The Dangers of Speaking a Second Language:

An Investigation of Lie Bias and Cognitive Load

Andre Dippenaar

DPPAND004

A full dissertation submitted in fulfilment of the requirements for the award of the degree Master of Social Sciences (Psychology)

Applied Cognitive Science and Experimental Neuropsychology Team (ACSENT)

Supervisor: Professor Colin G. Tredoux

Faculty of Humanities

University of Cape Town

2019

COMPULSORY DECLARATION

This work is my own. This work has not been previously submitted in whole or in part, for the award of any degree. Each significant contribution to, and quotation, in this dissertation, from the work(s) of others has been duly attributed, referenced and cited.

Signature: ___Signed by candidate___ Date: _____

Dedication

This thesis is dedicated to my grandparents.

Thank you for teaching me the values of commitment, perseverance and tenacity

Table of Contents

List of Tables

List of Figures

Acknowledgements

This thesis would not have seen the light of day if it had not been for the support and motivation of several people.

To my supervisor Professor Colin G. Tredoux, thank you for your guidance and support throughout this undertaking. From you I have learnt invaluable skills that I will forever cherish. You pushed me to be a better student, researcher and constantly encouraged me to write to the best of my abilities. Second, to my family, specifically my parents (Andrew and Christine Dippenaar), thank you for giving me the opportunity to further my studies. Thank you for all you have done, from the emotional support, for keeping me motivated during those late nights and for the continued understanding and for keeping me on track. Third, to my second family (Uncle Richard Sims, Aunty Catherine Sims, Dante Charlie Sims, Dylan Richard Sims and Robin Solomons) I cannot begin to say how grateful I am to you and your extended family (the Hector's, the Cloete's and the Solomon's). You all provided me with so much love and support, and an endless amount of coffee that saw me through many a late night. Fourth, to my guidance coach, Belinda Metlitzky Silbert, thank you for all your pearls of wisdom, and for keeping me motivated and on track. It is truly appreciated. To my partner, Jennah Louise Dollman, thank you for the love, support, and understanding.

I would be remiss if I also did not acknowledge all the members of the Eye-Witness group and ACSENT laboratory, and both the undergraduate and postgraduate secretaries (Mrs Mia Kariem and Mrs Rosalind Adams respectively) in the Department of Psychology at UCT, for their continued support and encouragement for the duration of this project. I would also like to extend my utmost thanks to all the participants who partook in this research project. Finally, I would like to extend my heartfelt thanks to my reviewers who gave valuable insight into this dissertation which resulted in it being the calibre that it now is.

Abstract

Today's world is an interconnected global village. Communication and business transactions are increasingly conducted in non-native languages. Literature suggests that biases are present when communicating in non-native languages; that a truth bias is present in first language communication, and a lie bias in second language communication. Less than 10% of South Africa's population identifies with English, the lingua franca of the country, as a first language. Not much research in the presence of bias in second language communication has been published in the South African multi-lingual context. This study evaluated the presences of bias within deception frameworks such as the Truth Default State and the veracity effect.

This study investigated whether deception detection can be improved by modifying the conditions under which statements are given by placing statement providers under cognitive load. The accuracy of veracity judgment language profiling software, LIWC2015, using published deception language profiles was compared against the results of the participating veracity judges.

Results of the study were mixed. It was consistent with extant literature in a presence of a truth bias overall, but mixed in terms of a lie bias. The results supported the Truth Default Theory and veracity effect frameworks. LIWC2015 performed marginally better than human judges in evaluating veracity.

Keywords: Lie Bias, Truth Bias, Deception Detection, Cognitive Load, Veracity Assessment.

Chapter 1 Introduction and Background

Deception for the most part is an innocuous daily act of human existence and human interaction and or communication. For the most part, getting away with deceit has limited consequences. There are incidents in which getting away with the deceit or lie would have serious consequences for any society. One example would be an asylum seeker being untruthful in an application for asylum to improve their chances to be granted asylum by a country or getting away with a fraudulent claim against a business. The world is becoming more globalised, and the requirement to communicate in one's non-native tongue is becoming more common. Studies have shown that people tend to communicate truthfully (Levine, 2014; Serota & Levine, 2014). However, a receiver should be cognisant that it could be profitable to the communicator to successfully deceive the receiver of a message or statement.

The act of deceit is not a simple act. To be successful in deceptive communication requires awareness of, and the management of, several verbal and non-verbal indicators of deception whilst communicating with the intent to deceive. Much of the research done to date focused on non-verbal indicators such as facial expressions and physiological responses that includes heart rate and pupil dilation. The use of polygraphs to measure non-verbal indicators of deceit is well known and studied. Other popular methods reported to identify deceit include audio- and visual recordings of interviews and analysis afterwards, looking at the actions of the deceiver for indicators of deceptive communication (Kolkman, 2012). Studies to determine the efficacy of some of these techniques have subsequently been discredited, reporting high occurrences of false-positives and false-negatives (Vrij, 2000).

Past research has found that people are deceitful on a daily basis in their communication with their fellow humans. At a bare minimum it was found that people were deceptive at least twice a day (DePaulo, Kashy, Kirkendol, Wyer, & Epsetin, 1996).

This study asked whether communicating in one's second language does have a negative effect on one's credibility, i.e. that second language communication, regardless of the veracity of the second language statement will preferentially be judged as untruthful. This study also investigated whether the converse holds true; that a statement provided in a first language resulted in a positive impact on one's credibility, i.e. the first language communication will preferentially be judged as truthful. This study investigated whether the presence of cognitive load (an increase in the demand placed on one's finite cognitive resources) whilst an individual was delivering a deceptive statement made it easier to detect the attempt to deceive. Finally, this study looked at whether Natural Language Processing Programmes (NLPs) out performed humans when it came to detecting deception, more specifically, if the software programme Linguistic Inquiry Word Count 2015 could achieve this differentiation.

The approaches used to assess the areas of enquiry of this study were to ask participants to provide a number of statements, some that were aligned with their own view on a topic (truthful) and some that were not aligned with their view on a topic (untruthful, or deceptive). Some of the statements were required in their first language and some in a second language. Some participants provided statements under cognitive load, some with no cognitive load imposed. The effect of language fluency, as represented by first and second language use to provide or assess statements on the credibility of statements were quantified statistically.

A study of this nature is particularly relevant in the South African context. South Africa is a country with eleven recognised official languages. But despite this and the fact that English is only spoken as a first language by 9.6% of the population, it remains the lingua franca (Statistics South Africa, 2012; Khokhlova, 2015). English is the language of government, medicine, commerce and the justice system. If indeed, communicating in one's second language results in a preference for extending a lie bias to the communicator, then approximately 90% of

South Africa's population's credibility could potentially be negatively influenced when communicating in English as a non-native language.

Research Problem

Research reports that people do not fare well in detecting the presence of deception in statements. Several studies have found that on average detection of deceit accuracy rate is not different to a chance outcome (Bond, & DePaulo, 2006; DePaulo et al., 2003; Vrij; 2000; van Vuuren, 2017). As such, various techniques have been put forward to improve the accuracy rates of detecting deceptive communication. Contemporary research finds that communication through a non-native language are often assumed to be deceitful. Research found that this lie bias exists regardless of the ability of the receiver to detect deception (Evans, & Michael, 2014). The presence of this bias is problematic for those who for no fault of their own, have to speak in their non-native language. However, studies conducted by researchers such as Cheng and Broadhurst (2005), Da Silva and Leach (2013) and Lev-Ari and Keysar (2010) has found that that if a person speaks in their native tongue, then they are afforded a truth-bias.

Significance of the study

This study contributes to the growing field of deception detection as well as improves the understanding of the dynamic of interpersonal communication. An improved understanding of the presence of biases, if present, can form the basis of future research to propose countermeasures to be followed to prevent deceptive communication from occurring. The results obtained from this study could be useful to the judicial and the police operations as well as financial institutions

If a proven means can be developed that increases the ability of human evaluators of statements to accurately detect deception, the aforementioned chance outcome typically associated with human evaluation can be improved. Studies have suggested that placing someone

under cognitive load whilst communicating makes it difficult for the communicator to deceive. If this assertion holds, this approach might provide an avenue to improve the accuracy of veracity assessment, in both human evaluators as well as computer-based language analysis programs.

Although much of the existing research indicates that deception is highly prevalent, it is important to understand the potential impact that a second language bias may have on the perception of honesty as it may have disastrous effects on the livelihoods of those who have no choice but to speak and interact in their non-native tongue. Furthermore, this study is of importance because despite the plethora of studies that deal with deception and the detection thereof few studies have been conducted on the notion that certain biases exists and are extended to people under certain instances, and that these bias in turn may hinder the ability to correctly classify deception from honesty in both high-stakes and low-stakes situations. The insight gained from this study shows us that it is not as simple as a proposed by past literature that a lie bias is extended to second language speakers, but in actual fact, this bias is more nuanced. That this bias is only present when second language reviewers are reviewing a second language statements.

Chapters Overview

Chapter 2 Literature Review. This chapter highlights previous research done in the field of deception. It addresses the problem of defining deception. It addresses the known existing cues as well as physiological indicators of deception. The literature review speaks to the various cues which have been used in the past to detect deception, from looking briefly at the physiological cues to looking at the verbal cues used in more contemporary empirical studies. The literature review addresses language processing frameworks characterising linguistic profiles of deceptive communication through the use of algorithms such as LIWC (Linguistic Inquiry Word Count). The literature review addresses studies that looked at the interaction of second language communication and deception assessment.

Chapter 2 ends with a summary of the rationale and the specific areas of enquiry of this study, stating the research hypotheses.

Chapter 3 Methodology. Chapter 3 addresses the methodology used in this empirical study. It describes the approach used for participant recruitment and data collection. The chapter elaborates on the procedure used to generate the statements and the controls deployed to ensure balanced representation across the key variables used for the analysis. The chapter touches on the statistical techniques used for analysis and concludes with the ethical considerations and approval received for the study.

Chapter 4 Results. Chapter 4 presents the results of the analysis conducted. The chapter is organised according to the research questions. Descriptive and inferential statistics, including validation tests for methods used are presented in tables and figures.

Chapter 5 Discussion. This chapter discusses the results derived from the analysis with reference to the extant literature discussed in chapter 2. The limitations identified in this study are listed. Recommendations for further areas of inquiry identified based on the results of this study concludes this chapter.

Chapter 2 Literature Review

This chapter presents a review of the existing literature in the field of deception. It offers a conceptual discussion of deception. It speaks to the cues (verbal and non-verbal) that are often used to detect deception. The chapter reviews several techniques used to detect deceit, paying specific attention to cognitive load and language techniques used to date. The chapter reviews studies focusing on the relationship between second language and deception, and the impact that language proficiency has on veracity assessments. This chapter positions this study within the reviewed literature and ends with the study's main research questions and hypotheses.

What constitutes deception?

Like many constructs and concepts in psychology, the views on what constitutes deception exactly are polarised. There have been a number of attempts to define deception and what constitutes deceptive communication (Masip, Garrido, & Herrero, 2004).

Mitchell (1986) argued that for a statement to be deceptive, it had to be a false statement from which the sender of that message accrued a benefit. Anolli and Ciceri (1997) argued that deception is characterised by two types of communicative intentions, the hidden intention (where the speaker intends to be deceptive, specifically by altering the content of the message) and the manifest intention (where the speaker intends to convincingly communicate with the receiver).

Vrij (2000) and Levine (2014) similarly argue that for a statement to be deceptive, the message has to be a specific and deliberate attempt by the sender to mislead the receiver of the message or, alternatively be conveyed with an intent to foster a sense of belief in the intended target. Masip and colleagues (2004) argue that for a message to be considered deceptive, two characteristics had to be present; (a) on the sender's part there had to be an attempt to conceal the fact that the sender has the intention to deceive, (b) there had to be some attempt at fabrication and/or manipulation of information to facilitate and sustain the deceit. McCornack, Morrison,

Paik, Wisner and Zhou (2014) argued that something is deceptive when there is some form of selective control over which information is disseminated to the receiver.

An earlier study concluded that a deceitful act presents in one of three forms, through the act of lying, the act of omission or through the act of equivocation (Buller, & Burgoon, 1996). Deceit via lying occurs when the message sent consists of a deliberate falsification of information with the intent to mislead the receiver. Deceit via omission means there was a conscious effort on the sender's part to withhold pertinent information and deceit via equivocation occurs when ambiguous language is used to mislead the intended receiver (Buller, & Burgoon, 1996).

The theme that thus emerges in literature is that a deceptive message requires an intention from the sender to deceive the receiver; the sender needs to know, or believe, that the content of the message is false and that sustaining deceptive communication requires ongoing efforts to conceal, fabricate or manipulate information (Mitchell, 1986; Buller, & Burgoon, 1996; Levine, 2014; Massip et al., 2004; McCornack et al., 2014;; Vrij, 2000).

For the purpose of this thesis, the definitions offered by Vrij (2000) and Levine (2014) are adopted - deception occurs when a deliberate attempt is made to manipulate the information provided to the recipient of the message with the intent to mislead the recipient.

The lenses through which deception can be viewed.

Multiple theories have been put forward to explain deception and provide reasons why human evaluators generally are not good at judging veracity. Two theories, pertinent to the empirical study of this thesis are briefly discussed. The first is Interpersonal Deception Theory (IDT) focussing on the act of deception within the processes of communication and the second is Truth-Default Theory (TDT) as it offers a framework in which to assess the possible presence of a truth-bias toward veracity assessment as described in deception literature.

Interpersonal Deception Theory. The IDT framework states deceptive communication is a complicated process with several steps required to deliver and maintain deceptive communication. Deceptive communication occurs between participants who each have their own motives and aims that are not aligned (Burgoon, & Buller, 2014). As deception is an active and purposeful act, both sender and receiver actively engage in the process. Both the sender and receiver of a deceptive message need to engage simultaneously in encoding and decoding of the message (Burgoon & Buller, 2014).

IDT holds that honest statements differ from deceptive statements, and that these differences can be observed (Burgoon & Buller, 2014). This assertion emanates from the fact that the speaker of the message is engaging in both strategic behaviour (intended behaviour) and non-strategic behaviours (unintended behaviours) concurrently. According to Bond, Levine and Hartwig (2014) IDT argues that people engage in both strategic and non-strategic behaviours whilst encoding and decoding messages. Buller and Burgoon (2008) argue that the strategic behaviours occur naturally when it comes to communication and are present since an early age. An example of this early strategic or intended behaviour is the use of gestures when describing how big something is. Non-strategic behaviours, in turn, are beyond conscious control, unintentional and result in behavioural leakage, which may indicate the presence of deception. IDT sees communication has being an interactive and dynamic event between two parties, a sender and receiver, this manipulation of information, places additional cognitive load on both parties (Buller & Burgoon, 1996).

Burgoon, Blair and Strom (2008) argue that a person's veracity assessment of something is influenced by four cognitive biases. The first cognitive bias is the visual bias. The visual bias argues that people rely too heavily on the non-verbal behaviours accompanying the speech of others rather than listening and looking at the language employed by the individuals (Burgoon et al., 2008). The second bias is the demeanour bias. The demeanour bias occurs when receivers

place emphasis on the communication style to make veracity judgements; it is based on the sender's personal character competency (Burgoon et al., 2008). The third bias is described as the expectancy violation bias (van Vuuren, 2017). This bias is concerned with social or cultural norms and is the result of receivers judging a statement as deceptive if it should violate social and cultural norms (Burgoon et al., 2008). The fourth and final cognitive bias which Burgoon et al., (2008) identified is known as truth bias. This is the predisposition of people to see people and their statements as truthful rather than deceptive i.e. receivers tend to judge senders as being innately truthful (Burgoon et al., 2008). Truth bias corresponds with IDT's idea that people expect others to be honest.

