurs in LIE texts, and how much of the indicative language of lies occurs in TRUTH texts. To do so, we first compare the entire LIE corpus (e.g., fake pro-lifers and fake pro-choicers) with the entire TRUTH corpus (e.g., real pro-lifers with real pro-choicers) to produce 45 dT-grams and 39 dL-grams. All 352 texts were then processed through the Gramulator's Evaluator module for values of dT-grams and dL-grams, resulting in the output of two variables for each text. These variables were normalized as in Study 1, thus controlling for differences due to length effects and the greater number of dT-gram targets (45 vs. 39).

Again, we confirm the general finding that the occurrence of dT-grams in LIE texts is higher (M = 2.27, SD = 2.60) than the occurrence of dL-grams in TRUTH texts (M = 1.79, SD = 2.14), $F(1, 350) = 3.96$, $p < 0.05$. Thus, we have further evidence to support the finding that participants employ more truth in deceptive statements than deception in truthful statements, suggesting that deceptive elements are embedded in what is mostly normal language.

Overall, the results suggest that the Gramulator is identifying thematic content that is more indicative of one corpus relative to another, given a particular context of communication. There is also no complete distinction between lies and truth, rather, elements of truthfulness pervade deception. And finally, despite a deceiver's goal to use "honest" language, this goal appears to fall short.

Inductive Interpretation

In the previous section, we confirmed that the Gramulator-produced *differentials* do indeed distinguish text types that, on the surface, are identical in topic, but vary substantially on some deeper psychological level. Again, the goal is to examine the differentials in greater detail to potentially reveal something new about the mental states of those that preserve or violate the truth. Thus, we provide an interpretation that is motivated by theory and grounded in a rigorous computational and statistical approach. And, unlike previous automated linguistic analysis of deception, the differentials are content-retrievable rather than numerical in nature. To reiterate our previous points, "hidden" intentions are not merely confined to composite linguistic measures of quantity, complexity, redundancy, etc., but can be examined in terms of the content that is actually expressed. In what follows, we highlight critical sets of differentials and their underlying themes, and end with a discussion of how these sets support and provide general insights into the cognitive states of deceivers and truth-tellers.

Distancing by equivocating. For the deceivers, the struggle to articulate a position is marked by increased differentials like *um* and *uh*, as well as phrases of equivocation like *I mean* (487, 101:126; 140, 30:50, $p = .008$). These differentials appeared in 80% of the deceptive texts compared to 60% of the true texts, with the overall rate of usage being over three times greater in deceptive texts than in true texts (487 versus 140). The use of *I mean* appears with common reformulations and hedges (see examples 1 to 6 below). Such usage may reflect relatively poor access to the topic matter that fake pro-lifers are trying to express, and may also explain why the examples appear more like an appeal to pathos than a presentation of an argument. This conclusion is supported by several other similar discourse markers, each of which appears as differentials: *um I*, *that uh*, *that um*, *and uh*, *um it.*

1. …kill a person, especially inside of you, <u>i mean, i mean</u> the person is alive…
2. … just have to look at it and it's wrong. <u>I mean</u>, abortion is wrong no matter…
3. …therefore, it, it, it's murder. <u>i mean</u>, you, the united states government…
4. …<u>I mean</u>, life, what a beautiful choice, man. <u>i mean</u>, it's a baby
5. …and, <u>i mean</u>, it's just, it's a horrible thing to do…
6. …wrong so it is wrong. uh, there's just, <u>I mean</u>, it's morally wrong

The use of *um* and *uh* are generally used to signal a delay in communication, with such delays arising from problems in planning, retrieving a word or idea from memory, or hesitation due to uncertainty about the appropriateness of what is being said (Brennan & Williams, 1995; Clark & Fox Tree, 2002; Goodwin & Goodwin, 1986; Levelt, 1989). Nevertheless, all interpretations converge on the general notion that deception places an *increased burden on processing.* These demands are also reflected in the use of *I mean*. Rather than signaling a delay, *I mean* is a prototypical editing expression that indicates a correction to, or justification of, the truth-value of some previous statement (Clark, 1996; Goffman, 1981). Given the uncertainty conveyed by *um* and *uh*, it is likely that deceivers feel compelled to use *I mean* as an attempt to clarify what they know to be a tenuous argument. These verbal cues also appear unintentional, unlike previous work that has found deceivers strategically minimize such cues to fend off the suspicion of others (Arciuli, Mallard, & Villar, 2010). It is possible that the emotional gravity of the present topic, i.e., whether abortion should or should not be allowed, limits strategic control that would normally be employed to avert suspicion.

Distancing by appealing to external agency. The differentials *it should, they should,* and *shouldn't be* occurred in 45% of deceptive texts compared to 28% in true texts (89, 57:126, 15, 14:50, p = .041). The question naturally arises: what then are the faker pro-lifers referring to when they use *should*-based differentials? By examining additional high-occurring differentials in deceptive language, the differential *to kill* is predominantly situated (30, 21:126; 2, 2:50, p = .025), as well as the presence of six additional differentials that correspond to "legal agency:" *the law, made legal, the bible, be allowed, allowed to, the government.* (88, 56:126; 24, 13:50; p = .027). Given this relationship, the underlying theme that appears to be most characteristic of fake pro-life arguments is *legal agency should prevent murder* (see examples 1 to 7 below).

1. …shouldn't be the government's decision to kill…
2. …we are not allowed to kill in our society…
3. …should be tried as if they had attempted to kill a full grown adult…
4. …no one should have the right to kill another defenseless human being
5. …people should not be allowed to murder a unborn child
6. …it should not be guaranteed *by law*
7. …they should be made legal in all states

When comparing these false arguments to real pro-life arguments, clear differences are manifest. Rather than an appeal to law (as an external force) and murder, real pro-lifers converge on more personal themes that *people should be responsible for their actions* and *the child should be given a chance/opportunity* (see examples 1 to 5 below). It is important to note that, as with fake pro-lifers, the use of "should" is a focal point in dictating obligations and responsibilities of the moral argument. But here, "should" is embedded in it's own set of *should*-based differentials; for example, the bigram *people should* (2, 2:126; 5, 4:50, p = .054) and slight variants of the trigram *should be given* (5, 34; 1, 83: p = .008); *be given*: (7, 34; 4, 83: p = .014); *given the*: (6, 34; 4, 83: p = .061); and *be given the*: (4, 34; 2, 83: p = .058).-But unlike the fake pro-lifers, the real pro-lifers are focused more on *the right, the opportunity, the chance, a fair chance, and every opportunity* (2, 2:126; 8, 7:50, p = .002) that one should be given, as well as the importance of *responsible/responsibility for their [own] actions* (3, 2:126; 4, 4:50, p = .055) (see examples 1 to 9 below).

1. …people <u>should be</u> responsible for their actions…
2. …morally and ethically wrong and I feel that <u>people should</u> take responsibilities for...
3. People <u>should be more careful</u> when having sex to prevent…
4. …take <u>responsibility</u> for their own action. It is not the baby's fault…
5. the child <u>should be given</u> <u>the chance</u> to live a life
6. the baby is innocent and <u>should be given</u> every <u>opportunity</u>… to have a happy and healthy life
7. …the child in the in the mother <u>should be given that right</u>
8. …<u>be given every opportunity</u> possible to have a happy and healthy life.
9. the child <u>should be given the chance</u> to live a life that god intended…

This Gramulator analysis suggests that deceivers tend to depersonalize their arguments by appealing to external standards of right and wrong (e.g., “the law”), whereas truth-tellers favor internal notions of moral imperatives. This external/personal divide is similar to the *distancing strategies* often attributed to deceivers (Bavelas, Black, Chovil, & Mullett, 1990; Buller & Burgoon, 1996; Newman et al., 1900). Much like in the earlier conversational corpora where deceivers find separation by situating their lies in remote time and space, distance in false arguments is achieved by referring to the standards of others, rather than one’s own sense of moral correctness. Thus, the argument is situated in the realm of what “others” generally think, rather than in what “I” personally feel. This finding is in line with deceivers’ reticence to give clear statements about their own opinions, and a tendency to refer to others and their actions in nonspecific terms (Burgoon et al., 1996). Even in our own data, deceivers used differentials like *they should* and *the fetus* rather than the more direct differentials *people should* and *a fetus* that were used by the truth-tellers.

Brief Synopsis

Combining the results of all analyses, the derived rhetoric appears to be one where fake pro-lifers (i.e., deceivers) struggle to articulate their position, and ultimately position their argument in terms of formal law. Conversely, true pro-lifers are less likely to appeal to external entities to justify their position. Instead, they rely on individual rights and personal responsibilities.

What is clear from this automated phrasal analysis and inductive interpretation is that it is difficult to adopt the true perspective of another when it violates one's own intrinsic beliefs. Major rifts were found in the content of deceptive and true essays/transcripts, even when the deceivers were attempting to adopt the opinion of actual pro-lifers.

Discussion

In this paper, we analyzed the distinctive phrases of deceptive and truthful discourse, doing so with a computational tool that extracts relevant text sequences based on their likelihood of appearing in one corpus relative to another. We found unique narrative elements used by deceivers in a domain where two strangers were simply getting to know each other; and in a second study, we found unique rhetorical elements in essays/transcripts where participants were asked to violate their true beliefs on abortion. In each set, the themes that emerged were supported by, and supplemented, existing theories of deception.

It is important to note that the themes for both studies were largely unique to each pragmatic domain. In the conversational narratives, the distinction between deception and truth was split along an event dimension, where the location, characters, and temporal/causal structure of events was central. Also, given the conversational nature of the interactions, unique strategies to appear more convincing, deter suspicion, or ease the cognitive burdens of deception, could be employed. Conversely, for the persuasive arguments, the focus was on rhetorical style, not on event narration. Thus, the distinction between deception and truth was by what criteria the argument could be best justified, and in the current study, it appeared to be one of external versus personal obligation.

These findings highlight the importance of evaluating linguistic behavior within targeted domains. Given the variability and flexibility of language, the quest to find universal linguistic patterns might be an exasperating venture, particularly with a behavior like deception that is employed for seemingly innumerable reasons. However, as more domains are evaluated, *meta-themes* might emerge that bring researchers closer to identifying near universals. One such meta-theme is the use of *distancing strategies*, mentioned earlier with the discussion of true and fake pro-lifers. In that study, deceivers distanced themselves from the content of their lies by giving externally-situated justification, as well as exhibiting greater equivocation. And likewise, with the conversational narratives, deceivers were more likely to emotionally distance by avoiding mention of people presumably closest to them (best friends and family), and by routinely reorienting the focus of the conversation to their conversational partner. If this general behavior holds, it might also suggest fundamental cognitive or affective constraints that generally influence deceptive behavior.