When people enter a conversation, IDT proponents argue that they do so under the influence of the truth bias. Burgoon and Buller (2014) argue the reason for successful deception is that receivers do not recognise leakage and behavioural cues. The communication process is a complex process that makes use of heuristics and as such, people only concentrate on information selectively and in such a manner as to confirm their assumptions and expectations (Buller & Burgoon, 1996).

Levine, Park and McCornack (1999) found when someone makes a veracity judgement of a message, the original veracity of the message is a crucial factor. Levine and colleagues (1999) argued that judgement accuracy would increase if a receiver saw or heard a greater number of truthful statements. This was coined the veracity effect by Levine and colleagues (1999). In essence the veracity effect is based on the idea that the original veracity of a message will exert an influence on the veracity assessment of a statement.

Truth-Default Theory. In 2014 a new theory pertaining to the detection of deception, and a new way of understanding why people get away with deception during communication arose. This theory was dubbed Truth-Default Theory (TDT) (Levine, 2014). The fundamental

premise of TDT is that people are more likely to believe others are honest, i.e. they have a truth bias. TDT argues that this bias is highly adaptive as this truth default state, allows for efficient communication and co-operation. However, these proponents do warn that although most of the time communication is honest, this presumption of honesty does make people vulnerable to being deceived (Levine, 2014). The adaptive nature of this default state was already previously suggested by Zukerman, DePaulo, and Rosenthal (1981).

Truth-default theory, as it pertains to this study, has two central constructs with several propositions to the constructs. These are the Truth-Default Theory constructs of the truth-default state and the closely related truth-bias. According to TDT, the truth-default state arises out of a passive assumption of honesty. This state follows from either failing to acknowledge the possibility of deception or relying on a cognitive state in which there is not enough information to confirm deception (Levine, 2014). Levine (2014) further argues, by default, this presumption occurs without conscious reflection and that it is a starting place to make inferences regarding communication.

Levine (2014) states that the possibility of deception does not present itself unless there is a reason to suspect the presence of deceit, and this notion of is founded on the Spinozan model of belief (Gilbert, 1991). In other words, all incoming information, unless actively disproved, are preferentially assessed as truths.

The truth bias can be understood as the tendency of receivers to assume that people are inherently honest at all times, and as such receiver's judge or expect that communication will be honest and truthful, regardless of the actual honesty conveyed in the communication. Central to understanding this bias, are the perceived truth-lie base rates (van Vuuren, 2017). The concept of a truth-lie base rate presumes that the expectations from receivers of messages influences the

outcome of the veracity assessment. Street and Richardson (2015) tested this idea and found that when judges expected a message to be true, the message was preferentially judged as truthful

The betrayers of deception

As early as 1969 Ekman and Friesen argued that deception, when present, could be detected by observation of certain non-verbal cues. Despite this finding, people in general do not fare much better than chance when it comes to correctly determining the presence of deception. The accuracy of correct veracity assessments by people is similar to a chance outcome (DePaulo et al., 2003). When looking at these cues of deceit, empirical research divides these cues into verbal cues and non-verbal cues. Verbal cues include tonality, vocalic pitch and the types of words used, etc. (Zuckerman, & Driver, 2014). Non-verbal cues are associated with the body's response such as galvanic skin conductance or individual responses such gaze avoidance, etc. Empirical research argues that these cues betray the presence of deception by 'leaking' through both the verbal and non-verbal channels (DePaulo, 1994; Ekman, & Friesen, 1969; Ekman, & Friesen, 1974; Ströfer, Noordzij, Ufkes & Giebels, 2015; Strömwall, Granhag, & Hartwig, 2004; Zuckerman, & Driver, 2014).

Body leakage as a cue to deception. Fundamental to this concept of using observation of leakage as an indicator to detect acts of deception, is the premise that despite conscious effort to control this leakage, we are not able to control our non-verbal behaviour with a large measure of success. It follows that the non-verbal behaviour which is leaked could therefore indicate whether the message conveyed is true or deceptive in nature (Zuckerman, & Driver, 2014). Zuckerman and Driver (2014) found that certain channels were more controllable than others were and that our ability therefore to successfully control leakage in the different channels presents on a continuum of control possibilities. Body language and tone of voice fall on the less controllable side of the spectrum, whereas verbal and facial expressions fall on the more controllable side of the continuum (Zuckerman, & Driver, 2014). This idea of body leakage

assumes that the channels that are less under our control are the ones that will betray the presence of deception in the message conveyed.

Evidence to support the idea the idea that deception cues are leaked via bodily channels can be found in some of Ekman and Friesen's earlier work. In 1974, Ekman and Friesen proposed that the source of the leakage of cues associated with deception stemmed more from the body than the face. The authors argued that this was the case for those who grew up in a Westernised society because in a Westernised society people are taught from an early age to control their facial expressions (Ekman & Friesen, 1974). When investigating body leakage, Ekman and Friesen (1974) developed two hypotheses. The first of these was that participants preferentially focused on facial expression cues to identify deception. It follows then that to conceal cues associated with deception, senders have to put the most effort towards controlling facial expression (Ekman & Friesen, 1974). The second hypothesis was that when making judgements concerning the veracity of statements, participants would attain a higher accuracy rates if they based veracity judgements on body rather than the face (Ekman & Friesen, 1974). Ekman and Friesen (1974) asked participants to be honest in one interview after viewing a pleasant film, by describing their true feelings and deceptive in a secondary interview after viewing an unpleasant film. The study fully supported hypothesis one, as participants stated that the face had to be controlled more than the body when being deceptive and hypothesis two was only partially supported (Ekman & Friesen, 1974). In 1994 DePaulo investigated how would could improve the accuracy of deception detection. In contrast to what Ekman and Friesen (1974) argued, she found that deception accuracy was greater if one looked at the bodily cues, such as fidgeting, pupil dilation and so forth, then looking at the face for cues associated with deception.

Studies have argued that there are behavioural cues associated with attempts to mask deception in the presences of active attempts by the deceiver to control non-verbal behaviour.

Because of this increase in attempted control, the deceiver appears too rehearsed and lacks spontaneity in body movements (DePaulo, 1994), and those movements that do occur come across as exaggerated (Zuckerman, & Driver, 2014).

Emotional leakage as a cue to deception. Closely linked to the notion of the body betraying deception by leaking cues of deceit, is the idea that emotions can also betray the fact that deceit is present. In other words, certain emotions, such as fear of being caught or the feeling of guilt for lying in the first place, are incongruent with the lie being told, and are thus leaked through various non-verbal channels (Ekman & Friesen, 1969; Ekman, 2001).

Porter and ten Brinke (2008) sought to determine if emotional leakage occurred when being deceptive. To investigate this the researchers asked participants to view several images designed to bring out different emotional responses, such as happiness, sadness, disgust, anger, irritability etc. In this experiment, the senders were asked to deceive via an emotional expression (i.e. they were asked to express an emotion that they were not feeling) (Porter & ten Brinke, 2008). In the study three types of responses existed in the deceptive condition; (a) a stimulating condition; in which they were asked to fake experiencing an emotion which was not being felt, (b) a neutralising condition, in which they had to display no emotion, and (c) a masking condition, in which they had to fake an emotion contradictory to what they were feeling (Porter & ten Brinke, 2008). Porter and ten Brinke (2008) concluded that emotional leakage did accompany the act of deceit, as emotional leakage was present in each of their study's experimental conditions. Additionally, the result of the study indicated that emotional leakage was more prevalent in the masking condition (Porter & ten Brinke, 2008).

Porter, ten Brinke and Wallace (2012) investigated the impact of emotional intensity on deception and if it would result in emotional leakage occurring. Participants viewed various images of varying degrees of emotional intensity (e.g. disgusting, sad, joyous etc.) and then had

to respond to the images. Five second videos were taken of their facial expressions and frames, spaced each 1/30s analysed. Results supported Darwin's claim that certain facial expressions linked to emotion could not be inhibited (Porter et al., 2012). This study also found high intensity deceptive emotions (emotional states in which it was expected to be greater inconsistencies between what was felt and the resulting facial expressions) were linked to more emotional leakage (Porter et al., 2012)

ten Brinke, Porter and Baker (2012) posited that deception was inherently a part of the communication process, and that it was always accompanied by unfelt emotions or the suppression of emotions which corresponded to the nature of the message. Based on this idea the researchers investigated the consequences of high-stake lies by looking at specific facial muscles (ten Brinke et al., 2012). Participants were exposed to real-life recordings of family members requesting the safe return of missing family members (half of which was deceptive as the person who was pleading was later convicted for murder) (ten Brinke et al., 2012). The facial muscles of the participants were coded frame-by-frame. Findings of the study supported the idea that whilst being deceptive emotional leakage did occur, and found that this emotional leakage stemmed from facial muscles. The study found that masking smiles and failed attempts to appear remorseful was more prevalent in deceptive pleas (ten Brinke et al., 2012).

Verbal indicators of deceit . Due to the two meta-analyses conducted by DePaulo et al (2003) and Bond and DePaulo (2006) which show the using non-verbal cues to detect deception does not improve accuracy, there has been a shift in empirical enquiry to start focusing more on the verbal cues associated with deception (Vrij, 2018). As with any cue of deception it is thought the verbal cues 'leak' out through the language used by the deceiver. In deception literature, before this shift to looking at verbal cues as potential indicators of deception it was often thought that detection accuracy would be higher if one focused solely on the non-verbal cues of deception, because it is easier to control physical behaviour than our voices (Vrij, 2008). There

are four main cornerstones to the idea that non-verbal indicators are better at indicating deceit. Firstly, there is unlike with behaviour, there is no automatic link between an experienced emotion and the types of words used (Vrij, Taylor, & Picornell, 2015b). Secondly, people are generally better at using words than using behaviour to get what they want (Vrij et al., 2015b). Thirdly, people are more conscious about what they are saying than their non-verbal behaviour (Vrij et al., 2015b). Finally, people can stop and plan the types of words they are going to use before they speak, but this is not the case with their non-verbal behaviour (Vrij et al., 2015b).

But despite this increased attention to looking at the verbal indicators of deception, no one universal set of diagnostic verbal cues has arose (Vrij et al., 2015b). A potential explanation for this is because deceiver's or liars do not always tell a lie in the same manner nor fashion (Picornell, 2013). When looking at the deceiver's language, the following four cues are typically investigated word quantity, usage of pronouns, cognitive complexity and emotional word usage (Picornell, 2013). This thesis will elaborate on these later in thesis chapter.

It should be noted that efficacy of verbal cues associated with deception are strongly influenced by the properties of and the nature of the experiment being conducted. Factors which exert and influence on the findings of these studies include the nature of the stimuli i.e. is it a video tape, or a written statement, or a web-based decision task (Bond, & Lee, 2005; Zhou, Burgoon, Nunamaker, Jr, & Twitchell, 2004), the nature of the topic at hand i.e. is it a topic which is highly emotionally strung or is it a more mundane topic (Hancock, Curry, Goorha, & Woodworth, 2008), how incentivised the deceiver to get away with the lie (Sporer, 1997) and the very nature of the experimental subjects i.e. are the convicted criminals or offenders versus university students (Kassin, Meissner, & Norwick, 2005; Porter, & ten Brinke, 2008). It has also been found that the nature of the medium influences how effective these cues will be i.e. is the communication being done in a face-to-face situation or is it taking place in the form of a

computer medium, the latter allowing for greater preparation and planning (Burgoon, Blair, Qin, Nunamaker. Jr., 2003; Woodworth, Hancock, & Goorha, 2005).

Despite all these so-called non-verbal and verbal cues of deceit having been identified, the one finding which has been consistent in deception literature, is that people are poor at correctly determining veracity as was found by DePaulo and colleagues (2003) in the meta-analysis. Empirical studies have offered some explanations to account for this low accuracy rating when it comes to detecting deceit. For the most part, people who are trying to discern truth from lie make use of cues that are poorly correlated with deception (Anderson, DePaulo, Ansfield, Tickle, & Green, 1999). In 1996 Vrij, Semin and Bull offered an explanation for this low accuracy rating. According to the authors, there is a difference between what they refer to as objective cues and the perception of these cues.

According to Vrij and colleagues (1996) actual objective cues are founded on empirically proven non-verbal indicators of deception, whereas perceived indicators are those cues which reviewers associate with deceit regardless of whether these cues are associated with deceit. Additionally, these poor detection rates have been attributed to common misconceptions of what an actual cue to deceit is (Carlucci, Compo, & Zimmerman, 2013). Vrij (2008) suggests the reviewers often seek out the cues which they assume are indicative of honesty or deceit. A more contemporary reason for this poor detection rating is offered by Vrij, Granhag, and Porter (2010). The researchers postulated that this low rating could be attributed to what they called the ‘Othello Effect’. Simply put, this effect occurs when nervousness which is present during an interview setting is attributed to be an indication of honesty or deceit (Vrij et al., 2010). Finally, this poor detection rating can be explained by the fact that for the most part deception studies have low ecological validity as they are seldom occurring in high-stakes situation, and as such there is little reason for the deceiver to get away with the deceit (Frank, Freeley, Paolotonio, & Servoss, 2004; Carlucci, Compo, & Zimmerman, 2013).

Despite this poor performance at detecting deception, researchers and studies continued to try to find individuals who excel at detecting deception. In 2004, O'Sullivan and Ekman conducted a study that sought to discover outstanding lie detectors. Out of 12,000 tested only 29 were identified as exceptional lie detectors and were aptly named 'Wizards' (O'Sullivan & Ekman, 2004). Bond and DePaulo (2008) counter argued that if one considers statistical techniques and arguments, these so called 'wizards' may just have occurred due to chance.

The evolution of deception detection techniques

Within the deception literature it is apparent that through the years various techniques have been tried and researched in order to find the optimal technique to correctly detect the presence of deceit (Ben-Shakhar, Bar-Hillel & Kremnitzer, 2002). Techniques have gone from a reliance on arousal-based techniques such as the polygraph to using various interview styles to detect the presence of deceit. Other techniques which have garnered increased empirical focus include inducing a higher level of cognitive load on the deceiver (Blandón-Gitlin, Fenn, Masip & Yoo, 2014; Christ, van Essen, Watson, Brubaker, McDermot, 2009; DePaulo et al., 2003; Patterson, 2009; Vrij & Ganis, 2014; Vrij, Fisher & Blank, 2015a; Zuckerman, DePaulo, Rosenthal, 1981), to focusing entirely on the language used by the deceiver (Akehurst, Köhnken, & Höffer, 2001; Johnson, & Raye, 1989; Oberlader et al., 2016; Parker, & Brown, 2000; Steller, & Köhnken, 1989; Strömwall, Bengtsson, Leander & Granhag, 2004; Vrij, Akehurst, Soukara, & Bull, 2002).

Deception detection through arousal. In the forensic field, the polygraph is often used as a diagnostic procedure to establish credibility (Palmatier, & Rovner, 2015). Polygraph testing has come under scrutiny with a number of empirical studies and much debate as to the accuracy of the assessments flowing from the interpretation of polygraph results (Palmatier, & Rovnar, 2015). The polygraph measures the psychophysiological response of the body when the individual is being interviewed (Ben-Shakhar et al., 2002; Raskin, & Honts, 2002). The

polygraph is based on the well-researched and well-documented flight-or-fight response (Raskin, & Honts, 2002). The flight-or-fight response refers to the body's complex response to a physiological or psychological threat, in which several complex reactions happen in the body that is readily measured by the polygraph. These responses include increases in blood pressure, decreases in respiratory response and constriction of blood vessels (Raskin, & Honts, 2002). The polygraph typically makes use of two distinct types of tests; the Control-Question Technique (CQT) and the Guilty knowledge Test (GKT) (Ben-Shakhar et al. 2002; Ben-Shakhar, 2002). With both types of testing, guilt is established by an increase in physiological response of the suspect (Ben-Shakhar, 2002; Icanon & Lykken, 1997).

One of the problems with polygraph testing and the use of its findings in criminal procedures is the ongoing debate with regards to the validity and reliability of polygraph's results. Both of the two approaches to polygraph testing (CQT, GKT), can fall victim to countermeasures being used by the examinee (Ben-Shakhar, 2002; Ben-Shakhar & Dolev, 1996; Ben-Shakhar & Elaad, 2002). All countermeasures deployed intend to mislead the examiner in the analysis of the polygraph results. Countermeasures work by producing physiological responses which results in the examiner interpreting the messages as truthful (Vrij, 2000). Countermeasures can take the form of either a mental or physical countermeasure. Physical countermeasures are actions which are undertaken during the control question phase to influence the baseline physiological responses are used. Examples include, biting one's tongue or pressing one's toes against the floor (Vrij, 2000). Mental countermeasures that have been reported include counting back sequentially or counting sheep (Vrij, 2000).

Interviewing to detect the presence of deceit. Empirical research has yet to prove the efficacy of arousal-based protocols such as polygraph interrogation, to detect the presence of deception (Mann, Vrij, & Bull, 2002; Vrij, 2000; Walczyk, Igou, Dixon & Tcholakian, 2013). As such, a considerable amount of the deception literature has looked at various interviewing

techniques to detect the presence of deceit. Two of these techniques will be discussed in this literature review, the Strategic Use of Evidence and the Devil's Advocate Approach.

Strategic Use of Evidence. One of the most researched interviewing styles tested empirically is referred to as Strategic Use of Evidence (hereafter SUE). SUE is founded on the psychology of self-regulation, the tendency of people to control their behaviour to avoid undesirable situations (Forgas, Baumeister & Tice, 2009). And this self-regulation is mostly unconscious and happens automatically (Bargh, & Chartrand, 1999).

SUE holds regardless of the veracity of statement, the sender will attempt to come across as innocent and deploy counter-interrogation strategies, be these verbal or non-verbal behaviour (Granhag, & Hartwig, 2008; Hartwig, Granhag, & Luke, 2014). SUE assumes that that the interviewer possesses incriminating evidence that can be used to counter the testimony of the examinee (Wagenaar, van Koppen & Crombag, 1993). The SUE technique comprises of various techniques which are concerned with how best to use the available evidence (Luke et al., 2012). Past research has provided empirical evidence to the support the SUE technique of withholding evidence. In order to elicit statements which are not supported by corroborating evidence, the SUE interviewer will only ask questions related to the evidence (Luke et al., 2012). By not leading with questions based on evidence before the suspect has provided a statement, the chances of a guilty suspect avoiding contrary evidence is reduced (Hartwig, Granhag, Strömwall, & Vrij, 2005; Hartwig, Granhag, Strömwall, & Kronkvist 2006). Hartwig et al., (2006) determined that those trained in the SUE technique were 85.4% accurate, compared to those who were not, who scored and accuracy rating of 56.4%.

The Devil's Advocate Approach. Leal, Vrij, Mann and Fisher (2010) developed a novel interview approach for discerning truth from lie. They named this new approach the Devil's Advocate Approach. Within an interview setting which is utilising this technique two different

questions are asked. The interviewer asks the interviewee both an opinion eliciting question, in which the interviewee will reveal their opinion on said question. This is then immediately followed by the Devil's Advocate question. This question expects the interviewees to generate and provide an opinion which is in contrast to their truly held beliefs or opinion. Hence this paradigm is formed on the idea that truths for liars will come to the bare during the Devil's Advocate question, and the converse holds for truth-tellers (Deeb et al., 2017).

The devil's advocate approach is founded on both confirmation bias and the principle of impression bias. According to Ajzen (2001) one's attitudes towards things and information are maintained via confirmation bias. It is a well-documented phenomenon that views or information which aren't aligned with one's personal views are focused on above information which is in direct contradiction to one's views or beliefs (Edwards, & Smith, 1996; Mercier, & Sperber, 2011). By selectively focusing on certain information, individual's reactions reduce the cognitive dissonance, as they are eliminating the discomfort of having to confront contradictory evidence (Petty, Wegener, & Fabrigar, 1997). By doing this the individual maintains their viewpoint and enhances their access to points which would substantiate their argument (Deeb et al., 2017). Thus, when people are generating arguments for their truly held views (i.e. telling the truth) they would find it easier to find corroborating evidence. (Nickerson, 1998). The opposite is true for liars.

The Verifiability Approach. A recent interviewing technique aimed at detecting deceit that is gaining attention is the Verifiability Approach. This interview technique exploits the differences between a truth and a lie by requiring from the speaker to provide as many verifiable details as possible during the interview (Nahari, 2019). This approach defines verifiable detail as any detail that can be objectively verified (Nahari, Vrij, Fisher 2014a). This approach exploits what is described as the liar's dilemma. When a person lies, they inadvertently enter the liar's dilemma; having to provide enough detail to convince the receiving party of their story, whilst

not providing too much detail that could potentially be verified to be untrue. The more detail provided, the greater the chance that the investigator might come across a detail which cannot be verified and would betray the deceit (Nahari, 2018). The liar's dilemma arises as deceivers are trying to convince the interviewer of their honesty by providing detail but needing to balance this requirement with likelihood of the receiver being able to verify it as untrue. (Nahari, 2019).

The Verifiability Approach interview technique begins with the interviewer instructing the interviewee to provide as much verifiable information as possible when recounting their story or providing a statement. This triggers the liar's dilemma if the interviewee intends to lie (Nahari, 2019). This approach is showing promise, as studies have yielded accuracy rates of 70% in discerning truth from lies. (Harvey, Vrij, Nahari, & Ludwig, 2016; Jupe, Leal, Vrij, & Nahari, 2017; Nahari et al, 2014b). The verifiability approach also seems to be resistant to the countermeasures typically undertaken by liars to facilitate them getting away with deceit, as evidenced by studies in which the liars where informed of the manner in which they were going to be interviewed (Harvey et al., 2016; Jupe et al., 2017).

The (cognitive) cost of being deceitful . In addition to the increased focus on the verbal indicators of deception, another major turning point when it came to detecting deception was the move from arousal-based techniques to focusing on the various cognitive processes or strategies employed by liars, and how-to best counter them (Vrij, 2018). According to Vrij (2018) this shift resulted from a report published in 2003 by National Research Council in the US. In this report the NRC concluded that no theory existed that could attribute an adequate reason as to the reason why deceivers would show a greater level of arousal (i.e. why they would be more nervous) than truth-tellers (Vrij, 2018). Secondly this report concluded that at the time it was not possible to ask a specific that would increase the level of nervousness of the liar (Vrij, 2008). Techniques built around cognition techniques are considered with the latter finding of NRC report. New

interview protocols were being develop in order to facilitate these differing cognitive processes between truth-tellers and deceivers, and in turn, enhance the verbal indicators of deception.

Although numerous physiological responses have been observed during a state of increased arousal, most researchers assert that this is more likely emanating from the fear of being caught out as being deceptive and that one must therefore consider the context in which the interrogation is done. Is the person fearful of being falsely found guilty or is the subject feeling guilty for no obvious reason or nervous? Each of these states will result in a physiological reaction (Patterson, 2009). Where the polygraph relies on physiological measures to detect deception, cognitive load approaches argue that being deceptive requires greater mental effort making it hard to get away with the deception (Walczyk et al., 2013).

The idea to use additional cognitive load to aid in detecting deception can be traced back to the observations made of Zuckerman, DePaulo and Rosenthal (1981), who argued that lying was more taxing on cognition than telling the truth. Because of this observation, many empirical studies have been and are still being conducted into the efficacy of cognitive-load approaches to detecting deception (Blandón-Gitlin et al., 2014). It is argued that the amount of cognitive load experienced by an individual whilst providing statements can be increased through the use of certain techniques or through specific interventions (Vrij et al., 2015a). The idea underpinning the use of cognitive load during the statement-gathering phase is that lying is more cognitively taxing, and that by increasing the cognitive load on the liar whilst providing a statement or during an interview makes it more challenging for the liar to mask the attempts at deception. The premise is that by increasing the cognitive load experienced by the deceiver, the total amount of cognitive load would be too much to control adequately. This in turn, would result in the interrogator having a better chance to detect deception. The interrogatee would betray that deception was present, by displaying more behavioural cues associated with deception because

of the reduced capacity to control behavioural cues following from the increased task complexity (Vrij et al, 2015a. Vrij, Fisher, Mann, & Leal, 2006).

Empirical research, with evidence stemming from real-life police settings, has found that the increase in cognitive load may result in physical manifestations. These manifestations include increased amount of blinking, increased fidgeting among others (Vrij et al., 2006). Because of these findings, the authors argue that if an interview style which utilises cognitive load were to be employed it would make it easier to detect deceit, as cognitive load will introduce a secondary task which would be more taxing cognitively, that sustaining the lie or deceit would be made more difficult. Several reasons have been offered as to why the increase in cognitive load facilitates the detection of deception. Without prior preparation of the lie or the fabricated story, deriving a feasible and/or a believable story requires a greater amount of cognitive effort (Vrij, Mann, Fisher, Leal, Milner, & Bull, 2008). Liars need to continually assess if their lies are being believed and if they are plausible claims or stories, and as such they must exert a greater amount of effort to ensure that no slip-of-the-tongue phenomena occur, which would betray the fact that they are lying (Vrij et al., 2008). Thirdly, liars do not enter a conversation under the assumption that they will appear credible and honest as truth-tellers do, as a result they are continuously monitoring their own demeanour and that of the receiver, to ensure that they are getting away with the deception, and this is draining on the amount of cognitive resources that they have (Vrij et al., 2008).

According to Blandón-Gitlin et al., (2014) at a neuro-cognitive level, on the sender's part, lying activates event-related memory, and this information is detrimental to the deceiver. The activation of the event-related memory causes problems for the deceiver as it is now required to manage the unwanted information in such a manner as to control any leakage which would undermine the credibility of the speaker (Blandón-Gitlin et al., 2014). Conversely, in the case of the truth-teller, the spreading of the activation of memory networks allows useful information to

become available (Blandón-Gitlin et al., 2014). Neuro-imaging studies further show there is significantly greater activation of certain brain regions when deception occurs. More specifically, the regions involved with working memory appear more active when a person is lying as opposed to being truthful (Christ et al., 2009). Blandón-Gitlin et al., (2014) established that on the behavioural-cognitive level, cognitive load approaches influenced the use of both cognitive and behavioural strategies to appear credible (Blandón-Gitlin et al., 2014). This idea is supported empirically by recalling events in reverse order. This is because an imagined event is challenging to manipulate during backward recall (Blandón-Gitlin et al., 2014).

Several studies have been conducted to test the idea that being deceptive increases cognitive load and attempted to determine the underlying mechanisms that contribute to this increased cognitive load. DePaulo et al., (2003) suggest that those who are being deceptive do not take their credibility for granted; therefore, they engage in impression management. Patterson (2009) argued that lying imposes an additional demand on one's finite cognitive resources available to the deceiver. Another mechanism identified is that liars are constantly monitoring their audience to ensure that they are 'getting away' with their lie, and therefore there is less available cognitive resources to suppress the truth (DePaulo et al., 2003). Furthermore, Patterson (2009) argued that this increase in cognitive load arose due to the nature of being deceptive. Patterson (2009) believed that as lying by nature is an intentional act, rather than a spontaneous one, it demands more cognitive resources.

Several studies have suggested various techniques which can be deployed to impose additional cognitive load. Techniques include asking interviewees to narrate a story in reverse chronological order, instructing participants to always maintain eye contact and finally have participants perform several tasks concurrently (Evans, Meissner, Michael, & Brandon, 2013; Vrij et al., 2008; Vrij, Leal, Mann, & Fisher, 2012; Zimmerman, Veinott, Meissner, Fallon, & Mueller, 2010).

In various studies participants reported that they experienced an increase in cognitive load when telling a lie. Participants felt that it was cognitively more taxing to lie than telling the truth (Hartwig et al, 2006; Vrij, & Mann, 2006). In 2007 Vrij, Mann, Kirsten and Fisher sought to determine if an increase in task complexity would result in an increase in the cognitive load reported by participants. They found that the greater the task complexity, the greater the amount of cognitive load reported by the participants.

In 2005, Walczyk et al., proposed and created an innovative approach to induce cognitive load called Time Restricted Integrity-Confirmation (Tri-Con). Tri-Con is founded on the differences in mental states between those who tell the truth and those who deceive (Walczyk et al., 2005). Tri-Con enhances cognitive load by asking unanticipated questions and requires quick response. Tri-Con works according to specific guidelines. Firstly, examinees are primed about area of enquiry of the upcoming questions. Secondly, despite this priming the specific questions are not disclosed until the examination. Thirdly, questions are deliberately kept vague and ambiguous to make it difficult to prepare lies. Fourthly, examinees are expected to respond immediately (Walczyk et al., 2005). The high cognitive load induced by the rapid questioning and responding increase the likelihood of cues leaking such as changes in pitch, pace and an increase in pupil dilation (Buller, & Burgoon, 1996; Walczyk et al., 2005; Walczyk et al., 2013). Tri-Con has been tested empirically and the results are reported to be promising.

Tri-Con's efficacy has been tested and established in empirical work. In 2003, Walczyk, Roper, Seemann and Humphrey asked adults to tell a lie and a truth their lives. By making use of responses time as the indicator of the presence of deceit, Tri-Con was successfully able to distinguish between truth and lie. In 2009, Walczyk, Mahoney, Doverspike and Griffith-Ross tested Tri-Con again, in which a rehearsal condition was added. Thus, participants allowed to prepare lies and truths. Liars and truth-tellers were classified 89% accurately. In 2012 Walczyk and colleagues tested Tri-Con again, this time in a forensic context. After watching a mock

crime, witnesses were required to report what they witnessed truthfully or witnesses were instructed construct a false witness report (Walczyk et al., 2012). Cognitive cues were used, these included response time, answer consistency, eye movements etc. Tri-Con correctly classified the lies and truths 69% of the time (Walcayk et al., 2012). Although, Tri-Con shows promise, it is not without its criticisms and limitations. One such criticism is offered by Sporer and Schwant (2007) who argue that the extended narratives provided by suspects allow for valuable verbal indicators of deceit to be used, and Walczyk et al., (2013) acknowledge that Tri-Con does not afford its examinees a chance to qualify their answers.

Uncovering the truth through language. Ekman (1992) claimed that liars tend to focus and be careful about the words they use since words are one of the most differentiated ways in which we communicate and are often scrutinised. A recent study found that language had some diagnostic value when used to differentiate between truth-tellers and liars (Jupe, Vrij, Leal, & Nahari, 2018). In a typical paradigm for studying language and deception, particularly in terms of communication, two groups of communicators are randomly assigned to two distinct groups where one group is instructed to be truthful and the other is instructed to be deceptive (Sporer & Schwandt, 2006). When it comes to detecting deception using language constructs several techniques and frameworks have been proposed. The most commonly used frameworks are that of Statement Validity Analysis (SVA) and Reality Monitoring (RM) (Vrij, Akehurst, Soukara, & Bull, 2002).

Statement Validity Assessment. Criteria-Based Content Analysis (hereafter CBCA) forms the basis of this language technique to determine veracity. SVA was originally conceptualised and designed to assess and determine the veracity of childhood sexual assault case statements. The Undeutsch hypothesis forms the basis of CBCA, which posits that statements based on actual experience are richer in detail when compared to those statements that are made on non-experienced based statements. This is because they can be readily be called

upon, rather than made up. CBCA comprises of a set of five categories (19 criteria) which are theorised to be more frequent in truthful as compared to deceptive statements (Vrij et al., 2015b). These five categories are general characteristics, specific characteristics, motivational components, content particularities, and offence-related components. The statement is read and then assessed against these criteria (Oberlader et al., 2016).