Another major contribution of this research is the development of a purpose built natural language processing tool. With this tool, the Gramulator, we were able to detect hidden chunks of content in the form of statistically relevant n-grams that we call *differentials*. The Gramulator has many of the advantages of previous computational analysis of deception. It uses computationally light algorithms of pattern extraction to bypass subjective selection that might compromise the identification of stable and reliable units of analysis. The resulting extracted units are submitted to standard inferential statistical tests, such as those used in these studies (linear mixed-effects ANOVA, paired t-tests). In doing so, we effectively integrate phrasal analysis with the rigorous control of a computational approach.

Given the importance of understanding the "what" being discussed; that is, the semantic content of the discourse, a great deal of emphasis within the current work has been on understanding linguistic features in context. In other computational approaches, linguistic features are typically reported as being a certain percentage of some linguistic set. Often it is unclear whether percentages are distributed across multiple texts in each corpus, or are confined to a few idiosyncratic instances. For example, if tracking the use of "ums" and "uhs," it might be the case that one or two texts in Corpus A are laden with um/uh fillers while elsewhere in Corpus A or Corpus B they rarely occur. To solve this problem, we employ an analysis that gives a weighted advantage to content features that are distributed over multiple texts within a corpus, and then apply a follow-up statistical test (i.e., Fisher's exact) to ensure that the appearance of any feature in one corpus is greater than its appearance in a comparison corpus. As further assurance that the evaluated language is being used similarly in multiple texts, the features are interpreted by explicitly evaluating all contexts in which the features appear. We look for consistent trends, but also polysemous meanings that should be taken into account. Indeed, in the conversational narratives of Study 1, the use of negation mapped onto a construct of noncommittal and a construct of event negation. Such graded differences would be nearly impossible to isolate with computational approaches that lump together all instances of a linguistic feature into a single percentage-based measure.

In comparing the Gramulator tool with other approaches, there are also potential limitations on our end. For example, deceivers might be quite good at monitoring and controlling the content-based “what,” thus making our approach easier to dupe when deceivers know what themes to avoid. On the other hand, for analyses that measure more stylistic and abstract features, it has been argued that such features are more difficult to control, thereby making them better candidates for deception detection (Chung & Pennebaker, 2007; Newman et al., 2005). We do not deny that such feature extraction has a great deal of potential, but we argue that the thematic “what” also has the potential to capture involuntary mental states. Further research is needed to determine if abstract features are indeed resistant to duping, and whether this resistance is superior to the phrasal analysis approach used here.

Although the current studies offer new insight into many features of truthful and deceptive discourse, a considerable amount of development lies ahead. For instance, uni-grams (1-word units) and quad-grads (4 word units) might also be assessed. In addition, the Gramulator tool has recently been modified to assess *off-grams*: sequences such as *word + [any word] + word* or even extended to *word + [any word + any word] + word*. And we are also implementing a parser with the tool so that n-grams can be specified for part of speech (POS) attributes. Off-gram analysis and POS attribute options open up numerous new avenues for investigation for Gramulator research. And it should also be acknowledged that although the Gramulator might come to offer numerous forms of n-gram analyses, that n-grams are not the only sequences of text that may reveal characteristic patterns of interest relative to registers. For example, the many various cohesion analyses generated from tools such as Coh-Metrix might yet be modified to highlight which cohesion patterns are contributing most to the final output. We also acknowledge that our interpretations of the differential n-gram arrays produced by the Gramulator are in need of automation. To this end, we are devising a flexible latent semantic analysis (Landauer et al., 2007) module that is designed to cluster the n-gram differential output. Clearly, a great deal is still in development; however, we argue that the results of the current studies offer compelling evidence that an automated n-gram analysis can make a significant contribution to textual assessment and perhaps to better revealing how participants represent truthful and deceptive representations of their beliefs.

Chapter 4: Conclusion

In this dissertation, I presented a comprehensive examination of the nonverbal and verbal signatures of low-level, fast-acting, and often-unintended cues that emerge during the course of deceptive responding. These cues were defined in terms of the cognitive processes involved when one must act in discordance with their knowledge of the truth. In the first section exploring nonverbal signatures, these processes were situated within an action dynamics framework, allowing an in-depth exploration of the competition involved when trying to resolve what is known to be true with what is intended to be false. Although this competition occurs over very short timescales, I was able to capture complex patterns by examining arm movements while participants responded falsely or truthfully to questions requiring a simple "no" or "yes" response. To answer, participants either navigated a Nintendo Wiimote-controller to "NO/YES" regions on a large projector screen (Study 1a) or a computer-mouse to "NO/YES" regions displayed on their personal computers (Studies 1b, 2a, 2b). Overall, trajectory analyses of the fine-grained arm movements show increased complexity in false responding over truthful responding, with the greatest difference in false "yes" answers. These motor movements also reveal greater strength of competition during the act of false responding, thereby extending traditional response time measures that capture latent competition alone.

Moreover, specific patterns of response time effects were identified as a result of varying the task-based cognitive demands associated with the content and manner in which questions were asked, as well as varying when false or true response prompts were presented. In Study 1b (which was a replication of Study 1a, using computer-mouse movements instead of Wiimote movements), the response prompts were presented with the final word of a question - questions whose content probed autobiographical knowledge. Conversely, in Study 2a, response prompts were presented with the initial word of an autobiographical question. And in Study 2b, the response prompts were again presented immediately at the final word of a question, but with questions now probing general semantic knowledge. These particular task demands were chosen to compare modulation of response competition, when participants can or cannot prepare for a false or true response (Study 1b vs. Study 2a), and when participants are asked questions involving autobiographical or semantic-based knowledge (Study 1b vs. Study 2b).

In the first of these two comparisons, it was found that participants who could anticipate how to respond (falsely or truthfully) experienced less response competition for true responses, but increased competition for false responses. The pattern of movements behind these responses, and the resolution of response competition over time, suggests that a "true prompt" decreases the bias to indiscriminately say "yes" when asked to tell the truth, but a "false prompt" increases a tendency to deny information, therefore making false responses that require a "yes" answer particularly difficult. In the second comparison, it was found that participants who responded to questions that probed semantic-based knowledge generally experienced less response competition for true and false responses. Presumably, falsifying information that is personal in nature is more difficult because this information tends to be more salient and readily believed, therefore making it harder to falsify. Across all studies, these results suggest that deception, operating under a variety of contexts, is uniquely detectable when action is allowed to covary with thought.

Turning now to the second major section, the focus is on the verbal signatures of deception. The words people use and the way they use them can reveal a great deal about their mental states when they attempt to deceive. The challenge for researchers is how to reliably distinguish the linguistic features that characterize these hidden states. In the first of two studies, I used a natural language processing tool called Coh-Metrix to evaluate deceptive and truthful conversations that occurred within a context of computer-mediated communication. Coh-Metrix is unique in that it tracks linguistic features based on cognitive and social factors that are hypothesized to influence deception. The results from Coh-Metrix were compared to linguistic features reported in previous independent research, which used a natural language processing tool called LIWC. The comparison revealed converging and contrasting alignment for several linguistic features and established new insights on deceptive language and its use in conversation.

In the final analysis, I conducted a similar natural language processing study as before, but now with a new tool called the Gramulator. However, there are notable differences that help provide a more comprehensive and complimentary linguistic analysis. The Gramulator uses statistical algorithms to extract snippets of text that are more likely to be used under one set of task conditions than in another. In Study 1, the task involved simple conversations amongst two strangers who were getting to know each other; in Study 2, the task involved participants arguing for a highly charged personal belief. In both studies, the manipulation of task instructions required participants to sometimes tell the truth and to sometimes lie. The resulting "snippets" of phrasal content found for each study, and for each condition (truth and lie), were then used to build an inductive interpretation of the psychological processes represented by the linguistic output. And unlike many other automatic natural language processing studies, these interpretations relied on how the output was used in its original context of use.

References

Abe, N., Suzuki, M., Mori, E., Itoh, M., & Fujii, T. (2007). Deceiving others: Distinct neural responses of the prefrontal cortex and amygdala in simple fabrication and deception with social interactions. *Journal of Cognitive Neuroscience, 19*, 287-295.

Anderson, M. L. (2003). Embodied cognition: A field guide. *Artificial Intelligence, 149*, 91-130.

Anolli, L., Balconi, M., & Ciceri, R. (2003). Linguistic styles in deceptive communication: Dubitative ambiguity and elliptic eluding in packaged lies. *Social Behavior and Personality, 31*, 687–710.

Arciuli, J., Mallard, D., & Villar, G. (2010). Um, I can tell you're lying: Linguistic markers of deception versus truth-telling in speech. *Applied Psycholinguistics, 31*, 397-411.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Baddely, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology, 49A*, 5-28.

Bavelas, J. B., Black, A., Chovil, N., & Mullett, J. (1990). Truths, lies, and equivocations: The effects of conflicting goals on discourse. *Journal of Language and Social Psychology, 9*, 135-161.

Beck, I. L., McKeown, M. G., Sinatra, G. M., & Loxterman, J. A. (1991). Revising social studies text from a text-processing perspective: Evidence of improved comprehensibility. *Reading Research Quarterly, 26*, 251-276.

Bok, S. (1999). *Lying: Moral choice in public and private life*. Vintage: New York.

Bond, F. C., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*, 214-234.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S. & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624-652.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433, 433-436.

Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron, 39*, 713-26.

Brennan, S. E., & Lockridge, C. B. (2006). Computer-mediated communication: A cognitive science approach. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics, 2nd Edition* (pp. 775-780). Oxford, UK: Elsevier Ltd.

Brown, B. (1977). Face saving and face restoration in negotiation. In D. Druckman (Ed.), *Negotiations: Socio-Psychological Perspectives* (pp. 275-299). Beverly Hills, CA: Sage.

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

Burgoon, J., Adkins, M., Kruse, J., Jensen, M. L., Meservy, T., Twitchell, D. P., Deokar, A., Numamaker, J. F., Lu, S., Tsechpenakis, G., Metaxas, D. N., & Younger, R. E. (2005). An approach for intent identification by building on deception detection. *In the Proceedings of the 38th Hawaii Internactional Conference on System Sciences*.

Burgoon, J., Buller, D., & Floyd, K. (2001). Does participation affect deception success? A test of the interactivity principle. *Human Communications Research, 27*, 503-534.

Burgoon, J. K., Blair, J. P., & Strom, R. E. (2008). Cognitive biases and nonverbal cue availability in detecting deception. *Human Communication Research, 34*, 572-599.

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

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.