According to Köhnken (2004) certain criteria when they are present are indicative of honesty as they are theorised to be too complex to fabricated, these criteria are; the statement has a coherent and logical structure, the statement consists of a high amount of detail, and finally the information in the statement is not structured in a chronological order. Other criteria according to CBCA theory which indicate honesty include, the presence of spontaneous corrections and the fact the provider will admit that they are not entirely sure of the memory via phrases such as "I think".

When testing the efficacy of SVA via the CBCA criteria Vrij (2008) found evidence that it could correctly differentiate between truth and lies. It was found that truthful statements typically scored higher in the amount of CBCA criteria. On average, CBCA could correctly differentiate between truth and lie 71% of the time Vrij (2008). Several studies looked at CBCA (Akehurst et al., 2001; Parker, & Brown, 2000). Generally, studies support the basic assumption of CBCA; that truthful statements are richer in detail and are based on actual experienced events (Akehurst et al., 2001; Parker, & Brown, 2000). Additionally, empirical evidence for CBCA comes from Vrij (2005) vote-counting study and the meta-analysis conducted by DePaulo and colleagues (2003). Truth-tellers generally included more details in their statements and had more corrections that are spontaneous and were more logically structured (DePaulo et al., 2003). The use of CBCA in forensic settings is not without its criticisms as stipulated in Hauch, Sporer, Masip and Blandon-Gitlin (2017) study. The researchers found that should CBCA be used in

forensics settings, there needs to be rigorous guidelines as to how the CBCA experts need to evaluate the presence of each criteria (Hauch et al., 2017).

Reality Monitoring. Reality monitoring is based on work which focuses on memory. Reality monitoring is more prevalent in deception research conducted by researchers rather than professionals working in the forensic fields. The central premise of reality monitoring is that differences exist between memories which are real and those which are fictious in nature (Johnson, & Raye, 1998). Although not originally designed to separate truth from lie, the reality monitoring approach is showing great promise (Oberlader et al., 2016; Vrij et al., 2015b). According to reality monitoring thought; memories that are based on actual experiences will contain perception details (i.e. information derived from the five senses) as well as contextual details (i.e. details concerned with where the incident occurred, or where an object or person was located) (Vrij et al., 2015b). Reality Monitoring also holds that memories based on real experiences tend to be richer in details pertaining to time (i.e. the story being told will have a sense of chronological order to it) (Vrij et al., 2015b). A false memory typically is typically less clear and tends to be vaguer in details (Vrij et al., 2015b). When applied to deception detection, reality monitoring believes that truths are based on external information where lies are based on internal information (Sporer, 2004).

Studies investigating the efficacy of reality monitoring generally find that its ability to distinguish between truth and lie was similar to that of CBCA (Masip, Sporer, Garrido, & Herrero, 2005; Vrij, 2008). On average, reality monitoring scored a 69% accuracy rating when differentiating between truth and lie, specifically, it was found that truthful accounts did have more perceptual and contextual details (Vrij et al., 2015b). Reality monitoring has shown greater amounts of inter-rater reliability when compared to that of CBCA (Strömwall et al., 2004). This has been attributed to its simplicity of application. Sporer (1997) also found that reality monitoring's criteria to be more precise and therefore easier to operationalise when compared to

CBCA's criteria. Contemporary research has found that certain RM criteria holds more efficacy in discerning truth from lies and can even be applied to discerning between intended future acts of deceit and honesty (Giolla, Ask, Granhag, & Karisson, 2019). The specific criteria are; *cognitive operations, clarity, reconstructability* and *realism* (Giolla et al., 2019), and argues that these criteria should be used using RM to distinguish between truth and lies, especially when using it to discern intended acts of deceit.

Psycholinguistics and deception. Linguistic cues are those features of language such as the types of words used, and the manner in which those words are stringed together, in other words, the structure and grammar of the messages. Furthermore, Burgoon (2018) argued that the linguistic features of language exist regardless and independent of the context or medium of the communication. There exists empirical evidence that systematic differences exist between truthful and deceptive witness accounts (Fuller, Biros, Burgoon & Nunamaker, 2013; Joffe, 1992; Zapamiuk, Yuille, Taylor, 1995). In addition, to these systematic differences, studies have shown that linguistic differences do exist between truthful and deceptive statements whether under controlled conditions (i.e. a laboratory environment) or under actual field conditions. (Burgoon & Qin, 2006; Hancock, Woodworth, & Goorha, 2010).

Newman, Pennebaker, Berry and Richards (2003) used psycholinguistics to identified several language or linguistic features that could be used to determine if deceit was present in a written statement or account. Those features identified by Newman and colleagues (2003) included; deceivers typically showing lower levels of cognitive complexity, used more negative emotion words, and had fewer references to both the self and others in their statements compared to truthteller's statements.

Walker, Vogl and Thompson (1999) found truth-tellers resorted to using more positive emotion words when discussing life events. Walker et al., (1999) explained this finding by

saying that these events were positively biased. Walker, Skowronski and Thompson (2003) explained this phenomenon by using the fading-effect bias. According to the fading-effect bias, negative memories fade faster than positive memories, thereby reducing the amount of negative emotionally charged words in a truthful account (Walker et al, 2003). Negative emotionally charged words are associated with shame and guilt (Eckman, 1992; Vrij, 2008), hence these types of words are more likely to be used by liars. In 2015 Vrij, Fisher and Blank concluded that because deceitful communication increases the cognitive load experienced, the words used by the deceiver are simpler. This finding was later replicated by Jupe, Vrij, Leal and Nahari (2019) that differences existed between truth-tellers and deceivers when cognitive load was a factor at play. Cognitive process words e.g. 'ought' etc and cause words such as 'hence', 'because' etc are often used by liars. (Sporer, & Schwandt, 2007).

A major problem with reviewing psycholinguistics and deception studies is that studies are reporting dissimilar results. It was found that deceivers used fewer words, but longer sentences (Haunch, Blandón-Gitlin, Masip & Sporer, 2015). Additionally, Haunch et al., (2015) found that contrary to what one would expect deceivers did not use fewer past tense verbs.

As such one solution offered by Burgoon (2018) to reconcile these different results is classify linguistic markers into groups and examine if any clear patterns emerge. In the case of non-verbal indicators there is and has been this misconceived notion that there exist universal cues to detecting deception, and these universal cues are context dependent and stable (Frank & Eckman, 1997). However, when it comes to the linguistic indicators of deceit, research argues that it is dependent on various factors e.g. the type of deception (for example is it a deliberate concealment of information or is it a fabricated story), the event type and the emotional valence (Buller, Burgoon, Buslig & Rogers, 1996). Furthermore, both Zhou and colleagues (2004) and Burgoon and Qin (2006) found that the linguistic markers or deceit are dynamic in nature. A brief overview of some linguistic indicators follows.

Quantity refers to the length of a statement. It is often measured using a count of syllables, types of words e.g. nouns, verbs etc. or loquacity (talk time) (Burgoon, 2018). The early idea was that truth-tellers would be forthcoming, whereas, deceptive communicators would not be so forthcoming when communicating with the intention to reduce the likelihood that they would be discovered (ten-Brinke & Porter, 2012). However, two meta-analysis conducted by DePaulo and colleagues (2003) and Hartwig and Bond (2011) respectively found mixed results when it came to looking at quantity as an indicator of deceit. One explanation offered is that there are many instances where deceivers intentionally decide to give longer statements (Burgoon, 2018) For example, interactive experiments have yielded results which indicate when afforded the chance to plan and prepare their lies, deceivers tend to provide longer statements (Burgoon et.al., 2015; Dunbar et al., 2014; Zhou et al., 2004). This increase in loquacity may occur due to the deceiver trying to be more persuasive.

Specificity is concerned with the level of detail provided by a statement (Burgoon, 2018). In essence specificity just serves as concrete operationalisation of quantity. Legal testimony is often coded using SVA, CBCA and Reality Monitoring (Johson & Raye, 1981; Undeutsch, 1989; Vrij, 2008). When determining veracity of statements, these techniques look at sensory details, spatiotemporal references etc. In truthful statements there is a greater amount of detail (Burgoon, 2018). This assertion is based on the premise that during a retelling of a truthful event, the person will able to recall a greater level of detail. Two theories support this claim namely, Information Manipulation Theory and Information Management Theory as these theories argue that to detect deception one needs to manipulate the quantity of details (McCornack, 1992; Burgoon, Buller, Guerrero, Afifi, & Feldman, 1996).

Lexical complexity refers to either lexical or syntactic complexity when it comes to using language to detect deception. (Burgoon, Buller, Ebesu, & Rockwell, 1994). Lexical complexity

is concerned with whether, simple of difficult words are used, measured by the number of letter or syllables (Burgoon et al., 1994). Syntactic complexity is concerned with whether simple sentences were used of if sentences with multiple clauses were used, measured by punctuations etc (Burgoon et al, 1994). Because of the cognitive complexity of lying there are less cognitive resources available and as such the speaker may opt to use less complex language (Qin, Burgoon & Nunamaker, 2004). However, when investigating lexical complexity Hauch et al., (2015) did not find a simpler lexical structure when looking at a deceptive statement.

Diversity speaks to each lexical item in a given statement and how unique or redundant they are. Diversity is often measured using the Type-Token Ratio (TTR). To obtain the TTR one simply "divides the number of distinct word types by the total number of words in the statement" (Porter, & Yuille, 1996 p. 446). According to Hollien (1990), in order to maintain credibility, people who are providing false testimony display low levels of lexical diversity. s their language behaviour becomes more stereotypical. Additionally, non-diverse language is indicated by repeated words and phrases. Repetitive language is considered a sign of a deceptive statement as it is the result of a reduced state of mental resources available to the speaker and therefore, the speaker reverts to using less advanced vocabulary and does not construct complex statement (Burgoon, 2018), Hauch et al., (2015) found that deceivers used sentences with fewer diverse wording and in general had a lower TTR.

Hedging or uncertainty speaks to the phrases or words and the sentence structure of the statements which introduce ambiguity, evasiveness and vagueness (DePaulo, 2018). According to Duran, Hall, McCarthy & McNamara (2010) deceivers may make use of ambiguity and vagueness to avoid detection by indicating lack of conviction when it comes to their stories being questioned. Thus, certainty in language conveys confidence in which is typical of truthful-tellers (Fuller, Biros & Wilson, 2009).

Immediacy and Personalism refers to how the language used creates a sense of closeness, or rather how that person takes ownership of what they say or think (Zhou et al, 2004). If the statement generated contains a mixture of the passive voice and using both the past and the future tenses then generally the person has provided a non-immediate sentence or statement (Wiener, & Mehrabian, 1968). Additionally, using the third person pronoun decreases the sense of immediacy. Deceivers, in order to disassociate themselves from their falsified narrative (Newman et al., 2003), often use non-immediate and impersonalised pronouns. Hauch, Blandón-Gitlin, Masip, and Sporer, (2015) found that the difference in language profiles between truthful and deceptive statements lies within the types of pronouns used. In the case of truth, there is a tendency to utilise first-person singular pronouns. In the case of deceit, more second and third person pronouns tend to be found.

Most displays of deception can be said to have some affective cause. It has been said that deceivers are seen to have been experiencing feelings of guilt, anxiety and the fear of being caught in the act of being deceitful. These fears then prompt uncontrollable and unconscious leakage of cues associated with deception (Buller & Burgoon, 1996; Ekman, 2009). What isn't clear, however, is if these emotional states which are leaked through, generate negative valanced language. Both Bond and Lee (2005) and Vrij, Edwards, Roberts and Bull (2000) suggest that guilt is reflected in high amounts of negative emotion words. However, Burgoon et al., (2015) found contradictory evidence to the previous findings. Burgoon and colleagues found that deceivers deliberately used more positive and hyperbolic language when they were trying to be persuasive. In the same year, Hauch et al., (2015) found the converse to Burgoon et al., (2015) as they found more negative emotion words trend towards more positive emotion words being used by liars. This finding led Hauch and colleagues (2015) to make the tentative assertion that compared to truth-tellers, deceivers used more emotionally charged words.

Burgoon (2018) asserts that both motivation and modality act as moderators to linguistic indicators of deception. In terms of motivation, the extant literature, centred on deception has argued that motivations effects both the sender in terms of their display of cues associated with deceit, and the receiver's ability to detect deception (Burgoon, & Floyd, 2000). However, a major criticism of this literature is that they focused on low-stake, mundane lies (Bachenko, Fitzpatrick, & Schonwetter, 2008; Buller, & Burgoon, 2008; Frank & Ekman, 1997, Vrij, 2008). As a result of this criticism, researchers have begun investigating whether motivation facilitates or inhibits detection accuracy in high-stake lies.

DePaulo, Kirkendol, Tang, & O'Brien (1988) furthered the research during the early years of deception by investigating the motivation impairment affect hypothesis. According to this hypothesis motivated liars become easier to detect or rather their intention becomes more transparent if judges can only make use of non-verbal cues. This implies motivation impairs non-verbal performance but improves verbal performance when someone is trying to be deceptive (DePaulo et al., 1988). In 2000 Burgoon and Floyd took this one step further, the researchers took several operationalised definitions of motivation and applied it to the deception detection. They concluded that if there was a high degree of motivation to detect deception then detection accuracy rates would be improved regardless if non-verbal cues or if verbal cues were used (Burgoon & Floyd, 2000). It was also found that highly motivated liars who made use mediums other than face-to-face communication such as instant messaging platforms could often manage to deceive their partners (Hancock, Woodworth, & Goorha, 2010). Thus, indicating that high levels of motivation may hinder the ability to detect the presence of a lie or deceit.

Modality and its influence on communication is a well-documented phenomenon. It has been found that the mode of communication influences how people speak. When speaking in person; it is often more informal, however, when using email, it is more formal (Burgoon, 2018). The study concluded that the modality in and of itself will result in differences in language used.

For example, in a face-to-face communication, a deceiver needs to create a credible story and therefore is more likely to be constantly aware of their impression and the feedback being received from the receivers (Burgoon, 2018). However, when using written modalities, communicators are able to edit and plan their messages before sending and can therefore focus on the linguistic aspect of the message (Burgoon, 2018).

A review of the literature has yielded the following characteristics that are often present in a deceptive statement. Liars tend to provide fewer details when speaking, and a result of this, their statements will often contain less word than truthful statements (Burgoon et al., 2003; Vrij, 2000). This is because of how cognitively taxing telling a lie is, and how hesitant they are in providing too much detail which they may be questioned on, which may betray they are lying if they are inconsistent in their stories. Secondly, liars typically use less pronouns overall and specifically less first-person pronouns to dissociate and create distance between themselves and the deceit (DePaulo et al., 2003; Newman et al., 2003). DePaulo and colleagues (2003) also concluded that liars also used a greater number of third-person pronouns to further facilitate this dissociation. Fourthly, liars typically used more positive and less negative words than truth-tellers in order to make their deception more believable (Burgoon et al., 2003)

One of the major developments that came from the increased focus of research in terms of psycholinguistics and deception, was taking machine learning and Natural Language Processes (NLPs) and using them to detect deception in communication.

Using computers to sort the lies from the truths. With the advent and rise of machine learning techniques, it is now possible to use computers to detect the presence of deceit, at least in theory (Nortje & Tredoux, 2019). These machine learning techniques are often used in natural language processing techniques. Examples of commonly used machine learning techniques include Bag-of-Words (BoW), n-grams and other styles which look at linguistic style approaches

(Hernández-Castañeda & Calvo, 2017). It should be noted, that although n-grams yield acceptable results, they generally perform better when they are augmented with other NLP techniques (Hernández-Castañeda & Calvo, 2017). NLPs make use of computational linguistics to detect deception by the text being processed and turned into vectors of various features (Hernández-Castañeda & Calvo, 2017). A wide variety of techniques have been developed in order to try and detect deception from using physiological sensors such as the polygraph as explained above to using neuroscience techniques (Davatzikos et al., 2005; Ganis, Kosslyn, Stose, Thompson & Yurgelun-Todd, 2003). However, more recently many techniques have been developed to detect deception at the communication level. However, finding universal cues is proving to be a challenge. One solution offered is to make use of statistical and machine learning techniques to detect these cues to deception.

Linguistic Inquiry Word Count (LIWC), a software programme, is widely used to construct language profiles and evaluate these to differentiate between truthful statements and deceptive ones. Additionally, when conducting research in this paradigm researchers use the categories as constructed and defined by LWIC (Pennebaker, Francois, & Booth, 2001).

LIWC analyses texts against predefined dictionaries to classify words into psychologically meaningful categories and reports the number of words in each category (Tausczik & Pennebaker, 2010). For an accurate classification, LIWC relies on meticulous and carefully constructed dictionaries (Nortje & Tredoux, 2019). LIWC has been used to study several areas including personality, psychological adjustment and mental health (Alpers et al., 2005; Mairesse, Walker, Mehl, & Moore, 2007; Rude, Gortner, & Pennebaker, 2004)

LIWC has successfully been used to correctly differentiate between truthful and deceptive texts. These combinations of verbal cues have resulted in a successful detection rate of between 67%-70%. Newman et al. (2003) collected a corpus of both truthful and falsified

statements across five experiments. In terms of the first three experiments that were conducted, participants were expected to express both their true opinions and a contradictory opinion on a topic. The first study made use of orally recorded transcriptions, study two and three made use of either written communication that was typed or hand-written respectfully. These were then put through LIWC and the authors achieved 61% correct classification of truthful compared to deceptive accounts score overall. The subset of LIWC which differentiated between truthful and deceptive communication were motion words, both self-references and references to another person, the total number of exclusive words and finally negative emotion words.

Newman et al. (2003) argued that this subset could be used to differentiate between a truthful and deceptive account because a high number of self-references show that the person is taking ownership of the expected quality of the communication process, whereas liars tended to make use of fewer self-references in order to dissociate themselves from the interaction. Newman and colleagues (2003) also found a lower number of references to another person in lies when compared to that of a truth. However, Vrij (2000) found that there is a greater number of other references when telling a lie. DePaulo et al. (2003), meta-analysis did not find any differences in references to others between lies and truths. As such, there is no consensus on this, Newman et al (2003) responded this and said that they may have had a higher number of references to other people due to their manipulation. They asked participants to give opinions on abortion attitudes, and this would result in a greater number of references to another person. Newman and colleagues (2003) believed that exclusive words (e.g. “but”, “rather”, “however”) were an indication of cognitive complexity as they were used to distinguish between alternating concepts. Research has indicated that when it comes the use of exclusive words, liars tended to use less (Newman et al., 2010). Newman et al (2003) also found that lies contained a greater number of negative emotion words. According to Vrij (2000) this increase in negative emotion words could be attributed to feelings of guilt and anxiety associated with the act of lying.

In 2005 Bond & Lee applied LIWC to a random sample of truthful and deceptive transcripts of oral statements provided by inmates. Overall, it was found that deceivers scored lower on the LWIC category sensory details than truth-tellers. Mihalcea and Strapparava (2009) used LIWC categories on a sample of one-hundred truthful and false opinions when it came to conversational topics. They obtained a 70% success rate of detecting deception by making use of NLPs.

Similarly, Pérez-Rosas and Mihalcea (2014) investigated if it was possible to use machine learning techniques to detect deception cross culturally, when one considered each culture individually and then built in specific cultural deception indicators. The study yielded a correct classification between 60-70%. Hence, it was concluded that one could use cross-cultural deception indicators to detect deception (Pérez-Rosas & Mihalcea, 2014).

Williams, Talwar, Lindsay, Bala, and Lee (2014) compared the lies told by children and those told by adults when it came to the language used. There were two groups and half of each group told a lie and the remaining half told a truth. The researchers used LIWC to generate the classification samples. The study showed that differences existed in the language profile of a deceptive versus a truthful statement. These differences mainly took the form of certain linguistic variables e.g., the use of both singular and plural self-references as well as, differences in both positive and negative valanced language (Williams et al., 2014).

There are advantages to using programmes such as LIWC to try and detect the presence of deception in a written account. One such advantage is that LIWC is able to determine emotionally charged words (Tauszcik, & Pennebaker, 2010). Another advantage of LIWC, and NLPs by extension is its inability to assimilate biases, as these programmes are not likely to be susceptible to stereotypes (Hauch et al. 2014). Furthermore, human evaluators or judges may become mentally taxed and exhausted after a prolonged interaction of trying to decide if a piece

is deceptive or not. For example, LIWC outperformed human judges when classifying opinions on abortion as truthful vs non-truthful. LIWC correctly classified these statements 67% of the time, whereas, human judges on 52% of time (Newman et al. 2003). According to Levine (2014) human judges tend to be truth-biased, in other words, they enter the conversation with the assumption that people are going to speak the truth.

It should also be noted that with the advances of our understanding of how machine learning works, machine learning has allowed for simple NLP models to accurately and correctly detect fake hotel reviews regardless if they these reviews were positive or negative (Ott, Cardie & Hancock, 2013). Using machine learning techniques Ott et al., (2013) scored an 86% accuracy rate when detecting fake negative reviews. Where in contrast, human judges scored an accuracy rating of 61.5% when detecting false negative reviews.

However, although this novel approach, shows promise, it is not without its criticism. It's major fall back is the fact that its algorithms are 'black boxes.' Simply put, this means that it is very difficult to discern how exactly these algorithms detect the presence of deception in written statements. Additional disadvantages of these types of programmes are that they ignore figurative language and contexts e.g. sarcasm, irony and idiomatic expressions.

Deception and second language

Recent developments in the field of detecting deception assess the impact that a speaker's language proficiency has on the speaker's credibility (Evans, Pimetal, Pena, & Michael, 2017). Several studies in this field, such as those conducted by Cheng and Broadhurst (2005), Da Silva and Leach (2013) and Lev-Ari and Keysar, (2010) indicate that veracity assessments of native language statements were more likely to be judged as truthful compared to statements provided in non-native languages.

However, these studies looking at the impact of second language has on veracity assessment has yielded inconsistent results. These dissimilar results have been explained by stating that these studies vary in both design and methodology, making comparisons between them difficult. In terms of their analysis certain studies have reported only on signal detection theory analysis and its relation to detecting deception only and others have reported on accuracy only.

Two main research directions emerge in attempts to clarify the differences between truthful versus deceptive statements made in one's native versus one's non-native tongue: (a) research that focuses on the cognitive load approach, and (b) research that stems from the domain of language psychology, which argues that there is greater emotional distance from the content of the statement when lying in a non-native language (Suchotzki, & Gamer, 2018).

Cladwell-Harris (2014) study is an example of the second research avenue indicated above; the researcher investigated the emotional content differences between native and foreign languages. According to the researcher, one's native tongue is acquired within an emotional context, where second languages are typically acquired or learned in a more rational academic context. The researcher proposed that words learned or acquired within a specific context, triggers an emotion related to the original context when used later. Researchers therefore argue that judging native and foreign languages relies on distinct processes. Evaluating second language communication is described as a more rational decision-making event whereas native tongue language communication evaluation shifts towards emotion-influenced events. This distinction between the different mental processes allows for emotional distancing to occur in evaluating second language communication (Cladwell-Harris, 2014) and this in turn could influence the veracity judgement of a message or statement.

Some studies argue that speaking in one's second language in and of itself introduces cognitive load onto the speaker (Da Silva, 2011; Evans et al., 2013, Evans, & Michael, 2014). Since the deceiver must suppress the truth, construct a credible story and constantly monitor their credibility increases the amount of cognitive load placed on the non-native speaker (Evans, & Michael, 2014). Furthermore, this increase occurs because the neural processing of the brain comes under increased demand when one speaks in a non-native tongue. This increased neural processing is attributed to more motor neurons being fired in the brain as these neurons are involved with the grammar of speech, thus resulting in it becoming more difficult for second language speakers to engage in conversation (Perani, & Abutalebi, 2005).

Research has indicated that relative to speaking in one's native tongue, speaking in a non-native tongue is more cognitively taxing (Akca, & Elkilic, 2011; Ardila, 2003). Ullman (2001) explains this phenomenon by saying this increase in neural activity is linked to explicit memory processes when conversing in one's non-native tongue. According to Ullman (2001) when conversing in one's second language it is harder to actively engage in word recall. Therefore, indicating that second language speakers have fewer words to their disposal when constructing a deceptive statement.

It has been found that differences exist between native and non-native speakers. Studies have argued that speaking in one's second language reduces one's working memory span (Service, Simola, Metsanheimo, & Maury, 2002). And as a result of this may be slower at recognising words or slower to make lexical decisions (Gollan, & Acenas, 2004). Thus, contributing to the increase in cognitive load experienced by non-native speakers. Research has also found that differences in arousal level exist when looking at a person when they are speaking their first language compared to when they are speaking their second language. Both Dewaele (2008) and Keysar, Hayakawa and An (2012) found that those who are speaking in their second language experience less emotional intensity. Behaviourally speaking, it was found

that when lying in their non-native language, participants exhibited increased pupil dilation and had longer utterances (Duñabeitia, & Costa, 2015)

Although much of the research done to date has primarily focused on the idea that lying is more cognitively taxing than being honest, it has been suggested by both lay and scientific enquiry that lying is also more emotionally arousing (Zuckerman, DePaulo, & Rosenthal, 1981). When asked people how they felt when lying in their non-native language, people reported that they felt tenser and experienced greater levels of stress when lying when compared to being honest (Caso, Gnici, Vrij, & Mann, 2005). Abe, Suzuki, Mori, Itoh, and Fujii (2007) found further evidence to suggest that lying is emotionally more arousing using functional MRIs. Abe and colleagues (2007) found that there was an increase in activity in the amygdala, the brain region which is involved emotional processing when participants were being deceptive.

Research from the disciplines of linguistics, psychology and psychophysiology indicated that when participants spoke in the second language, they experienced lower levels of emotional arousal when compared to speaking in their home language (Dewaele, 2004). Furthermore, when listening to emotionally charged phrases such as reprimands when younger, results in less skin conductance if these reprimands are delivered in a second language when compared to when they are said in the native tongue (Caldwell-Harris & Ayçiçeği-Dinn, 2009). Hence, Suchotzki and Gamer, (2018) conclude that being deceptive in one second language results in lower levels of emotional arousal, which in turn facilitates the act of lying or deceit.

Research further suggests that second language speakers have a harder time getting away with deception due to this accompanying increase in cognitive load, as one would need to split the cognitive resources between working memory and that of creating a story. This makes attention-switching tasks more challenging. Hence, second language speakers are argued to leak a greater number of cues which indicate deception (Da Silva, 2011). This statement is

problematic though. According to Gregersen (2005), those who speaking in their non-native tongue likely engage in activities associated with being deceptive simply because they are nervous, and because of this they may appear to be being deceptive.

In 2005 Cheng and Broadhurst investigated the ability of receivers of a message to detect deception if they heard both statements made in the speaker's native language and their (speaker's) non-native language. Statements were given by Cantonese speaking individuals, both in their native tongue (Cantonese) and in English but those who identified as second language English speaking individuals. The researchers found that second language speakers displayed behaviours attributed to lying and experienced a lie bias towards them. It was also found that those who spoke in their non-native tongue i.e. their second language they would be judged as being deceptive compared to first language speakers (Cheng, & Broadhurst, 2005).

Caldwell-Harris and Ayçiçeği-Dinn (2009) conducted a study which sought to establish whether there was a preference in language used when telling a lie. Participants listened to emotionally charged phrases in both their native and non-native tongues. Results of this study found that participants reported lower skin conductance rates when hearing emotional phrases which were spoken in their non-native tongue. The authors also asked the same participants to read both truthful and untruthful statements in both their first language and their second language. Results from the study indicated that there was a higher level of skin conductance, when participants were being deceitful in their non-native tongue compared to when they were lying in their native tongue Thus, lending strength to the argument that people may leak cues associated with deceit when conversing in their non-native tongue.

Da Silva and Leach (2013) conducted a study in which Canadian students' accuracy when judging deception was evaluated. They found that students tended to judge those who provided a statement in their first language where seen as being truthful i.e. there was a truth bias

extended towards them. Furthermore, Da Silva and Leach (2013) found that when making veracity assessments, participants extended a lie bias towards those statements which were generated in the speakers second language. This lie bias was also found by Meissner and Kassin in 2002. Meissner and Kassin (2002) also found that this lie bias was extended even if you were a professional who was trained forensically e.g. law officials also extended this lie bias.

Some studies have reported that a speaker's language proficiency plays a role in an individual's ability to either correctly identify the presence of a lie or correctly reject the absence of a lie (Evans et al., 2017). A contemporary study conducted by Leach, Snellings and Gazaille (2017) sought to investigate whether language proficiency influenced one's ability to detect the presence of deception. Both first language and second language English speakers were filmed whilst telling a truth or a lie. Both native and non-native English laypersons were asked to watch these recordings. Leach and colleagues (2017) found that when participants watch and heard the native English speakers, they extended a truth bias towards them i.e. those recordings were judged as truthful. When reviewing second language recordings, participants judged these as being lies in other words, they extended a lie bias.

Snelling (2013) investigated if a truth bias was extended to those speaking their native tongue. The results yielded by the study indicated that level of language proficiency did indeed impact the ability to discern truth from lie. In addition to this finding, the study also found that native speakers were extended a truth bias when their statements were being reviewed by a reviewer with a high level of proficiency for that language.

Solodukin (2015) investigated if sender's language proficiency could influence receiver veracity judgement bias. Solodukin (2015) tested the receiver's ability to make correct veracity assessments on senders who had varying degrees of proficiency of the English language. As with previous research Solodukin (2015) found that receivers were truth-biased to native speakers.

This could then imply that there is indeed a lie bias extended to those who speak in their second language. The studies result also suggest that language proficiency play a role in receiver's judgements as judges were least accurate in their veracity assessments of the basic second language category.

Studies which have investigated responses biases and second language have yielded more consistent results than those looking at the impact of second language on veracity assessments (Costillo, Tyson, & Mallard, 2014; Evans & Michael, 2014) and that a lie-bias is extended to non-native speakers (Castillo et al., 2014) or at the very least a very minor truth-bias to native speakers (Evans & Michael, 2014).

A review of the extant literature has offered some potential reasons why these bias (truth and lie) phenomena exists toward native vs non-native speakers. According to Wang, Zue, Chen, Xue and Dong (2007) when one speaks in their non-native language must inhibited neural control mechanisms which would have them speaking and responding in their native tongue. This then results in an increase in cognitive load experienced by the second language speaker. As a result of this increase in cognitive load, non-native speakers may exhibit cues which are associated with lying such as being less fluent, nervousness and fidgeting (Wang et al., 2007). These behaviours are typically only displayed when second language speakers are expected to communicate in their second language. Furthermore, these biases may potentially exist because those speaking in their second language may provide stories which are less rich in details, which lie-tellers also tend to do (Cheng & Broadhurst, 2005; DePaulo et al., 2003).

As a result, people may misattribute these naturally occurring tendencies when the speaker speaks in their second language as being indicative of lying. This could potentially explain the truth-bias extended to people who communicate in their first language, and the lie bias extended to those who speaking in their second language.