Byrne, R. W. (2010). Deception: Competition by misleading behavior. *Encyclopedia of Animal Behavior, 1*, 461-465.

Carlson, J. R., George, J. F., Burgoon, J. K., Adkins, M., & White, C. H. (2004). Deception in computer-mediated communication. *Group Decision and Negotiation, 13*, 5-28.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032-1053.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive functioning and theory of mind? Contribution of inhibitory control and working memory. *Infant and Child Development, 11*, 73-92.

Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development, 69*, 672-691.

Carrion, R. E., Keenan, J, P., & Sebanz, N. (2010). A truth that's told with bad intent: An ERP study of deception. *Cognition, 114*, 105-110.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science, 280*, 747.

Charniak, E. (1997). Statistical techniques for natural language processing. *AI Magazine, 18*, 33-44.

Christ, S. E., Van Essen, D. C., Watson, J. M., Brubaker, L. E., & McDermott, K. B. (in press). The contributions of prefrontal cortex and executive control to deception: Evidence from activation likelihood estimate meta-analyses. *Cerebral Cortex.*

Christiansen M. H., & Chater, N. (2002). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23*, 157-205.

Chung, C., & Pennebaker, J. (2007). The psychological functions of function words. In K. Fiedler (Ed.), *Social Communication* (pp. 343-359). New York, NY: Psychology Press.

Clark, A. (1997). *Being there: Putting brain, body, and the world together again*. Cambridge, MA: MIT Press, 604-606.

Clark, H. H. (1996). *Using language*. Cambridge, England, Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1987). Collaborating on contributions to conversations. *Language & Cognitive Processes, 2*, 19-41.

Cobb, T. (2003) Analyzing late interlanguage with learner corpora: Québec replications of three European studies. *The Canadian Modern Language Review, 59*, 23-43.

Coltheart, M. (1981). The MRC psycholinguistics database. *Quarterly Journal of Experimental Psychology, 33A*, 497-505.

Colwell, K., Hiscock, C. K., & Memon, A. (2002). Interviewing techniques and the assessment of statement credibility. *Applied Cognitive Psychology, 16*, 287-300.

Crossley, S., Louwerse, M., McCarthy, P. M., & McNamara, D. S. (2007). A linguistic analysis of simplified and authentic texts. *The Modern Language Journal, 91*, 15-30.

Dale, R., & Duran, N. D. (2009). Dynamical characterization of semantic coupling in child-caregiver interaction. *Paper presented at the 15th International Conference on Perception and Action*, Minneapolis, MN.

Dale, R., & Duran, N. D. (in press). The cognitive dynamics of negated sentence verification. *Cognitive Science*.

Dale, R., Duran, N. D., & Roche, J. (in press). Action dynamics in language processing. In R. K. Mishra & N. Srinivasan (Eds.), *Language-Cognition Interface: State of the Art.*

Dale, R., Kehoe, C. E., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory and Cognition, 35*, 15-28.

Dale, R., Roche, J., Snyder, K., & McCall, R. (2008). Exploring action dynamics as an index of paired-associate learning. *PLoS ONE, 3*.

Damerau, F. J. (1993). Generating and evaluating domain-oriented multi-word terms from texts. *Information Processing & Management, 29*, 433–447.

Danker, J. F., & Anderson, J. R. (2010). The ghosts of brain states past: remembering reactivates the brain regions engaged during encoding. *Psychological Bulletin, 136*, 87-102.

Darlington, R. B. (1970). Is kurtosis really “peakedness”?, *American Statistician, 24*, 19-22.

DeCarlo, L. T. (1997). On the meaning of kurtosis. *Psychological Methods, 2*, 292-307.

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin, 111*, 203-243.

DePaulo, B. M. (1994). Spotting lies: Can humans learn to do better? *Current Directions in Psychological Science, 3*, 83-86.

DePaulo, B. M., Ansfield, M. E., & Bell, K. L. (1996). Theories about deception and paradigms for studying It: A critical appraisal of Buller and Burgoon’s lnterpersonal Deception Theory and Research. *Communication Theory, 6*, 297-310.

DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology, 74*, 63-79.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*, 979-995.

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

Donders, F. C. (1868). Over de snelheid van psychische processen. Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogenschool, 1968-1869, Tweede reeks, II, 92-120. Transl. By W. G. Koster, *Acta Psychologica, 30*, 412-431.

Duran, N. D., Bellissens, C., Taylor, R., & McNamara, D. S. (2007). Quantifying text difficulty with automated indices of cohesion and semantics. In D.S. McNamara and G. Trafton (Eds*.), Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 233-238). Austin, TX: Cognitive Science Society.

Duran, N. D., & Dale, R. (2009). Anticipatory arm placement in the statistical learning of position sequences. *Proceedings of the 31st Annual Conference of the Cognitive Science Society.* Amsterdam, Netherlands: Cognitive Science Society.

Duran, N. D., Dale, R., & McNamara, D. S. (2010). The dynamics of overcoming the truth. *Psychonomic Bulletin and Review, 17*, 486-491.

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.

Duran, N. D., & McCarthy, P. M. (2011). *Using n-gram differential diagnotics to reveal the thematic content of deception*. Unpublished manuscript.

Duran, N. D., McCarthy, P. M., Graesser, A. C., & McNamara. D. S., (2007). Using temporal cohesion to predict temporal coherence in narrative and expository texts. *Behavior Research Methods, 39*, 212-223.

Eapen, N. M., Baron, S., Street, C. N. H., Richardson, D. (2010). The bodily movements of liars. In In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 233-238). Austin, TX: Cognitive Science Society.

Ekman, P. (1997). Lying and deception. In N. L. Stein, P. A. Ornstein, B. Tversky, & C. Brainerd (Eds.), *Memory for Everyday and Emotional Events* (pp. 333–348). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Ekman, P. (2007). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. New York, NY: Henry Holt.

Ekman, P., & Friesen, W. V. (1972). Hand movements. *The Journal of Communication, 22*, 353-374.

Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist, 46*, 913-920.

Ekman, P., & O'Sullivan, M. (2006). From flawed self-assessment to blatant whoppers: The utility of voluntary and involuntary behavior in detecting deception. *Behavioral Sciences & the Law, 24*, 673–686.

Ekman, P., O'Sullivan, M., Friesen, W. V., & Scherer, K. R. (1991). Face, voice, and body in detecting deceit. *Journal of Nonverbal Behavior, 15*, 125–135.

Ekman, P., O'Sullivan, M., & Frank, M. G. (1999). A few can catch a liar. *Psychological Science, 10*, 263-266.

Evans, J. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255-278.

Evans, A. D., Xu, F., & Lee, K. (2011). When all signs point to you: Lies told in the face of evidence. *Developmental Psychology, 47*, 39-49.

Farmer, T. A., Cargill, S. A., Hindy, N. C., Dale, R., & Spivey, M. J. (2007). Brief reports: Tracking the continuity of language comprehension: Computer mouse trajectories suggest parallel syntactic processing. *Cognitive Science, 31,* 889-909.

Feeley, T. H., & deTurck, M. A. (1995). Global cue usage in behavioral lie detection. *Communication Quarterly, 43*, 420-430.

Feldman, R. S., Forrest, J. A., & Happ, B. R. (2002). Self-presentation and verbal detection: Do self-presenters lie more?. *Basic and Applied Social Psychology, 24*, 163-170.

Fodor, J. 1983. *The Modularity of Mind.* Scranton, PA: Crowell

Fodor, J. A., & Pylyshyn, Z. W. (1981). How direct is visual perception? Some reflections on Gibson's "Ecological Approach". *Cognition, 9*, 139-196.

Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General, 137*, 673-690.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosopical Transactions of the Royal Society Society B: Biological Sciences, 358*, 459–473.

Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L., & Yurgelun-Todd, D. A. (2003). Neural correlates of different types of deception: An fMRI investigation. *Cerebral Cortex, 13*, 830-836.

Garavan, H., Ross, T. J., Murphy, K., Roche, R. A. P., & Stein, E. A. (2002). Dissociable executive function in the dynamic control of behavior: Inhibition, error detection, and correction. *NeuroImage, 17*, 1820-1829.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27*, 181–218.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46*, 107-119.

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology, 59*, 601-613.

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology, 65*, 221-233.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20*, 1-55.

Gombos, V. A. (2006). The cognition of deception: The role of executive processes in producing lies. *Genetic, Social, and General Psychology Monographs, 132*, 197-214.

Graber, M. L., Tompkins, D., & Holland, J. J. (2009). Resources medical students use to derive a differential diagnosis. *Medical Teacher, 31*, 522–527.

Graesser, A. C., McNamara, D. S., & Louwerse, M. M. (2003). What do readers need to learn in order to process coherence relations in narrative and expository text? In A.P. Sweet and C.E. Snow (Eds.), *Rethinking reading comprehension* (pp. 82-98). New York, NY: Guilford Publications.

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

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.

Granger, S. (1998). *Learner English on Computer*. London: Longman.

Greene, J. O., O'Hair, H. D., Cody, M. J., & Yen, C. (1985). Planning and control of behavior during deception. *Human Communication Research, 11*, 335-364.

Gregg, A. P. (2006). When vying reveals lying: The timed antagonistic response alethiometer. *Applied Cognitive Psychology, 21*, 621-647.

Grice, P. H. (1975). Logic and converation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics* (Vol 3, pp. 41-58). New York, NY: Academic press.

Grice, P. H. (1989) *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

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

Hancock, J. T., Thom-Santelli, J., & Ritchie, T. (2004). Deception and design: The impact of communication technologies on lying behavior. *Proceedings, Conference on Computer Human Interaction* pp. 130-136. New York, NY: ACM.

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cogntiive Sciences, 5,* 394-399.

Hempelmann, C., Rus, V., Graesser, A., & McNamara, D. (2006). Evaluating state-of-the-art treebank-style parsers for Coh-Metrix and other learning technology environments. *Natural Language Engineering, 12*, 131-144.

Hempelmann, C. F., Dufty, D., McCarthy, P. M., Graesser, A. C., Cai, Z, & McNamara, D. S. (2005). Using LSA to automatically identify givenness and newness of noun phrases in written discourse. *In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 941-946). Mahwah, NJ: Erlbaum.

Holsti, O. R. (1969). Content analysis for the social sciences and humanities. Reading, MA: Addison-Wesley.

Hughes, C. J., Farrow, T. F., Hopwood, M. C., Pratt, A., Hunter, M. D., & Spence, S. A. (2005). Recent developments in deception research. *Current Psychiatry Reviews, 1*, 271–279.