Evans et al. (2017) asked participants to provide four statements which were auto-biographical and opinion-based statements. These statements were either truthful or deceptive in nature. Observers were asked to rate if they felt if these were either truth or false statements. Observers accuracy at correctly classifying the statement as a truth or a lie was highest for group which was the least proficient English language group. This finding suggests that lie detection would be facilitated if a speaker were to provide a statement in their non-native language (Evans et al., 2017).

People tend to go into a conversation with the intent to believe someone is telling them the truth, in other words they are extended a truth-bias. And in turn this bias results in a truth-based credibility judgement (DePaulo et al, 2003; Vrij, 2008). A potential explanation for this truth-bias is offered by O'Sullivan, Ekman, and Freisen (1988). According to O'Sullivan and colleagues (1988) on average people are exposed to a greater amount of truthful behaviour daily and therefore people believe that deceptive behaviour is rare. As a result, they expect people to be honest in their interactions. Street and Richardson (2015) argued that because of this greater based truth rate, people judged others as being truthful rather than deceptive. Vrij and Baxter (1999) also offer another explanation for people being truth-biased, and their argument lies in social norms. According to Vrij and Baxter (1999) it would be considered impolite to constantly question another's credibility

Contemporary research has found a clear bias being extended to non-native speakers. Results from these contemporary studies indicate that the less proficient one is in speaking the language, the less truthful they are perceived to be (Dragojevic, & Giles, 2016; Hansen, & Dovidio, 2016). This is because if a speaker has an accent the statement may be judged deceptive as it harder to understand (Akehurst, Arnhold, Figueiredo, Turtle & Leach, 2018). It has been found that perceptual fluency increases believability of a message (Unkelbach, 2007).

Research has also suggested that non-native speakers exhibit language which is associated with deceit. Non-native speakers tended to use simpler words, as well as more concrete words as it was easier for them to access these words and non-native speakers avoided using abstract terms, as these are more cognitive complex to use (Newman et al., 2003). Kormos and Dénes (2004) found that non-native speech was less diverse than native speech and was more redundant and was repetitive. These are the same characteristics and patterns which are associated with deceptive speech according to studies (Arciuli, Mallard, & Villar, 2010; Zhou et al., 2014)

Conclusion of literature review

The bulk of the deception literature finds that people are generally poor at detecting when someone is lying to them (Bond & DePaulo, 2006; DePaulo et al., 2003). Various techniques have been proposed, refined and tested as research in the area of detection of deception. Initially physiological measures and the observation of non-verbal cues associated with deception were thought to be the best indicators of deceit. Contemporary avenues of research focus on the language and language constructs used by deceivers. There is also a focus on various interviewing styles to induce measures of cognitive strain on the deceiver to facilitate the detection of the deceit (Vrij, 2018). The literature indicates that the field of study concerned with the use of computer programmes to aid in deception detection is promising, but also that these evolving techniques are not without their criticisms. Further, fields of enquiry into the influence of language proficiency on veracity assessments indicates that a communicator's language proficiency impacts on a receiver's veracity assessment of the message.

Rationale

This study was restricted to the three dominant languages used in the Western Cape being those of Afrikaans, isiXhosa and English. The 2011 census, as reported by southafrica-info.com,

found that these three languages represent 98.8% of the reported first languages in the Western Cape province. Table 1 summarises the percentage usage for the three languages as a first language at provincial level (Western Cape only) and at a national level both first and second language percentages.

Table 1
isiXhosa, Afrikaans and English Language % users in South Africa

	Western Cape (% of WC population)	National Statistics (% of RSA population)		
	1st language users	1st language	2nd language	Combined share
isiXhosa	24.7%	16.0 %	21.2%	37.2%
Afrikaans	53.8%	13.5 %	19.9%	33.4%
English	20.3%	9.6%	21.2%	30.8%

Note:
% use of total population of 51.77 million people (2011 census)
Source: https://southafrica-info.com

Despite having 11 official languages, and only being spoken by 9.6% of the overall population in South Africa as a first language and 20.3% at Western Cape provincial level, English remains the lingua franca of the country. This means that most people (90.4% at national level, 79.7% at Western Cape provincial level) have to communicate or transact in a non-native language, being English, on day-to-day routine and business dealings.

Although sparse, the literature pertaining to second language use and perceptions of deception indicates that receivers of communication are prone to demonstrating biases that are functions of whether the communication being received is by a native or non-native language speaker of the communication language. The studies report that a truth bias is likely to be extended to those speaking in their native-tongue and a lie-bias when the speaker is engaging in their non-native tongue (Cheng, & Broadhurst, 2005; Da Silva, & Leach, 2013; Lev-Ari, & Keysar, 2010)

If indeed a lie bias is preferentially extended to non-native speakers it disadvantages the non-native speaker. Within the South African context, both at national (90.2% of the population) and at Western Cape provincial level (79.7% of provincial population) of people could be prone to be subject to bias.

The ability to correctly detect when someone is lying has often been compared to that of a coin toss i.e. there is a fifty percent chance of identifying it such. Literature reviews indicate that the notion of placing cognitive load on the individual while they are trying to lie, improves the ability of someone to correctly discern whether they are being lied to or not (Vrij et al, 2008; Walczyk et al., 2013). If the addition of cognitive load at the time of statement gathering makes it easier to discern whether someone was lying or not when evaluating a written, or transcribed verbal statement, it could be of benefit to a number of applications including, but not limited to forensic investigations.

The number of studies published within the last decade to detect deception in written statements has grown significantly. This task is often undertaken by making use of natural language processing (NLP) programmes. These programmes typically outperform humans when it comes to correctly identifying when deceit is evident in the written statement. This study used the Linguistic Inquiry Word Count program with its 2015 dictionary version to assess this line of research.

In summary, the results of the study are relevant within the multi-lingual South African context as it would determine whether biases (lie or truth) are prevalent in South Africa based on whether the communication is being done in a native (first) or non-native (second) language. The study also assesses whether cognitive load and the use of NLPs in assessing transcribed statements improves the detection of deception accuracy rates compared to human judges; the outcome thereof is relevant to professions reliant on information collected from statements.

Specific Aims and Hypotheses

The aims of this study were four-fold. The first avenue of enquiry determines the extent to which either a lie or a truth bias is afforded to South African speakers when the veracity of a statement is assessed. This first avenue of enquiry is assessed through three refinements to the enquiry: (a) whether veracity assessment outcomes are biased towards assessing veracity as untruthful for statements provided in a non-native (second) language, (b) whether the veracity assessment outcomes are biased towards assessing veracity as truthful for statements provided in a native (first) language, and (c) what the differences are in veracity assessment outcome accuracy comparing the outcomes of first and second language statements. The second avenue of enquiry assesses whether the introduction of cognitive load during the recording of a verbal statement impacts the veracity assessment of a transcribed version of statement. The third avenue of enquiry assesses the extent to which the original statement veracity (truth or lie) impacts the veracity judgement outcome. The final avenue of enquiry aims to discern whether linguistic features and profiles of the transcribed statements, as assessed by LIWC2015, are better at detecting deception in transcribed accounts, when compared to that of human judges.

To assess these four avenues of enquiry, six research questions and hypotheses were posed.

Research Questions

Research question 1. Is there a lie bias present when making veracity assessments of second language statements? This research question investigates whether second language statements are preferentially assessed as untruthful, regardless of the original statement veracity.

Research question 2. Is there a truth bias present when making veracity assessments of first language statements? This research question investigates whether first language statements are preferentially assessed as truthful, regardless of the original statement veracity.

Research question 3. Is there a difference in the accuracy of veracity assessments observed when comparing the veracity assessment outcomes of first language statements against the veracity assessment outcome accuracy of second language statements? This research question investigates the possible extent to which language usage (reflected in this study through identification with the statement language as a first or second language) could influence veracity assessment outcomes.

Research question 4. Does the original statement veracity impact the veracity assessment outcome? This research question investigates whether what is known as the veracity effect is influencing veracity assessment outcomes.

Research question 5. Does an increase in cognitive load whilst providing a verbal statement in a second language make it easier to correctly assess the veracity of the transcribed statement? This research question considered the impact that cognitive load has on the veracity assessment of statements.

Research question 6. Does the computer program LIWC, using the 2015 dictionary definitions perform better than human evaluators of statements in detecting deception in second language speakers?

Hypotheses

Hypothesis 1. Reviewers are subject to a lie bias when reviewing second language written statements.

Hypothesis 2. Reviewers are subject to a truth bias when reviewing first language written statements.

Hypothesis 3. There is a difference in veracity assessment outcome accuracy between first and second language statement assessments.

Hypothesis 4. The original statement veracity influences the veracity assessment outcome.

Hypothesis 5. The presence of cognitive load, when providing the statement makes it easier for the subsequent assessment by a veracity judge to detect deception.

Hypothesis 6. LIWC2015 will be more accurate in determining the veracity of written statements compared to human judges.

The next chapter outlines the methods used in this thesis. The chapter describes the procedure followed in the two data gathering phases of this project. The chapter thereafter reports the data analysis conducted and the results of the analysis.

Chapter 3 Methods

Research Design and Setting

This exploratory study adopted an experimental design to assess whether statement veracity assessment outcomes are impacted by (a) considering whether the statement was provided in a non-native or a native language of the statement provider, (b) placing cognitive load on the statement provider, and (c) the original statement veracity, being either truthful or not. The study also explored whether language profiles of the English statements in this study, as determined by the software program LIWC 2015, could be used to assess veracity outcomes with higher accuracy compared to the human judges.

In order to assess whether the statement language, identified as either a native or non-native language of the statement provider, affects the veracity assessment outcome, three research questions were posed. The first of these were whether statements provided in a non-native language are more likely to be judged as untruthful (i.e. that a lie bias is present when the veracity of statements provided in non-native languages are assessed). The second was an assessment of the converse, i.e. assessing whether statements provided in a native language is more likely to be judged as truthful (i.e. a truth bias is present when judging the veracity of statements provided in native languages). The third research question stepped away from the investigation of biases and asked whether the accuracy of veracity assessments is impacted by whether the statement was provided in a non-native or a native language of the statement provider.

To provide a more nuanced understanding of the factors which influence veracity assessment outcomes, this study was expanded to include an additional research question to also assess whether the original provided statement's veracity impacted the veracity judgment accuracy. The empirical findings of this study formulated under this research question were

evaluated against empirical concepts and theories such as the veracity effect and the Truth-Default Theory.

This study sought to replicate previous empirical findings on the impact of cognitive load as described in the preceding chapter. This study assessed whether cognitive load would enhance the ability to correct differentiate between truth and lie as suggested in previous studies.

Finally, this research sought to determine if the software programme LIWC2015 is able to differentiate between truth and lie, by evaluating linguistic profiles reported in the literature associated with deceptive statements. The study compared the software veracity assessment outcomes with the assessment outcomes of human evaluators.

Participants

This study's participants were recruited through both convenience and stratified random sampling for both phase 1 (statement generation) and phase 2 (veracity assessment). Participants for both phases were expected have a good command of English (either as a first or as a second language) as well as a good command of either Afrikaans or isiXhosa as the alternate language. These three languages are the dominant local languages in the Western Cape where the study was conducted.

Given the three languages selected for this study, six combinations as first and second language pairings were possible. The combinations are: English first language, isiXhosa second language ("E1X2"); English first language, Afrikaans second language ("E1A2"); isiXhosa first language, English second language ("X1E2"); isiXhosa first language, Afrikaans second language ("X1A2"); Afrikaans first language, English second language ("A1E2") and Afrikaans first language, isiXhosa second language ("A2X1). The Afrikaans, isiXhosa combination of first and second languages, as well as English first language with isiXhosa second language were not well represented within the target recruitment population of the UCT campus humanities

students. Only three of the six language combinations were successfully recruited for the statement generation phase of the study as indicated below.

Table 2
Number of First and Second Language combinations recruited.

First/Second language pairing (phase 1 participants)	N
English 1st language, Afrikaans 2nd language (E1A2)	26
Afrikaans 1st language, English 2nd language (A1E2)	16
isiXhosa 1st language, English 2nd language (X1E2)	12
'Other' combinations with English	8
Total number of statement generators	62

The recruitment was done through a call for participation placed on the university's intranet site to which all UCT students have access to. Sixty-five interviews were scheduled following the time allowed for responses; three participants did not show up for the interviews and one candidate chose not to continue during the interview session. This resulted in sixty-one (n = 61) participants' statements recorded and carried forward to the veracity assessment stage. The age and gender profile for the 61 participants in the statement generation phase was 40 female (M_{age} = 19.90; SD_{age} = 1.92) and 21 males (M_{age} = 20.62; SD_{age} = 2.20).

An excellent response of five hundred and sixteen (n = 516) students were achieved for participation in the veracity assessment stage of the study. Statement providers (phase 1 participants) were excluded from taking part in the veracity adjudication stage of the study. The age and gender profile for the 516 adjudicators was 385 females (M_{age} = 19.60; SD_{age} = 2.63) and 126 males (M_{age} = 19.83; SD_{age} = 2.32). Five participants provided non-gender specific responses to the gender question (Mage = 19.60; SD_{age} = 0.55).

Procedure

Pilot studies (statement generation). Two pilot studies were performed to assess the practicality of the envisaged statement generation process prior to actual data collection. The first, conducted in a group setting with post-graduate students and faculty members in attendance, recommended two changes to the intended statement generation procedure. It was suggested that the odd numbered Likert scales provided participants with an option to respond neutrally to questions aimed at determining the extent to which the participant agreed or not with a view expressed on a topic. The original secondary task expected of the participants to induce cognitive load was deemed cognitively too taxing with a risk of rendering statements provided under cognitive load to be of no use.

An undergraduate student from the psychology department was recruited to trial run the amended statement gathering procedure the following week. Two additional minor changes followed the second trial run: the statement duration was shortened from five to three minutes, and the (secondary) cognitive load task further simplified. The secondary task agreed on at this stage was that participants providing statements under cognitive load were to tap their finger in sync with a metronome during the interview process.

Statement Generation. The primary researcher gathered statements in individual face-to-face interviews over a two-week period, with six one-hour sessions scheduled every day (60 meetings). Five additional ad-hoc interviews followed the following week, targeting first-language Afrikaans speaking participants to achieve a balanced representation of language combinations in the design. Three potential participants did not show up for the interviews

during the three-week statement generation phase and one candidate chose not to continue during the interview session.

At the start of the phase 1 interviews, the researcher outlined the purposes of the study and explained that participants will first be required to indicate the extent to which they agreed with opinions expressed on a range of contentious topics, where after two topics would be selected from this list. Participants would then be required to provide two statements on these topics in their first language, followed by two statements in their second language. The nature of these two statements were not provided at this stage. The researcher provided the participant an opportunity to clarify matters that might arise from the introduction where after participants completed and signed a consent form indicating their agreement to participate in the study (Appendix 1 and 2).

Two bags containing six tokens each (three black, three white in one bag, three green, three blue in the second bag) were prepared for each day. The tokens would determine the sequence in which Truth/Lie statements would be provided by participants, as well as whether the statements were to be provided under cognitive load or not. The draws from the bags at the beginning of each session were done without replacement of tokens, which resulted in an even balance of conditions implemented as determined by the draws at the end of each day.

Topic selection. After consenting to take part in the study, the researcher provided participants with a list of twelve topics, each followed with an opinion expressed either for or against the topic. Participants were asked to indicate the extent to which they agreed or disagreed with the opinion expressed on each topic using a six-point Likert scale. (Appendix 3). The last question on the opinion-evaluating questionnaire was a self-reflection question, asking participants to indicate how honest they were in completing the ratings.

Once all twelve statements were responded to, the researcher established with the participant whether any of the topics on the list were deemed too contentions for the participant if they selected to continue in the next phase of the interview. If a participant indicated discomfort with any topic, the specific topic was removed from the list of statements available for selection for the individual participant. No justification for removal of topics was required.