Jensen, M. L., Meservy, T. O., Burgoon, J. K., & Nunamaker, J. F. (2010). Automatic, multimodal evaluation of human interaction. *Group Decision and Negotiation, 19*, 367-389.

Johnson, M. K. (1988). Reality monitoring: An experimental phenomenological approach. *Journal of Experimental Psychology: General, 117*, 390-394.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67-85.

Johnson, R., Barnhardt, J., & Zhu, J. (2004). The contribution of executive processes to deceptive responding. *Neuropsychologia, 42*, 878-901.

Johnson, R., Barnhardt, J., & Zhu, J. (2005). Differential effects of practice on the executive processes used for truthful and deceptive responses: An event-related brain potential study. *Cognitive Brain Research, 24*, 386-404.

Jurafsky, D., & Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics.* New York, NY: Prentice-Hall, 2nd Edition.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*, 228–238.

Juul Andersen, T. (2005). The performance effect of computer-mediated communication and decentralized strategic decision making. *Journal of Business Research, 58*, 1059-1067.

Kahn, J., Tobin, R., Massey, A., & Anderson, J. (2007). Measuring emotional expression with the Linguistic Inquiry and Word Count. *The American Journal of Psychology, 120*, 263-286.

Kaufer, D., Ishizaki, S., Ishizaki, K., & Butler, B. (2000). *Docuscope: An Environment for Document Production and Analysis.* Pittsburgh, PA: Carnegie Mellon University.

Knowles, E. S., & Condon, A. C. (1999). Why people say yes: A dual-process theory of acquiescence. *Journal of Personality and Social Psychology, 77*, 379-386.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51*, 481-537.

Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.), (2007). *LSA: A road to meaning*. Mahwah, NJ: Erlbaum.

Langacker, R. (1991). *Concept, Image, and Symbol*. Berlin, Germany: Mouton de Gruyter

Langleben, D. D., Schroeder, L., Maldjian, J. A., Gur, R. C., McDonald, S., Ragland, J., O'Brien, C. P., & Childress, A. R. (2002). Brain activity during simulated deception: An event-related functional magnetic resonance study. *Neuroimage, 15*, 727-732.

Leal, S., Vrij, A., Mann, S., & Fisher, R. P. (2010). Detecting true and false opinions: The Devil's Advocate approach as a lie detection aid. *Acta Psychological, 134*, 323-329.

Lee, K. (2000). Lying as doing deceptive things with words: A speech act theoretical perspective. *Minds in the making: Essays in honor of David R. Olson*, 177–196.

Lee, K., & Ross, H. J. (1997). The concept of lying in adolescents and young adults: Testing Sweetser. *Merrill-Palmer Quarterly, 43*, 255-70.

Levine, T. R., & McCornack, S. A. (1991). The dark side of trust: Conceptualizing and measuring types of communicative suspicion. *Communication Quarterly, 39*, 325-340.

Lippard, P. V. (1988). "Ask me no questions, I'll tell you no lies,": Situational exigencies for interpersonal deception. *Western Journal of Communication, 52*, 91-103.

Mandler, J., & Johnson, N. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9*, 111-151.

Manning, C. D., & Schütze, H. (2000). *Foundations of Statistical Natural Language Processing.* Cambridge, MA: MIT Press.

Masip, J., Garrido, E., & Herrero, C. (2004). The nonverbal approach to the detection of deception: Judgemental accuracy. *Psychology in Spain, 8*, 48-59.

Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the reality monitoring approach: a review of the empirical evidence. *Psychology, Crime & Law, 11*, 99-122.

Mayo, R., Schul, Y., & Burnstein, E. (2004). "I am not guilty" vs "I am innocent": Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology, 40*, 433-449.

McCarthy, P. M., Dufty, D., Hempelman, C., Cai., Z., Graesser, A. C., & McNamara, D. S. (in press). Evaluating givenness/newness. *Discourse Processes*.

McCarthy, P. M., Lewis, G. A., Dufty, D. F., & McNamara, D. S. (2006). Analyzing writing styles with Coh-Metrix. *In Proceedings of the Florida Artificial Intelligence Research Society International Conference (FLAIRS)* (pp. 764-769), Menlo Park, CA: AAAI Press.

McCarthy, P. M., Myers, J. C., Briner, S. W., Graesser, A. C., & McNamara, D. S. (2009). A psychological and computational study of sub-sentential genre recognition. *Journal for Language Technology and Computational Linguistics, 24*, 23-55.

McCarthy, P. M., Renner, A. M., Duncan, M. G., Duran, N. D., Lightman, E. J., & McNamara, D. S. (2008). Identifying topic sentencehood. *Behavior Research Methods, 40*, 647-664.

McCarthy, P.M., Watanabe S., & Lamkin, T. (in press). The Gramulator: A tool for the identification of indicative linguistic features. In P. M. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution*. Hershey, PA: IGI Global.

McCornack, S. A. (1992). Information manipulation theory. *Communication Monographs, 59*, 1-16.

McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science, 19*, 22–24.

McNamara, D. S., Kintsch, E., Songer, B. N., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from texts. *Cognition and Instruction, 14*, 1-43.

McNamara, D.S., Louwerse, M.M., McCarthy, & P.M. & Graesser, A.C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes, 47*, 292-330.

McNamara, D.S., Ozuru, Y., Graesser, A.C., & Louwerse, M. (2006). Validating Coh-Metrix. In R. Sun & N. Miyake (Eds.), *In the Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 573-578). Austin, TX: Cognitive Science Society.

Meijer, E. H., Verschuere, B., Vrij, A., Merckelbach, H., Smulders, F., Leal, S., Ben-Shakhar, G., et al. (2009). A call for evidence-based security tools. *Open Access Journal of Forensic Psychology, 1*, 1-4.

Meltzer, B. M. (2003). Lying: Deception in human affairs. *International Journal of Sociology and Social Policy, 23*, 61-79.

Mertens, D. M., & Hopson, R. K. (2006). Advancing evaluation of STEM efforts through attention to diversity and culture. *New Directions for Evaluation, 109*, 35-51.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. London, UK: Sage.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Miller, G. A. (1990). Introduction to WordNet: An on-line lexical database. International *Journal of Lexicography, 3*, 235-312.

Min H. C., & McCarthy, P. M. (2009). Assessing the discourse features of Korean scientists writing using contrastive corpus analysis. *In Proceedings of the Florida Artificial Intelligence Research Society International Conference (FLAIRS)*, Menlo Park, CA: AAAI Press.

Moens, M., & Steedman, M. (1988). Temporal ontology and temporal reference. *Computational linguistics, 14*, 15–28.

Monrose, F., & Rubin, A. D. (2000). Keystroke dynamics as a biometric for authentication. *Future Generation Computer Systems, 16*, 351–359.

Munro, R., Bethard, S., Kuperman, V., Tzuyin Lai, V., Melnick, R., Potts, C., Schnoebelen, T., & Tily, H. (2010). Crowdsourcing and language studies: the new generation of linguistic data. *In the Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk* (pp. 122–130), Los Angeles, California.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle?. *Psychological Bulletin, 126*, 247–259.

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.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology, 21*, 337.

Noonan, H. W. *Personal Identity*. New York, NY: Routledge.

Nunez, J. M., Casey, B. J., Egner, T., Hare, T., & Hirsch, J. (2005). Intentional false responding shares neural substrates with response conflict and cognitive control. *Neuroimage, 25*, 267-277.

Ozuru, Y., Best, R., & McNamara, D. S. (2004). Contribution of reading skill to learning from expository texts. In K. Forbus, D. Gentner & T. Regier (Eds.), *Proceedings of the 26th Annual Cognitive Science Society* (pp. 1071-1076). Mahwah, NJ: Erlbaum.

Paivio, A. (1969). Mental Imagery in associative learning and memory. *Psychological Review, 76*, 241-263.

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

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77*, 1296-1312.

Pennebaker, J. W., Mayne, T. J., & Francis, M. E. (1997). Linguistic predictors of adaptive bereavement. *Journal of Personality and Social Psychology, 72*, 863-871.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, Our selves. *Annual Reviews in Psychology, 54*, 547-577.

Pickering, M., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 169-226.

Pickering, M., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*, 105-110.

Piefke, M., Weiss, P. H., Zilles, K., Markowitsch, H. J., & Fink, G. R. (2003). Differential remoteness and emotional tone modulate the neural correlates of autobiographical memory. *Brain, 126*, 650-668.

Port, R., & van Gelder, T. (Eds). (1995). *Mind as Motion*. Cambridge, MA: MIT Press.

Porter, S., & ten Brinke, L (2010). The truth about lies: What works in detecting high-stakes deception? *Legal and Criminological Psychology, 15*, 57-75.

Premack, D. (2007). Human and animal cognition: Continuity and discontinuity. *Proceedings of the National Academy of Sciences, 104*, 13861.

Quan-Haase, A., Cothrel, J., & Wellman, B. (2005). Instant messaging for collaboration: A case study of a high-tech firm. *Journal of Computer-Mediated Communication, 10*, 23-34.

Rahati, A., & Kabanza, F. (2010). Persuasive dialogues in an intelligent tutoring system for medical diagnosis. *In Proceedings of the 10th Annual Intelligent Tutoring Systems International Conference* (pp. 51-61), Berlin, Germany: Springer.

Riessman, C. K. (1993). *Narrative analysis*. Newbury Park, CA: Sage.

Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S.D., & Castiello, U. (2008). How to accurately assess autobiographical events. *Psychological Science, 19*, 772–780.

Schacter, D. L., Normam K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology, 49*, 289.

Schober, M. F. (2005). Conceptual alignment in conversation. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 239-252). New York, NY: Guilford Press.

Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. *Handbook of Discourse Processes*, 123–64.

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. Amsterdam, Netherlands: John Benjamins.

Seymour, T., Seifert, C., Shafto, M., & Mosmann, A. (2000). Using response time measures to assess guilty knowledge. *Journal of Applied Psychology, 85*, 30-37.

Sip, K. E., Roepstorff, A., McGregor, W., & Frith, C. D. (2008). Detecting deception: the scope and limits. *Trends in Cognitive Sciences, 12*, 48–53.

Smith, C. P. (2001). Content analyis and narrative analysis. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology*. Cambridge, UK: Cambridge University Press.

Spärck Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation, 28*, 11–21.

Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., & Woodruff, P. W. R. (2001). Behavioural and functional anatomical correlates of deception in humans. *NeuroReport, 12*, 2849-2853.

Spence, S. A., Hunter, M. D., Farrow, T. F. D., Green, R. D., Leung, D. H., Hughes, C. J, & Venkatasubramanian, G. (2004). A cognitive neurobiological account of deception: Evidence from functional neuroimaging. *Philosophical Transactions of the Royal Society B: Biological Sciences, 359*, 1755-1762.

Spivey, M. J. (2006). *Continuity of Mind*. New York, NY: Oxford University Press.

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science, 15*, 207-211.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences, 102*, 10393-10398.

Spivey, M., Richardson, D., & Dale, R. (2008). Movements of eye and hand in language and cognition. In E. Morsella & J. Bargh (Eds.), *The Psychology of Action, Vol. 2*. New York, NY: Oxford University Press.

Sporer, S. L. (1997). The less traveled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology, 11*, 373-397.

Steller, M., & Kohnken, G. (1989). Criteria-based content analysis. In D. C. Raskin (Ed.), *Psychological Methods in Criminal Investigation and Evidence*. New York, NY: Springer-Verlag.

Sternberg, S. (1969), The discovery of processing stages: Extension of Donders' Method. *Acta Psychologica, 30*, 276-315.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum.

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). The general inquirer: A computer approach to content analysis. Cambridge, MA: MIT Press.

Sweetser, E. E. (1987). *The definition of lie. Cultural models in language and thought*. Cambridge, UK: Cambridge University Press.

Talwar, V., Gordon, H. M., & Lee, K. (2007). Lying in the Elementary School Years: Verbal Deception and Its Relation to Second-Order Belief Understanding. *Developmental Psychology, 43*, 804.

Taslitz, A. E. (2006). Wrongly accused: Is race a factor in convicting the innocent?. *Ohio State Journal of Criminal Law, 4*, 121.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53*, 1-25.

Underwood, B. J., & Schulz, R. W. (1960). *Meaningfulness and verbal learning*. Philadelphia, PA: Lippincott.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General, 132*, 331-350.

Van Orden, G. C., & Paap, K. R. (1997). Functional neuroimages fail to discover pieces of mind in the parts of the brain. *Philosophy of Science, 64*, S85-S94.

Vendemia, J. M. (2005). Reaction time of motor responses in two-stimulus paradigms involving deception and congruity with varying levels of difficulty. *Behavioural Neurology, 16*, 25-36.

Vendemia, J. M., & Buzan, R. F. (2004). Neuronal mechanisms of deception and response congruity in a visual two-stimulus paradigm involving autobiographical information. *Psychophysiology, 40*, 25-36.

Vendemia, J. M., Buzan, R. F., & Green, E. P. (2005). Practice effects, workload, and reaction time in deception. *The American Journal of Psychology, 118*, 413-429.

Verschuere, B., Prati, V., & de Houwer, J. (2009). Cheating the lie detector: Faking in the Autobiographical Implicit Association Test. *Psychological Science, 20*, 410-413.

Vogeley, K., & Fink, G. R. (2003). Neural correlates of the first-person perspective. *Trends in Cognitive Science, 7*, 38-42.

Vrij, A. (2001). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. Chichester, UK: Wiley.

Vrij, A. (2004). Why professionals fail to catch liars and how they can improve. *Legal & Criminological Psychology, 9,* 159-181.

Vrij, A. (2005). Criteria-Based Content Analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law, 11*, 3-47.

Vrij, A., Akehurst, L., & Morris, P. (1997). Individual differences in hand movements during deception. *Journal of Nonverbal Behavior, 21*, 87-102.

Vrij, A., Edward, K., Roberts, K., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior, 24*, 239-263.

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences, 10*, 141.

Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the liars: Toward a cognitive lie detection approach. *Current Directions in Psychological Science, 20*, 28-32.

Vrij, A., & Heaven, S. (1999). Vocal and verbal indicators of deception as a function of lie complexity. *Psychology, Crime, & Law, 5*, 203–215.

Vrij, A., & Mann, S. A. (2004). Detecting deception: The benefit of looking at a combination of behavioral, auditory, and speech content related cues in a systematic manner. *Group Decison and Negotiation, 13*, 61-79.

Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior, 32*, 253-265.

Walczyk, J. J., Mahoney, K. T., Doverspike, D., & Griffith-Ross, D. A. (2009). Cognitive lie detection: Response time and consistency of answers as cues to deception. *Journal of Business and Psychology, 24*, 33-49.

Walczyk, J. J., Roper, K. S., Seemann, E., & Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: response time as a cue to deception. *Applied Cognitive Psychology, 17*, 755-774.

Walczyk, J. J., Schwartz, J. P., Clifton, R., Adams, B., Wei, M., & Zha, P. (2005). Lying person-to-person about life events: A cognitive framework as a cue to deception. *Personnel Psychology, 17*, 755-774.

Wason, P., & Johnson-Laird, P. (1972). *Psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.

Weinberger, S. (2010). Intent to deceive? *Nature News, 465*, 412-415.

Wiener, M., & Mehrabian, A. (1968). *Language within language: Immediacy, a channel in verbal communication.* New York, NY: Appleton-Century-Crofts.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques (Second Edition)*. Boston, MA: Morgan Kaufmann.

Woodruff, G., & Premack, D. (1979). Intentional communication in the chimpanzee: The development of deception. *Cognition, 7*, 333-362.

Wundt, W. (1874). *Grundzuge der Physiologischen Psychologie*. Leipzig, Germany: Engelmann.

Ybarra, O., Winkielman, P., Yeh, I., Burnstein, E., & Kavanagh, L. (2010). Friends (and sometimes enemies) with cognitive benefits: What types of social interactions boost executive functioning. *Social Psychology and Personality Science, 2*, 253-261.

Zhou, L. (2005). An empirical investigation of deception behavior in instant messaging. In *IEEE Transactions on Professional Communication, 48*, 147-160.

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.

Zhou, L., & Lutterbie, S. (2005). Deception Across Cultures: Bottom-Up and Top-Down Approaches. *Intelligence and Security Informatics, 3495*, 3-14.

Zipf, G. K. (1948). *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge, UK: Addison-Wesley.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 14, pp. 1–59). New York, NY: Academic Press.

APPENDICES

Appendix A. Question stimuli with the stem *Have you ever…* and 120 autobiographical format completions.

1) Have you ever been…
…parachuting? …backpacking? …bowling? …canoeing? …hungover? …mugged? ...married? …quarantined?

2) Have you ever been to…
…Disneyland? …Europe? …Hawaii? …a zoo? …a park? …a supermarket? …a restaurant? …a circus? …Asia? …an ER?

3) Have you ever been on…
…TV? …an elephant? …a camel? …a horse? …a motorcycle? …the radio? ...medication? …a cruise?

4) Have you ever eaten…
…buffalo? …octopus? …insects? …spaghetti? …a hamburger? …breakfast? …artichokes? …apricots? …cheese?

5) Have you ever gone…
…skiing? …surfing? …fishing? …hunting? …sailing? … snowboarding? …paragliding?

6) Have you ever made…
…cheesecake? …gumbo? …sushi? …toast? …cookies? …a dress? …a scrapbook? …a website? …a joke?

7) Have you ever slept…
in a cave? … in a tent? …on a bed? …on a train? …in Memphis? …in Chicago? …on a plane? …in a hammock? …in a barn?

8) Have you ever seen…
…a tornado? …the Simpsons? …an ostrich? …the ocean? …a movie? …the Olympics? ...Stonehenge?

9) Have you ever met…
…Oprah? …Elvis? ...a professor? ...a mayor? ...a neighbor?

10) Have you ever built…
...an igloo? …a snowman? …a sandcastle? …a cabinet? …a fire? …a computer? …a robot?

11) Have you ever tried…
…knitting? ...origami? …paintball? …snorkeling? …tofu? …to waltz? …rollerblading? …skateboarding? …basketball? …surfing?

12) Have you ever collected…
…antiques? …butterflies? …clocks? …coins? …knives? …spoons? …stamps? …rocks? …dolls?

13) Have you ever played…
…backgammon? …checkers? …the tuba? …the violin? …dominoes? …poker? …Monopoly? …Scrabble?

14) Have you ever bought…
…a house? …a car? ...a CD? ...a computer? …a piano? …milk? …a sofa?

Appendix B. Example layout for experimental design.

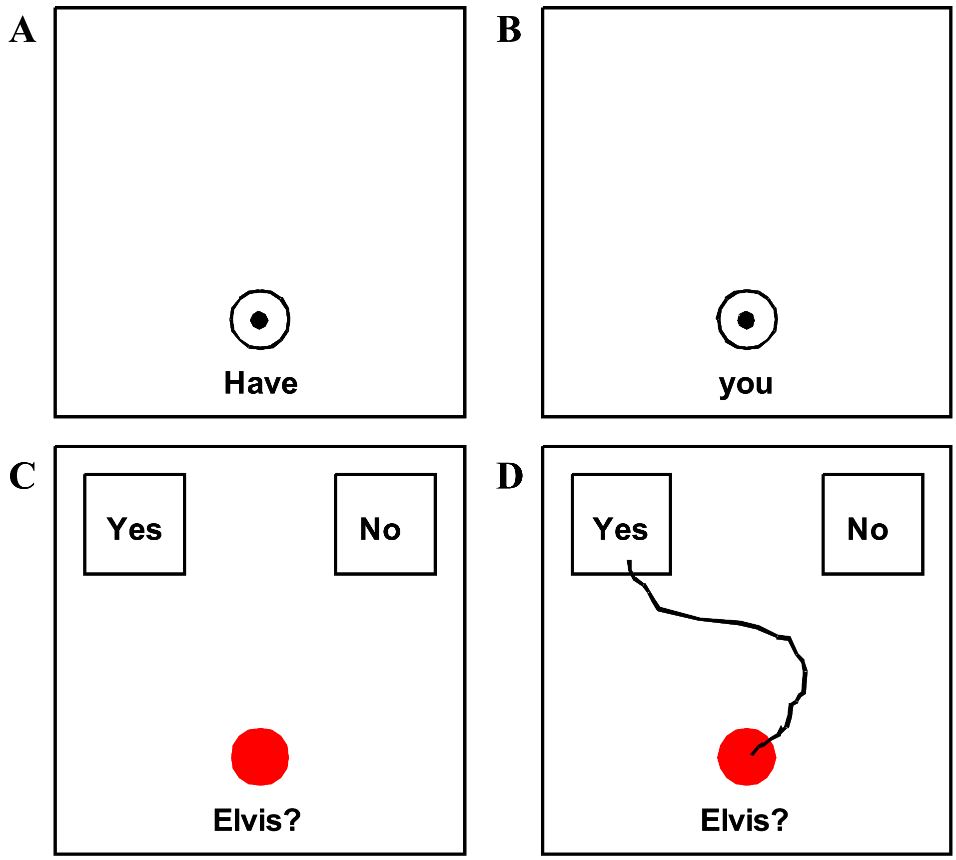

Figure B1. Sample trial sequence for the question "Have you ever met Elvis?". In panel (A), the first word of the question appears. In panel (B), the next word of the question appears after participants click on the bull's-eye-shaped circle. In panel (C), the last word is encountered and the bull's-eye turns to red and response options appear. In panel (D), a sample trajectory shows a participant falsely responding "Yes".