After eliminating topics that the participant was not prepared to engage on, the researcher informed the participants with additional information, not provided before rating the twelve statements. Participants were now informed that two topics would be selected from their responses and that one truthful and one untruthful statement in the participant's first language will be required for the first topic, followed by another set of one truthful and one untruthful statement in the statement provider's second language for the second topic. The participants were also informed at this stage that the statements would be audio recorded, transcribed and presented for review and veracity adjudication in a subsequent phase by a different group of participants. The researcher explained that in the subsequent phase the statement reviewers would be asked to assess whether the statements provided by the participant were truthful or deceptive and provided assurances to the participant that no identification of the statement providers would be possible. Having now being informed that continuing in the study would require participants to be untruthful in statements and that they would be audio recorded, participants were asked to provide consent again before continuing (Appendix 3).

The second consent form was required as the procedure initially described leading up to the first consent form was incomplete. This was by design. The requirement that participants would be asked to provide untruthful statements was withheld initially to mitigate against the possibility that prior knowledge of this requirement could influence how participants assess a topic in anticipation of having to provide an untruthful statement on a topic. All except one participant agreed to continue at this stage. Once the amended consent was provided, the

researcher selected two topics from the statement evaluation list, giving preference to selecting topics where extreme scores of agreement or disagreements were indicated.

First-Language Statements. The statement generation procedure followed a technique developed by Leal, Vrij, Mann and Fisher (2010), known as the Devil's Advocate Approach. The Devil's Advocate approach starts with an opinion-eliciting question, establishing a participant's opinion (either agreement or disagreement) on a topic. Once this position is established, the participant provides a statement in defence of the opinion expressed. The Devil's Advocate question follows after this statement requiring the participant to justify an opposing view to the one expressed before in as convincing a manner as possible. This results in a pair of statements, one a truth and one a lie on the same topic by each statement provider. This process was then repeated for the second selected topic in the participant's second language.

In recognition that participants may be more settled by the time the second statement is required, the sequence of whether the first statement was to be truth or lie was randomised by drawing tokens from the first bag of six tokens. As the tokens were drawn from the bag without replacement, and given that each bag was prepared each day with three black and three white tokens, the collection of truth first verse untruth statements first were balanced over the statement generation phase. The participants were asked not to reveal to the researcher whether their first statement was to be a truth or a lie initially.

This statement generation followed. The researcher asked the participant to express an opinion on the statement provided for the first of the selected topics. If the participant was required to tell a truth (as indicated by the token drawn), the participant would respond with the view expressed on the topic questionnaire; if the participant was required to be untruthful (as indicated by the token drawn), the participant would respond with a contrary view to the one expressed on the topic questionnaire.

Once this position has been established, the researcher asks the opinion-eliciting question; “What do you think led you to have that opinion,” referring to the response offered to the first question. Participants then justified the opinion offered (truthful or untruthful depending on the response to the first question) in a three minutes statement. A timer, visible to both researcher and participant, kept track of time during the actual statement generation.

At the end of this first statement, the researcher continued the interview by asking the devil’s advocate question as follows “Playing devil’s advocate, is there anything you can say in favour of, (or against) the topic?” Participants had to say something in agreement of the topic at this stage if their first statement expressed disagreement, alternatively if their first statement was in agreement with the topic, the devil’s advocate question required of them to provide a statement disagreeing with the topic.

Second-Language Statements. The procedure was repeated for the second topic, the only difference being that the statements were now required in the statement provider’s second language.

Cognitive Load. At the same time as drawing the ‘start with truth/lie’ token to establish the sequence in which truth/lie statements was required, participants also drew a token from a second bag. This token would indicate whether their statements (both first-language and second-language) would be provided under cognitive load or not. The drawing without replacement of the second bag ensured that a balanced representation between statements provided under cognitive load and statements with no cognitive load was collected during the data collection phase. The secondary task required from participants selected to provide statements under cognitive load was to listen to and tap a finger on the desk in sync with the beat of a metronome whilst providing the statements.

Each of the 61 phase 1 participants provided four statements, one truth, one lie in the participant's home language and one truth, one lie in the participant's second language. Half of the participants (n = 30) provided the four statements under cognitive load, and half (n = 31) under no cognitive load. One participant (having drawn a cognitive load token) withdrew during the interview. This yielded 264 statements available for veracity adjudications in the second phase of data collection.

Transcription. The isiXhosa statement providers transcribed their own statements immediately post interview if sufficient time remained. All the first-language isiXhosa-speaking participants agreed to this. The researcher asked a native isiXhosa speaker, not involved in the study, to verify the transcriptions by listening to the audio recording whilst reading the transcriptions of each isiXhosa participant; and to complete transcriptions in the few instances where a participant had insufficient time after the scheduled interviews to complete the transcription. The Afrikaans and English statements were transcribed by the primary researcher.

The following meta-data was recorded with each statement, a) the statement language i.e. English, Afrikaans, isiXhosa; b) whether the statement language was a native (first) or non-native (second) language of the statement provider; c) whether the response to the opinion eliciting question was a truth or a lie; and d) whether the provided statement was given under cognitive load or not.

General procedure for phase 2. Combining the three language combinations recruited with the two possible statement veracities (truth/untruth) and the presence/absence of cognitive load (2 possibilities) yields 12 possible combinations of attributes. The lowest number of statements within one of these twelve groups of attributes (Afrikaans as first language) was 5 statements. The 244 statements were then reduced to 120 by randomly dropping statements from

oversampled groups to five statements within a group. This yielded a balanced representation of statements attribute combinations for adjudication in phase 2 as set out in table 3.

Table 3
Number of statements retained across all fixed effect attributes

Statement Language	Truth / Lie	Cognitive Load?	Number
Afrikaans First language	Lie	No Cognitive Load	5
		With Cognitive Load	5
	Truth	No Cognitive Load	5
		With Cognitive Load	5
Afrikaans Second Language (English first)	Lie	No Cognitive Load	5
		With Cognitive Load	5
	Truth	No Cognitive Load	5
		With Cognitive Load	5
English First Language	Lie	No Cognitive Load	5
		With Cognitive Load	5
	Truth	No Cognitive Load	5
		With Cognitive Load	5
English Second Language (Afrikaans first)	Lie	No Cognitive Load	5
		With Cognitive Load	5
	Truth	No Cognitive Load	5
		With Cognitive Load	5
English Second Language (isiXhosa first)	Lie	No Cognitive Load	5
		With Cognitive Load	5
	Truth	No Cognitive Load	5
		With Cognitive Load	5
isiXhosa First language	Lie	No Cognitive Load	5
		With Cognitive Load	5
	Truth	No Cognitive Load	5
		With Cognitive Load	5

The one hundred Afrikaans and English statements (n = 100) were distributed into twenty surveys, each survey containing five statements. Each questionnaire consisted of one truthful first language statement, one untruthful first language statement, one truthful second language

statement and one untruthful second language statement. The fifth statement's combination of statement attributes varied across questionnaires but when viewed across the 20 questionnaires (for 100 statements) all possible combinations of statement attributes were balanced. Similarly, four isiXhosa / English surveys were designed, each with five statements, as was the case for the English / Afrikaans statements.

The adjudication of the veracity of statements was done using an online survey tool, SurveyMonkey (www.surveymonkey.com). Students indicated their interest in participating in the statement adjudication phase by responding to a call for participation via email. In response, the researcher sent participants the appropriate link to a survey (i.e. a link to an Afrikaans/English survey; or alternately to an English/isiXhosa survey). Each of the 24 surveys (5 statements per survey, 120 statements total) also contained a set of demographic questions. The demographic questions included the participants' age, gender, home province and the participant's first and second language identification. The request for demographic information was followed by the 5 statements. Following each individual statement, the same ten questions pertaining to the statement was asked. The same ten questions were asked for each of the statements in the questionnaire.

The following instructions were given to reviewers "*You are expected to read the five statements which follow and then answer the same questions per statement. It may be useful to think of yourself as a judge at a debating contest when you are answering the questions per statement."* Reviewers were deliberately not informed that they would be deciding if a statement was truthful or deceptive. This was a conscious decision made on the researcher's part as to avoid the influence of social desirability bias i.e. to ensure that participants did not review a question in a specific way which they perceived to be correct.

The SurveyMonkey responses were collated into a single file and combined with the statement attributes for subsequent analysis. Twenty-nine attributes (metadata variables) were recorded for each statement review.

Data Analysis

Both fixed effects and random effects had been identified in the data collected for analysis. The original design of the experiment considered three main effects, the original statement veracity, whether the statement was provided in a first or second language and whether the statement was collected under cognitive load or not. During analysis it became evident that understanding whether the language is being reviewed in a first or second language (of the veracity judge) could also be germane to this study and a fourth main effect, that of Reviewer first/second language was added to the original three anticipated main effects.

Given that all fixed effects variables, as well as the veracity-adjudication outcome variable were categorical and dichotomous, a logistic regression GLMM model was required to extrapolate the experiment's findings to the bigger population and also to assess which of the effects are significant in affecting the modelled outcome of the fixed effects. The most suitable GLMM model definition was determined through iterative comparisons of different variable specifications in the modelling. As the input variables are categorical in nature, logistic regression models were required and the criteria used to determine whether one model provides an improved fit or not above another was the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) criteria. The 'better' model in an AIC criteria comparison is the model with the lower AIC score. If the actual AIC scores difference between models are less than two, the two models are deemed to have similar 'goodness of fit' scores of the logistic regression of the models. Where appropriate and the degrees of freedom allowed for it, the

model outputs were compared with ANOVA assessments. Chi-square significance testing was conducted for differences in the states of the dichotomous variables where required.

All statistical analysis was done using R version 3.5.3 (R Core Team, 2019). R packages and modules used for the GLMM modelling and the assessment of the models included lme4 (Bates, Maechler, Bolker, Walker, 2015), psycho (Makowski, 2018), merTools (Knowles & Frederick, 2019), DHARMa (Hartig, 2019), data.table (Dowle & Srinivasan, 2019) and sjPlot (Lüdecke, 2018). Graphs were generated using module ggplot2 (Wickham, 2016). All R commands were run in RStudio version 1.1.463.

Incomplete statement responses (not all questions following a statement answered), as well as statement reviews received from reviewers that reported either their first language or their second language as a language not within the group of Afrikaans, English or isiXhosa was excluded (n = 453) from subsequent analysis. One thousand nine hundred and seventy-two (n = 1,972) statements were retained for further analysis.

The main effects considered to assess the first five research hypotheses were; whether the veracity of the statement was a truth or a lie; whether the statement was provided in the first or second language of the statement provider; whether the statement was provided under cognitive load or not, and whether the assessment of the provided statement was done in the first or second language of the statement reviewer. Table 4 on page 67, depicts the frequency distribution of the 1,972 adjudicated statements into the sixteen possible combinations of the 4 dichotomous statement attributes. The number of statements across combinations down to the eight possible combinations of the three statement generation attributes are balanced. The inability to recruit sufficient second language isiXhosa participants, as well as insufficient Afrikaans second language combined with isiXhosa first language participants for the statement adjudication phase skews representation between first and second language reviews when considering the

reviewers' language correspondence to statement language, the fourth dichotomous factor in the distribution table.

The random effects considered in the study were (a) possible influences of individual statement providers on the overall fit of a GLMM model, (b) possible influences associated with individual statement reviewers, or (c) the subject matter (topic) of the statements.

The age and gender distribution of statement providers were re-assessed after reducing the number of statements from 2,425 to 1,972. Thirty-nine (n = 39) of the original sixty-one statement providers remained. The 39 retained statement providers consisted of twenty-six females (M_{age} = 19.73; SD_{age} = 1.61) and thirteen males (M_{age} = 21.00; SD_{age} = 2.58).

The age and gender evaluation of statement reviewers were re-assessed after the number of statements were reduced from 2,425 to 1,972. Four hundred and seventeen (n = 417) statement reviewers remained. The re-assessed age/gender profile for the 417 reviewers of the retained statements was 313 females (M_{age} = 19.74; SD_{age} = 2.78) years and 99 males (M_{age} = 19.89; SD_{age} = 2.50) with 5 participants responding non-gender specific (M_{age} = 19.60; SD_{age} = 0.55).

Table 4
Frequency scores of statements reviewed across fixed effect variables

Statements for adjudication (N = 1,972)															
Statement Veracity: TRUE *n* = 995								Statement Veracity: UNTRUE *n* = 977							
First language statements *n* = 491				Second language statements *n* = 504				First language statements *n* = 477				Second language statements *n* = 500			
CL: Yes *n* = 246		CL: No *n* = 245		CL: Yes *n* = 252		CL: No *n* = 252		CL: Yes *n* = 241		CL: No *n* = 236		CL: Yes *n* = 242		CL: No *n* = 258	
RvL1 *n* =148	RvL2 98	RvL1 163	RvL2 82	RvL1 148	RvL2 104	RvL1 149	RvL2 103	RvL1 152	RvL2 89	RvL1 142	RvL2 94	RvL1 162	RvL2 80	RvL1 180	RvL2 78

Notes
First/Second language statements: Statements provided in statement provider's first/second language
CL: YES – Statement provided under cognitive load;
CL: NO – No cognitive load imposed during statement provision
RvL1: Statement language is same as statement reviewer's first language
RvL2: Statement language is same as statement reviewer's second language

To address the sixth research question, the accuracy of the veracity adjudication outcomes of human judges compared against the accuracy of the veracity judgement outcomes using a Natural Language Processing (NLP) programme was done using LIWC2015 (Linguistic Inquiry and Word Count, 2015 version). All available English statements (n = 102) were used for the comparison.

Every word in each English statement was compared against LIWC2015's dictionary of approximately 6,400 words, word stems and emoticons. For each statement, a language profile consisting of 90 output variables were created in the LIWC2015 analysis. The output variables consist of summary variables, a number of individual word category variables as well as grouped word categories variables. Analysis (comparisons against known linguistic profiles) can be done on grouped categories, such as "negative emotions" or on the individual constituent word categories within a grouped category such as "sadness" words, "anxiety" words or "anger" words within the "negative emotions" word category. The LIWC2015 linguistic profile of each statement was compared against three different linguistic profiles, one using a reality monitoring framework for deceptive statements as proposed by, Bond and Lee (2005) the original NP framework linguistic profile for deceptive statements proposed by Pennebaker and colleagues (2003)and against a linguistic profile created from a post-hoc analysis of the English statements used in this study.

Ethical considerations.

This research was carried out in accordance to the guidelines of the Health Professions Council of South Africa (HPCSA) and the University of Cape Town's Codes for Research. The study received ethical approval in February 2018 (reference PSY2018-003). All participants had to indicate informed consent by signing a consent. All consent forms outlined the requirements and expectations for the particular phase.

Participants were assured that no participant identifying information would be made available within the study. All identifying data was kept confidential, all hard copy information (consent form, demographic forms) were kept in a secure cabinet to which only the researcher had access during the study. Audio recordings of participant's statements contained no identifying information and were kept in password-protected folders. Participants were informed that they can withdraw at any stage without any penalties or repercussions. Participants were informed that upon withdrawal they would be provided an opportunity to instruct that the statements or any other information provided by them be removed from the study's data repository.

Chapter 4 Results

This chapter reports the statistical analysis performed of the data gathered as described in the previous chapter. Descriptive statistics of selected meta-data as well as the fixed and random effect variables are followed by summary overviews of the veracity assessment outcomes for the four fixed effect variables inspected individually. The interactions between the fixed effect variables on the veracity assessments then follow as does the impact that the three random effect variables have on the veracity assessment outcomes. All reported confidence intervals, whether indicated in tables or on figures are at 95%

Following the descriptive analysis, the stepwise evaluation of generalised linear mixed models (GLMM) constructed to extrapolate the study findings to a generalized population is discussed. The effect of the fixed effects and the impact (effect ranges) of considering different random effects are evaluated step-wise to determine the generalized linear mixed modelling (GLMM) model with the lowest AIC scores as described in the data analysis section.