Appendix C. Question stimuli for Study 2c: semantic knowledge.

1a) Is Moscow the capital of Russia?
1b) Is Moscow the capital of England?

2a) Did Shakespeare write the play Hamlet?
2b) Did Shakespeare write the play Rent?

3a) Is insomnia the inability to sleep?
3b) Is insomnia the inability to eat?

4a) Is Wimbledon an event associated with the sport tennis?
4b) Is Wimbledon an event associated with the sport hockey?

5a) Do cheetahs run faster than humans?
5b) Do humans run faster than cheetahs?

6a) Is "lamb" the name for a young sheep?
6b) Is "lamb" the name for a young horse?

7a) Is treason the crime in which people betray their country?
7b) Is treason the crime in which people betray their spouse?

8a) Did Einstein propose the theory of relativity?
8b) Did Einstein propose the theory of evolution?

9a) Is the unit of sound intensity called a decibel?
9b) Is the unit of sound intensity called a watt?

10a) Are people that make maps called cartographers?
10b) Are people that make maps called cardiologists?

11a) Is the largest flightless bird called an ostrich?
11b) Is the largest flightless bird called a crow?

12a) Is the longest river in South America called the Amazon?
12b) Is the longest river in South America called the Nile?

13a) Is the villainous captain in the story "Peter Pan" named Hook?
13b) Is the villainous captain in the story "Peter Pan" named Jafar?

14a) Was the first U.S. President named Washington?
14b) Was the first U.S. President named Lincoln?

15a) Is the legendary one-eyed giant in Greek mythology named Cyclops?
15b) Is the legendary one-eyed giant in Greek mythology named Samson?

16a) Is the skirt worn by men in Scotland called a kilt?
16b) Is the skirt worn by men in Scotland called a sarong?

17a) Is the name of Dorothy's dog in the 'Wizard of Oz" named Toto?
17b) Is the name of Dorothy's dog in the 'Wizard of Oz" named Yeller?

18a) Are dried grapes called raisins?
18b) Are dried grapes called figs?

19a) Is the thick layer of fat on a whale called blubber?
19b) Is the thick layer of fat on a whale called wool?

20a) Was the unsinkable ship that hit an iceberg called the Titanic?
20b) Was the unsinkable ship that hit an iceberg called the Mayflower?

21a) Are dried plums called prunes?
21b) Are dried plums called apricots?

22a) Is deer meat called venison?
22b) Is deer meat called beef?

23a) Is the cartoon character that eats spinach named Popeye?
23b) Is the cartoon character that eats spinach named Garfield?

24a) Is the horse-like animal with black and white stripes called a zebra?
24b) Is the horse-like animal with black and white stripes called a donkey?

25a) Is Paris the capital of France?
25b) Is Paris the capital of Germany?

26a) Are doctors that specialize in diseases of the skin called dermatologists?
26b) Are doctors that specialize in diseases of the skin called radiologists?

27a) Is the process by which plants make their food called photosynthesis?
27b) Is the process by which plants make their food called inhalation?

28a) Is the artist that painted the Sistine Chapel's ceiling named Michelangelo?
28b) Is the artist that painted the Sistine Chapel's ceiling named Picasso?

29a) Is "chick" the name for a young bird?
29b) Is "chick" the name for a young fish?

30a) Is "Old Faithful" located in the national park called Yellowstone?
30b) Is "Old Faithful" located in the national park called Denali?

Appendix D. Computer-mouse movements in the laboratory, with tests of continuity, examination of practice effects, and discriminant analysis (supplementary study to Study 1b)

Computer-mouse Movement (Laboratory)

Participants

Thirty-seven undergraduate students participated for extra credit in an introductory psychology class. Only native English speakers with normal-to-corrected vision were eligible to participate. Of those who participated, 18 were female and 9 were male. All participants performed the experimental task with their right hand.

Procedure and Question Stimuli

The procedure was identical to Study 1a; however, instead of using a Nintenedo Wiimote to collect data, a computer-mouse cursor was used. The sampling rate for the mouse movement trajectories averaged 42 Hz. These recordings, as well as the stimuli presentation, were conducted using Matlab software with the Psychophysics Toolbox (Brainard, 1997).

The same question set as used in Study 1a (see Appendix A) was also used here. It should also be noted that the questions were selected to elicit an equal number of false “no” responses, true “no” responses, false “yes” responses, and false “no” responses. A pilot study was conducted to confirm that the responses were equally distributed within our target population of undergraduate college students.

Summary of Data Analysis

A series of analyses were conducted on the mouse cursor trajectories for the four possible response conditions: a) false answers while making “yes” responses (hereby referred to as “false/yes” responses), b) true answers while making “yes” responses (i.e., “true/yes” responses), c) false answers while making “no” responses (i.e., “false/no” responses), and d) true answers while making “no” responses (i.e., “true/no” responses). The analyses are categorized into six complementary dependent measures presented below. Each area is relevant to identifying signatures of motor movement that are hypothesized to distinguish the continuous true and false response patterns. As a brief summary before further explanation, the first analysis examines the shape of each trajectory within the possible movement space, the second analysis examines the location of the trajectory over time, the third analysis validates the hypothesis of continuous trajectory movements, the fourth analysis identifies various properties that characterize the trajectories, the fifth analysis uses the trajectory properties to predict response the four possible prompt/response categories, and the sixth analysis evaluates whether the trajectory variables are affected by practice effects.

Results

Trials were excluded if responses in the verification task were incongruent with the original responses. Two participants were removed because incongruent responses exceeded 50% of total trials. Among the remaining participants, 40 trials were removed, or 1.39% of the trials for the included participants. Trials were also excluded if *total time* was above 3 standard deviations. This exclusion criterion removed 55 trials, or an additional 1.94% of the trials. Of the 2784 trials remaining, 639 trials occurred in the “false/no” condition, 726 trials occurred in the “truth/no” condition, 728 trials occurred in the “false/yes” condition, and 691 trials occurred in the “truth/yes” condition.

The data for the subset of participants who viewed the reversed position “YES” and “NO” response boxes did not differ from the larger group of participants. As such, the trajectories for the subset were mirror-reversed and combined with larger group to form one dataset.

Trajectory Shape

The first analysis examines the shape of each trajectory taken from the bottom-center bull's-eye to the final response at the top left or top right of the screen. To conduct this analysis, the trajectories for each participant were initialized to x, y coordinates (0, 0) and interpolated to 101 time steps (see Dale et al., 2007; Spivey et al., 2005). In this way, each trajectory was *time-normalized* to the same number of x, y coordinate positions across all trials. Normalization allows the trajectories to be averaged for each participant and for comparison across the four combinations of prompt (true vs. false) and response (yes vs. no) conditions. For the statistical comparison between conditions, we performed paired t-tests at corresponding x-coordinate time steps, resulting in 101 t-tests with 35 observations (i.e., participants) in each response group.

The trajectories for "false/no" trials and "true/no" trials statistically diverged for 27 time steps ($p < .05$) between the 47th and 73rd steps, while the trajectories for "false/yes" and "true/yes" statistically diverged for 71 time steps ($p < .05$) between the 5th and 75th steps (see Figure D1). The divergence for each comparison exceeds the minimum number of 8 consecutive time steps reported as a standard for statistical significance (Dale et al., 2007). Accordingly, the *trajectory shape* analysis reveals false and true responses that are conspicuously different. True response movements appear to travel a more direct route to the target response, whereas false response movements take a more curved route. Not surprisingly, the bend of the curve is always in the direction of the competing response option (i.e., the "true" response).

Given that the false responses reveal the greatest competition, we compared the false "no" and "yes" responses to determine whether there is any further differentiation. Figure D1 suggests that the "false/yes" responses exhibit greater pull towards the truth region than "false/no" responses. To conduct this analysis, the trajectories for "false/no" responses were mirror-reversed so the trajectories were in the same region of movement space as "false/yes" responses. Paired t-tests were conducted once again at each of the 101 time steps. The analysis revealed that the "false/yes" responses were indeed closer to the truth region than the "false/no" responses for 16 time steps ($p < .05$) between the 40th and 55^{th} time steps. Thus, "false/yes" responses appear to be the most susceptible to the influences of the truth.

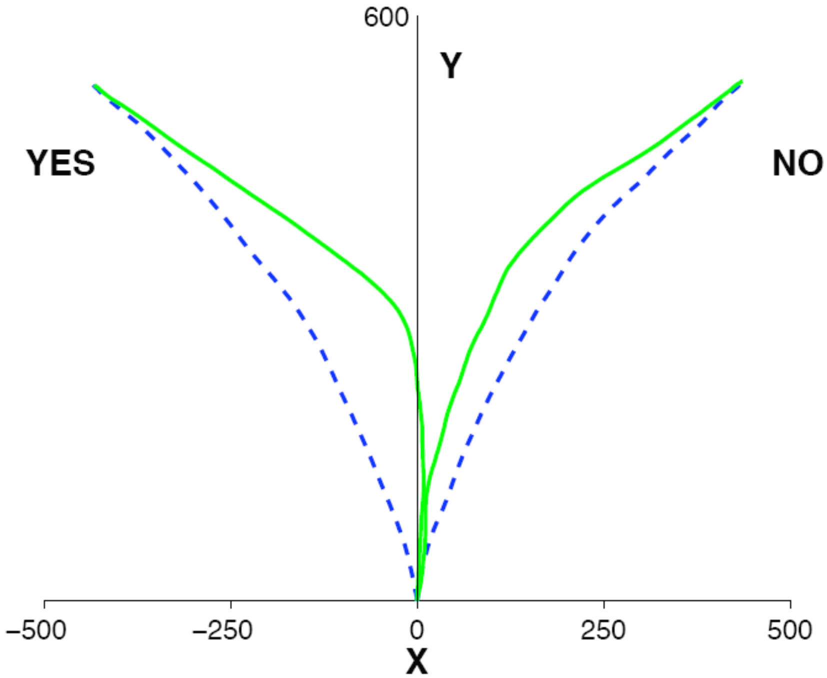

Figure D1. Time-normalized computer-mouse trajectories for each condition. False answers (solid lines) display a greater arc towards the competing response option than true answers (dashed lines).

Trajectory Location

The second analysis examines the location of the trajectory over time. To conduct this analysis, the trajectory coordinates were normalized to initiate at *x, y* coordinate (0, 0) and finalize at *x, y* coordinate (1, 1). Unlike the first analysis that normalized temporal change, this analysis is *space-normalized* and preserves the temporal organization of the trajectory. As the trajectory traveled along the *x*, *y* coordinates in the normalized range, these coordinates were captured in temporal time bins of 0 to 500 ms, 500 to 1000 ms, and 1000 to 1500 ms (see Figure D2). For statistical comparison between conditions, the x-coordinate positions within each time bin were submitted to a repeated-measures

ANOVA. The goal was to determine if the greater competition for false “no” and “yes” responses was consistent across time.

The 2 (prompt type: true vs. false) x 3 (time bin: 0-500 ms vs. 500-1000 ms vs. 1000-1500 ms) repeated-measures ANOVA revealed that for “no” trials there was a statistically significant effect for prompt type, $F(1, 34) = 26.94$, $p < .001$, a statistically significant effect for time bin, $F(2, 68) = 145.94$, $p < .001$, and a statistically significant interaction between prompt type and time bin, $F(2, 68) = 7.90$, $p = .001$. To explore this interaction further, planned comparisons were conducted between false and true prompt types at each time bin. There was a statistically significant difference of the x-coordinate position at the first time bin (0-500 ms) between true ($M = .24$, $SD = .22$) and false ($M = .13$, $SD = .21$) answers, $F(1, 34) = 12.22$, $p = .001$, at the second time bin (500-1000 ms) between true ($M = .55$, $SD = .28$) and false ($M = .30$, $SD = .31$) answers, $F(1, 34) = 26.88$, $p < .001$, and at the third time bin (1000-1500 ms) between true ($M = .72$, $SD = .24$) and false ($M = .51$, $SD = .27$) answers, $F(1, 34) = 15.02$, $p < .001$.

A similar pattern of results was found for “yes” trials. There was a statistically significant effect for prompt type, $F(1, 34) = 78.77$, $p < .001$, a statistically significant effect for time bin, $F(2, 68) = 159.90$, $p < .001$, and a statistically significant interaction between prompt type and time bin, $F(2, 68) = 3.90$, $p = .025$. Planned comparisons between false and true prompt types at each time bin revealed statistically significant differences at the x-coordinate position for the first time bin (0-500 ms) between true ($M = -.34$, $SD = .23$) and false ($M = -.08$, $SD = .19$) answers, $F(1, 34) = 33.02$, $p < .001$, at the second time bin (500-1000 ms) between true ($M = -.64$, $SD = .27$) and false ($M = -.28$, $SD = .31$) answers, $F(1, 34) = 86.57$, $p < .001$, and at the third time bin (1000-1500 ms) between true (M = -.75, SD = .21) and false (M = -.49, SD = .28) answers, $F(1, 34) = 35.51$, $p < .001$. These results indicate that the x-coordinate position for false answers is closer to the competing response options (“yes” or “no) than it is for true answers. It also appears that this competition is consistent throughout the decision process.

As in the first analysis, we can again compare the false responses for just the “no” and “yes” responses. The trajectories for “false/no” responses were mirror-reversed to be comparable with the “false/yes” responses. Paired t-tests were conducted at each time bin for the false responses. Although the “false/yes” responses were consistently closer to the truth region than the “false/no” region, the tests were not statistically significant.

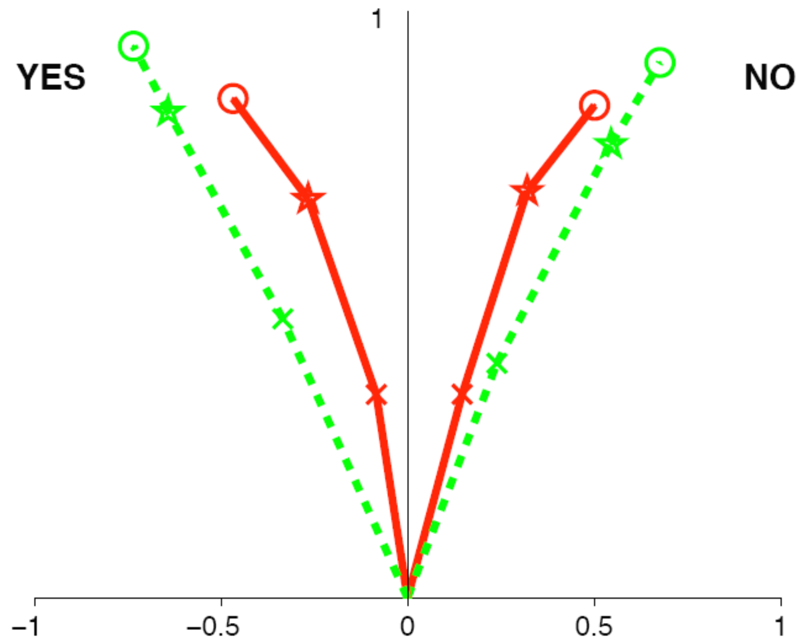

Figure D2. Space-normalized computer-mouse trajectories for each condition. The x-coordinate position is plotted at 500 ms (cross), 1000 ms (star), and 1500 ms (circle). The false answers (solid line) are slower and closer to the competing response option than true answers (dashed line).

Trajectory Continuity

The third analysis validates the hypothesis of continuous trajectory movements. Following the work of Spivey et al. (2005), we evaluated whether continuous or discrete movements predominate trajectory behavior in decision tasks. In related studies, trajectory movements toward a target answer are considered to be simultaneously influenced by a competitor response option. The resulting movement appears to gravitate towards the competitor as a smooth arc en route to the target. Alternatively, the observed continuity can also be explained as the averaging over multiple distributions of discrete, sequential movements. If so, the illusion of continuity would occur for the following trajectory behavior: a) direct movements toward target false responses, summed with b) direct movements toward competitor truth responses before switching to target false responses. To test the possibility of a bimodal distribution, we calculated the area of each trajectory (in pixels) from a hypothetical straight line plotted from the initial and final x, y trajectory coordinates. The area between these two lines serves as a measure of trajectory curvature. The trajectory areas for each participant were then converted to z-scores and pooled across participants. If there were two distinct trajectory curvatures, the distribution of z-scores would reveal evidence of bimodality. We submitted the distributions of z-scores to various tests of normality, including significance testing of a bimodality coefficient and the Kolmogorov-Smirnov test.

Figure D3a shows the distributions of the z-scores of trajectory area for "false/yes" trials (*kurtosis* = .49, *skewness* = -.40) and "true/yes" trials (*kurtosis* = 1.22, *skewness* = -1.13). Figure D3b shows the distributions of z-scores for "false/no" trials (*kurtosis* = .11, *skewness* = -.29) and "true/no" trials (*kurtosis* = 1.16, *skewness* = -.27). These distribution statistics all appear to fall within the range of unimodal normality.

Following Spivey et al. (2005), we tested for bimodality using the bimodal coefficient (Darlington, 1970; DeCarlo, 1997). The coefficient for the "false/yes" trials was .33 and the coefficient for "true/yes" trials was .54. Both bimodality coefficient values are below the .55 critical cut-off point. The coefficient for the "false/no" trials was .35 and the coefficient for the "true/no" trials was .26. Again, both bimodality coefficients are below the critical cut-off point.

We also used a Kolmogorov-Smirnov test of normality to compare the distributions of false and true answers for each category of “yes” and “no” responses. We tested the null hypothesis that distributions for false and true answers are drawn from similar populations. We have no theoretical reason to suspect bimodal distributions for truthful answers, meaning that truthful answers are likely unimodal. A Kolmogorov-Smirnov test will fail to reject the null hypothesis when false answers are also unimodal. Indeed, the Kolmogorov–Smirnov test failed to reject the null hypothesis when comparing false and true “yes” responses, $\chi^2 = .08$, $p = .07$, and when comparing false and true “no” responses, $\chi^2 = .05$, $p = .50$. As such, the distributions for false answers appear to be as unimodal as for truth answers.

It should be noted that the marginal statistical significance of the “yes” responses in the Kolmogorov–Smirnov test, as well as the high “true/yes” value in the bimodal coefficient test, can both be explained by the large kurtosis of “true/yes” distributions. Both tests are adversely affected by high kurtosis. However, high kurtosis in the present data confirms that distributions are concentrated over a single movement pattern. This concentration lends further evidence for unimodal distributions of false answers.

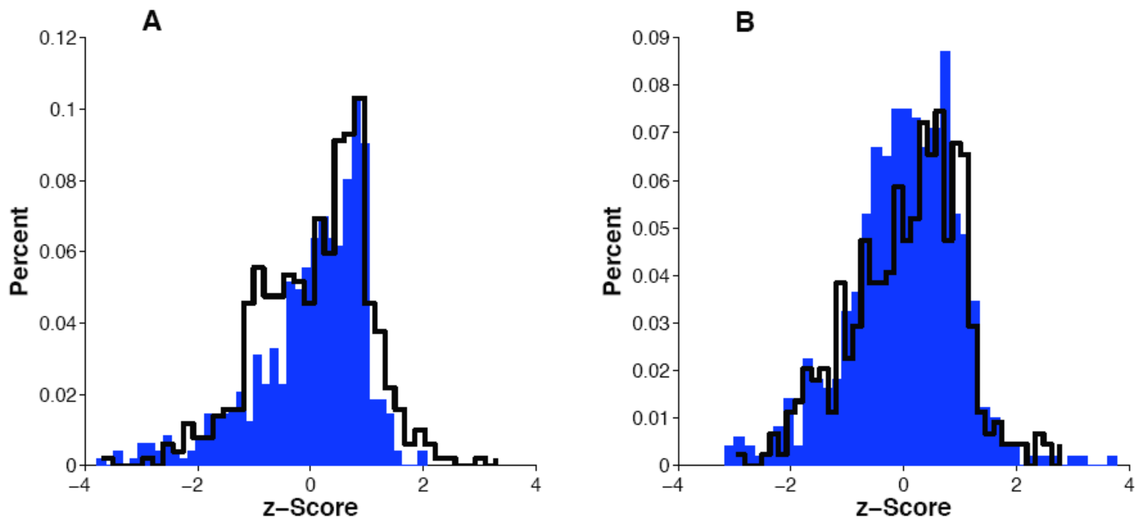

Figure D3. The computer-mouse *z*-score distributions of the area between the trajectory and a straight line from the trajectory beginning and end points. For (A), the solid gray distribution represents "true/yes" answers and the black outline represents "false/yes" answers. For (B), the solid gray distribution represents "true/no" answers and the black outline represents "false/no" answers.

Trajectory Properties

The fourth analysis identifies various *properties* that characterize trajectory behavior. We computed eight properties along continuous scales of measurement (Dale et al., 2007). These property variables were analyzed in a repeated-measures ANOVA to test differences as a function of prompt (true vs. false) and response (yes vs. no) conditions. The trajectory characteristics are the same as those used in Study 1a: *Total Time*, *Latency*, *Distance*, *Motion Time*, *High x-value*, *Low x-value*, *x-flips in Latency*, and *x-flips in Motion*.

The *latency* variable is analogous to *reaction time* in standard response tasks and is not strictly considered a "trajectory" variable. The remaining variables capture dynamical processes that occur at an order of magnitude smaller than reaction time, except for *total time* that captures the entire dynamical response motion. For example, *x-flips in latency* and *x-flips in motion* provide an intuitive measure of moment-to-moment disorder during reaction time, and *high x-value* and *low x-value* variables are indicators of competing attractor strengths that occur while reacting either to "yes" falsely (e.g., moving rightward on the x-axis toward the competing "NO" response region) or to "no" falsely (moving leftward on the x-axis toward the competing "YES" response region).

A 2 (prompt type: truth vs. false) x 2 (response type: yes vs. no) repeated-measures ANOVA was conducted for each of the seven dependent variables. Each variable, as well as their mean values and SEs for each condition, are provided in Table D1. The results of the repeated measures ANOVA are provided in Table D2.

Table D1. *Means and SEs for the computer-mouse trajectory variables by prompt and response type.*

	Yes				*No*			
	False		*True*		*False*		*True*	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Total time (ms)	2146.30	478.62	1537.98	387.92	2173.52	454.53	1749.48	423.31
Latency (ms)	741.99	440.60	530.46	325.25	761.07	464.06	622.01	386.20
Distance (pixels)	1068.56	356.76	855.89	197.33	1065.87	326.93	920.14	266.85
Motion time (ms)	1404.30	393.61	1007.52	231.26	1412.45	356.65	1127.46	288.76
High x-value	157.43	109.96	71.44	65.05	502.62	27.86	502.25	32.87
Low x-value	-495.74	28.73	-499.65	27.04	-145.45	103.16	-98.66	80.72
x-flips in latency	1.57	2.62	1.27	2.26	1.45	2.33	1.50	2.54
x-flips in motion	1.64	0.86	1.19	0.66	1.66	0.88	1.35	0.70

Table D2. *F scores for the repeated measures analysis using computer-mouse movements.*

Variable	*Yes vs. No Response*	*Truth vs. False Prompt*	*Response x Prompt*
Total time	8.62**	132.44**	6.51*
Latency	5.09*	31.56**	2.73
Distance	1.26	32.85**	1.50
Motion time	3.01	74.49**	2.80
High x-value	890.79**	30.76**	30.32**
Low x-value	803.87**	9.00**	14.28*
x-flips in latency	0.71	3.34	5.54*
x-flips in motion	1.07	38.01**	0.72

Note: * indicates statistical significance at $p < .05$; ** indicates statistical significance at $p < .001$; the degrees of freedom for all analysis are 1, 34.

Overall, the trajectory properties reliably differentiated false and true answers. The main effects for all variables indicate that truthful answers (compared to the false answers) took more time overall, took more time to initiate, traveled a greater distance, took more time while in motion, and had a greater number of x-flips while in motion.

The four statistically significant interactions between prompt and response type also provided additional analysis into the difference of false and true answers while making either “yes” or “no” responses. The results of these planned comparisons are reported in Table D3. For “yes” responses, false answers compared to true answers had statistically larger values on all trajectory variables except *low x-value*. The lack of an interaction for *low x-value* was expected because the “yes” option was located in the region of lowest x-values, and thus all “yes” responses had a low x-value. However, even though false “yes” responses had a low x-value, the interaction with *high x-value* also demonstrates that false “yes” responses were more likely to move toward the competing true “no” option than were the true “yes” responses. For “no” planned comparison results, false answers compared to true answers had statistically larger values on all trajectory variables except *x-flips in latency* and *high x-value*. Again, the lack of an interaction for *high x-value* was expected because the “no” option was located in the region of highest x-values, and thus all “no” responses had a high x-value. And like the previous results, even though false “no” responses had a high x-value, the interaction with *low x-value* also demonstrates that false “no” responses were more likely to move toward the competing true “yes” option than were the true “no” responses.

Table D3. *Mean value differences of false minus true responses for variables with a statistically significant prompt (truth vs. false) x response (yes vs. no) interaction (computer-mouse analysis).*

	Yes	*No*
Variable	*False - True*	*False - True*
Total time	608.32**	424.04**
High x-value	85.99**	0.37
Low x-value	3.91	46.79**
x-flips in latency	0.30**	-0.05

Note: * indicates statistical significance at p < .05,
** indicates statistical significance at p < .01;
the degrees of freedom for all analysis are 1, 34.

Discriminant Analysis

The fifth analysis supplements the previous analysis by evaluating the importance of trajectory properties and reaction time in differentiating the categories of “false/no”, “true/no”, “false/yes”, and “true/yes” responses. To do so, we used a discriminant analysis (DA) to simultaneously predict the four prompt/response categories. The DA provides estimates of how well dynamical trajectory variables and combinations of trajectory variables contribute to the overall analysis. Moreover, the primary value of a DA here is the ability to compare the prediction accuracy of trajectory variables to the accuracy of discrete variables like reaction time. On the one hand, reaction time could potentially subsume the trajectory properties, therefore obviating the contribution of the trajectory properties in predicting false and true responses. On the other hand, the trajectory properties may capture unique cognitive and behavioral processes that are ignored when reaction time is used as the sole unit of analysis. If the latter is true, the classification accuracy for dynamical trajectory variables will be far superior to the classification accuracy of reaction time.

To conduct a DA based on “dynamical” properties, the variables obtained from the trajectory analysis were used as predictors of category membership. Concerns with multicollinearity prompted the removal of variables that correlated with other variables above a 0.70 threshold. When a highly correlated pair was identified, the variable with the lowest F-score from the repeated-measures ANOVA (see Table D2) was excluded. As a result, the final DA included a reduced set of four predictors: *total time*, *distance*, *high x-value*, *and low x-value*.

Discriminant scores loaded onto two functions that delineated the four prompt/response categories into distinct regions of state space. The objective of the discriminant scores is to maximize between-category differences while minimizing within-category variation. The Wilks’ Lambda test of function discrimination was statistically significant across the two functions, $\Lambda = .169$, $X^2(12) = 4938.65$, $p < .001$, indicating that, overall, the variables in the study predicted category membership.

We used a “one-against-all” classification technique for predicting group membership of each trial. In this technique, a discriminant analysis model is built using all but one trial and the model is then used to predict the excluded trial. This process of exclusion and prediction continues until all trials have been assessed. Prediction success for each category is reported as recall, precision, and F1 scores. The recall scores measure the number of trials that were classified correctly (hits) divided by the total number of trials in each corresponding category (hits + misses). The precision scores measure the number of trials that were classified correctly (hits) divided by the number of correct classifications and misclassifications (hits + false alarms). The F1 score is the harmonic mean between the recall and precision scores.

In general, the accuracy of the classifications greatly exceeded the .25 baseline for each prompt/response category (Table D4). Accuracy was highest for “true/yes” trials, followed by “true/no” trials, followed by “false/yes” trials, and lowest for “false/no” trials.

Table D4. *Recall, precision, and F1 scores for discriminant analysis using computer-mouse.*

	Yes		*No*	
	False	*True*	*False*	*True*
Recall	0.43	0.79	0.36	0.68
Precision	0.57	0.61	0.46	0.59
F1	**0.50**	**0.70**	**0.41**	**0.64**

For the purposes of comparison, we then built a DA that included the four trajectory variables and the *latency* variable that served as an approximate measure of reaction time. The *latency* variable captures the temporal duration between stimuli onset and the initial movement toward a response choice, and does not correlate with the other variables above the critical .70 threshold. This expanded model showed no improvement in the recall, precision, and F1 scores over that of the initial model. Another model was then constructed with latency as the sole classification predictor. The results indicate that classification for "false/yes" and "true/no" categories were far below baseline, while classification for "false/no" and "true/yes" categories were moderately above baseline (Table D5). However, the overall classification accuracy was below the initial model (56.6% vs. 27.5% cases correctly classified). Thus, it appears that the trajectory variables are better at discriminating the four prompt/response categories than a reaction time measure based on the *latency* variable.

Table D5. *Recall, precision, and F1 scores for only latency variable in discriminant analysis (computer-mouse analysis).*

	Yes		*No*	
	False	*True*	*False*	*True*
Recall	0.04	0.66	0.37	0.06
Precision	0.33	0.28	0.26	0.28
F1	**0.19**	**0.47**	**0.32**	**0.17**

Practice Effects

The sixth and final analysis examined the possibility that competition between false and true answers is attenuated with practice. If so, trajectory properties for false answers might shift toward values that characterize true answers during the course of the experimental session. In real-world applications, there is always the possibility that deceivers will become more proficient in responding falsely when given repeated opportunities to do so. To test this possibility, the sequence of trials for each participant was divided into three increasing time sets (first set: trials 1-26; second set: trials 27-53; third set: trials 54-80). For each of the four prompt/response categories, the trajectory properties were compared between trial sets using a repeated-measures ANOVA. The potential for practice effects will be mitigated if there are not significant effects as a function of trial sets.

Based on the repeated-measures ANOVA, only the variable *motion time* was affected by practice effects. For *motion time,* "false/yes" responses decreased from 253.80 ms from the first set to the third set of trials, $F(2, 68) = 4.37$, $p = .02$. However, while false answer trajectories for motion time showed improvements in the direction of true answer trajectories, there remained a statistically significant difference between these false answers with the true answers ($p < .001$). This finding suggests that the response patterns to false and true questions are robust against practice effects, and when practice effects are found, the improvement is negligible. The practice effects analysis also indicates that "false/yes" responses were more affected by practice than "false/no" responses.

Brief Summary

The movements of the arm revealed distinct signatures of movement during false response behavior. These action dynamics were slower and more disorderly than true responses, thus suggesting greater difficulty in cognitive processing. The false response movements were also curved towards a competitor "truth" region that was visually co-present with the target response region. This curvature suggests the presence of a truth-bias attractor during false responding (Gilbert et al., 1993; McKinstry et al., 2008). We also found that the curvature is composed of smooth and graded changes that occur in parallel, rather than a combination of discrete response "modes" that are influenced by separate processing factors. Finally, we found that dynamical trajectory properties predicted false responding with greater accuracy than did a reaction time measure.