The GLMM modelling was done using the lme4 (Bates et al. 2015) package in R. The R-package dHARMa (Hartig, 2019) is used to evaluate the GLMM models' residuals to assess whether the GLMM model presents problems such under/over dispersion, etc.

Descriptive statistics

Survey Duration

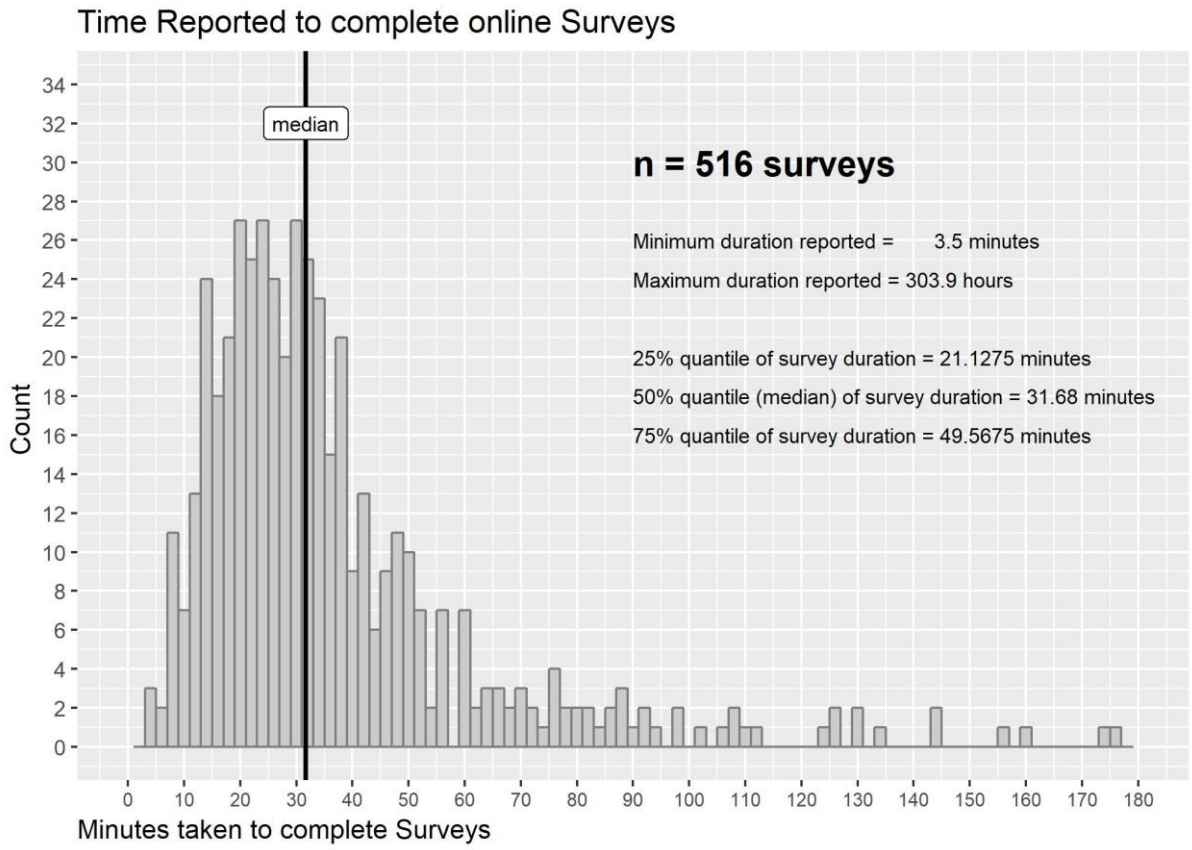

Figure 1 Reported duration of survey completion

Figure 1, above is a histogram of the duration of the on-line questionnaires accessed. It is likely that the outliers on the high side represents surveys left logged in and that eventually timed out or possibly another user logging in on a communal PC. The outliers to the left represent incomplete surveys.

The selection of whether a reviewed statement was carried forward into the analysis phase was based on criteria such as completeness (all questions answered for a statement) and whether the self-reported first and second language identification of participants were within the language combinations of Afrikaans, English and isiXhosa. Ninety-nine (*n* = 99*)* respondents' surveys were eliminated based on the reported language criteria, leaving four hundred and seventeen (*n* = 417) surveys. An additional fifty-three statements were eliminated from this

collection of 2,035 statements based on whether all questions for a statement was answered. One thousand, nine hundred and seventy-two (*N*=1,972) completed statements were thus retained from the original number of 2,580 statements received from the 516 surveys returned by the on-line surveying tool.

Fixed and Main Effects

The binary outcome variable, the four dichotomous fixed effect variables, the three random effect variables considered in the study are summarized in table 5.

Table 5
Summary of fixed and random effect variables

	N	*Value ranges*
Outcome	1,972	Correct; Incorrect
FE: Statement Veracity	1,972	Truth; Lie
FE: Statement 1st/2nd Lang	1,972	1st Language; 2nd Language
FE: Reviewer1st/2nd Lang	1,972	1st Language; 2nd Language
FE: Cognitive Load	1,972	CL Present; CL Not Present
RE: Statement Provider	1,972	ID# 1 – 39
RE: Statement Topic	1,972	Topic # 1 – 12
RE: Statement Reviewer	1,972	ID# 1 - 417

Notes:
FE: Fixed effect variable; RE: Random Effect Variable

Fixed Effects The ratio of veracity judgement outcomes (correct vs. incorrect) expressed as percentages of the 1,972 statements judged are depicted as a function of each of the four fixed effect variables individually in figure 1, page 73. The confidence intervals reported are 95% confidence intervals. Only the veracity judgement outcomes grouped by original statement veracity shows a marked difference between correct and incorrect assessment outcomes. The number of correct veracity judgement outcomes, expressed as a percentage of the total number of assessments, in the case of a truthful statement was 73.9% CI [71.0, 76.6]. In the case where the statement was untruthful the number of incorrect judgements were 70.3%, CI [67.3, 73.2]. Statement adjudicators preferentially judged the veracity of statements as ‘truths’ regardless of

the actual statement veracity. The differences between truth and lie assessments (expressed as a percentage) were 40.6% in the case of assessing untruthful statements and 47.8% in the case of assessing truthful statements. The difference (expressed as a percentage) between correct and incorrect judgements of veracity is reported as significant $\chi^2(1) = 383.83$, $p < 0.001$.

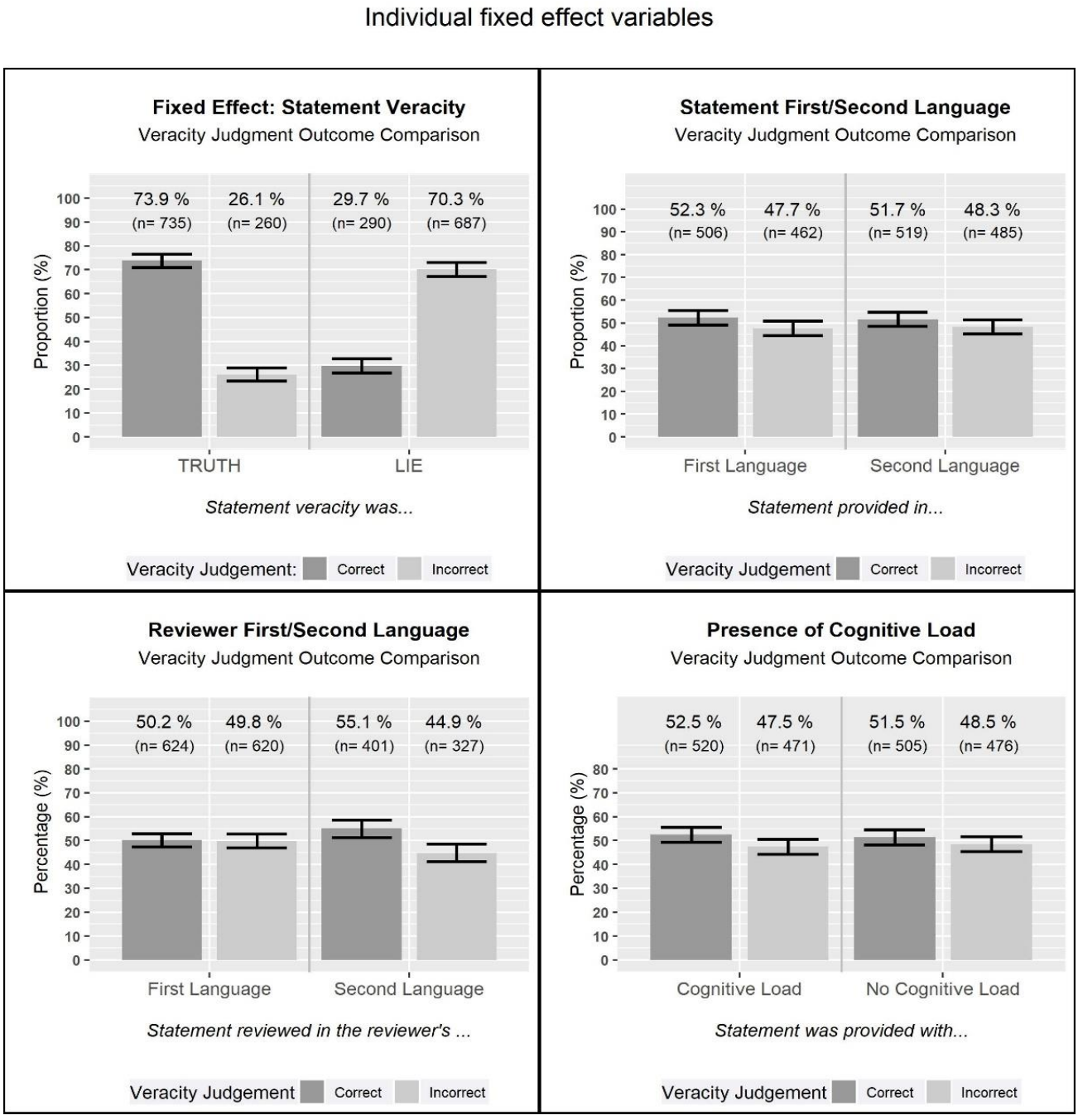

Figure 2 Fixed Effect variables, individually summarised veracity assessment outcomes

The other three fixed effect variables, first/second language statements, first/second language reviews and presence of cognitive load does not indicate marked differences between correct and incorrect veracity judgement outcomes when evaluated individually and independent of one another.

Random Effects The three random effects considered, in this study, for possible impact on the accuracy of the veracity assessments were (a) statements belonging to individual statement providers, (b) statements assessed by individual statement reviewers, and (c) statements grouped by topic.

The veracity assessment outcomes viewed as a function of individual statement provider (n = 39) in figure 3, page 76 shows the differences between correct and incorrect veracity assessment outcomes when the veracity statement outcomes are grouped by each of the 39 statement providers retained after the filtering process described earlier.

The veracity assessment outcomes viewed as a function of the different statement topics (n = 12) in figure 4, page 77 shows variation in the differences between correct and incorrect veracity judgements when the veracity statement outcomes are grouped by the 12 different topics available to statement providers. Equal representation of the topic number was not controlled for. Topics were selected based on the extent to which participants indicated agreement with, or opposition to a list of topics. Care should therefore be taken in comparing the percentage differences observed across topic numbers. As an example, topic 5 (Sport inequality in South Africa) whilst indicating the largest difference in the proportional expression of correct vs. incorrect assessments (30.6% correct vs. 69.4% incorrect veracity assessments), only 36 of the 1,972 (< 1%) statements assessed dealt with topic 5.

Figure 5, page 78 depicts the comparison of correct vs incorrect veracity judgement outcomes as a function of the individual reviewers (n = 417). The outcomes are presented as raw

frequency counts in this instance. Each statement veracity judge (phase 2 reviews) were asked to review five statements. Where a statement judge skipped one or more of the questions posed after each of the five statements, the statement was removed as described earlier. This explains why there are a few reviewers depicted in figure 5 with less than 5 statements remaining. The maximum number of 5 statements possible per reviewer represents less than 0.01% of the 1,972 statements retained for analysis.

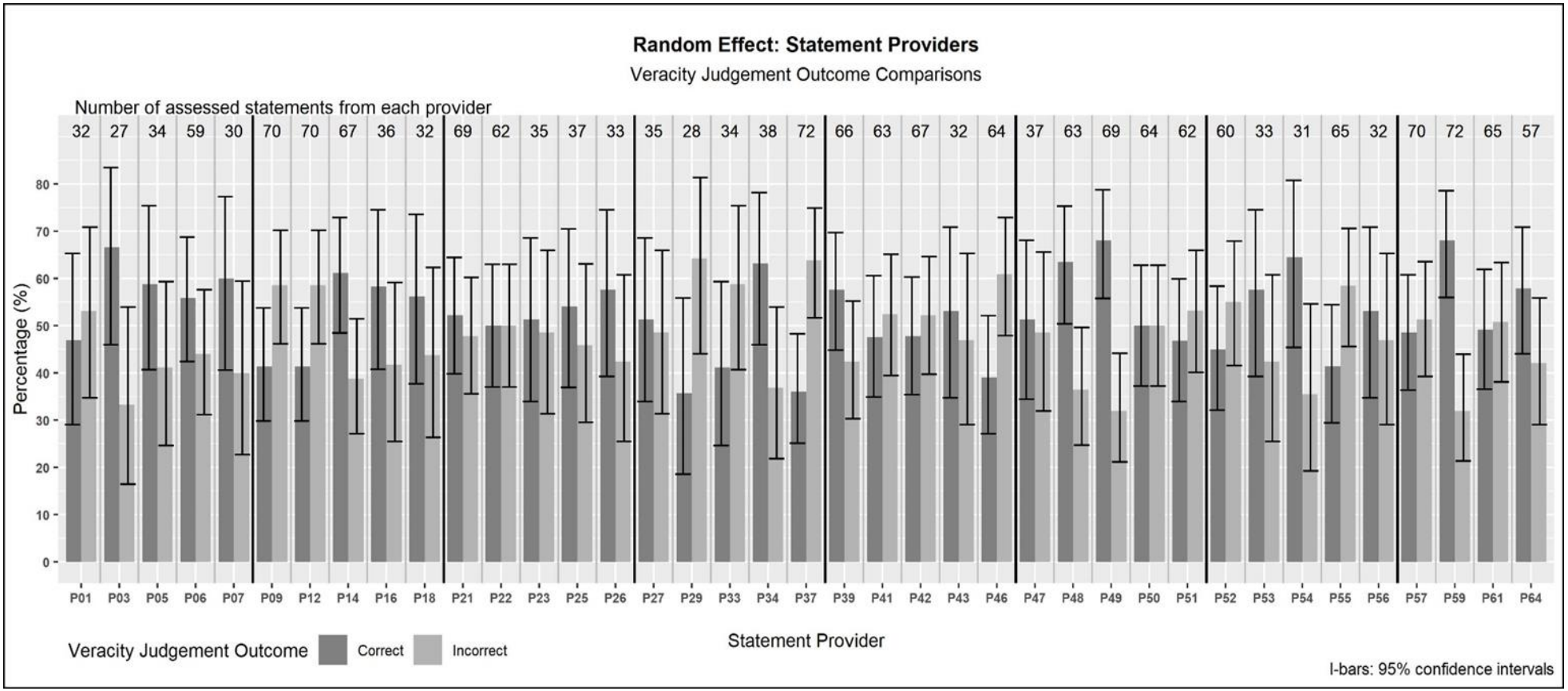

Figure 3 Outcome assessments grouped by the 39 individual statement providers remaining post filtering

Note: X-axis reports original Participant number

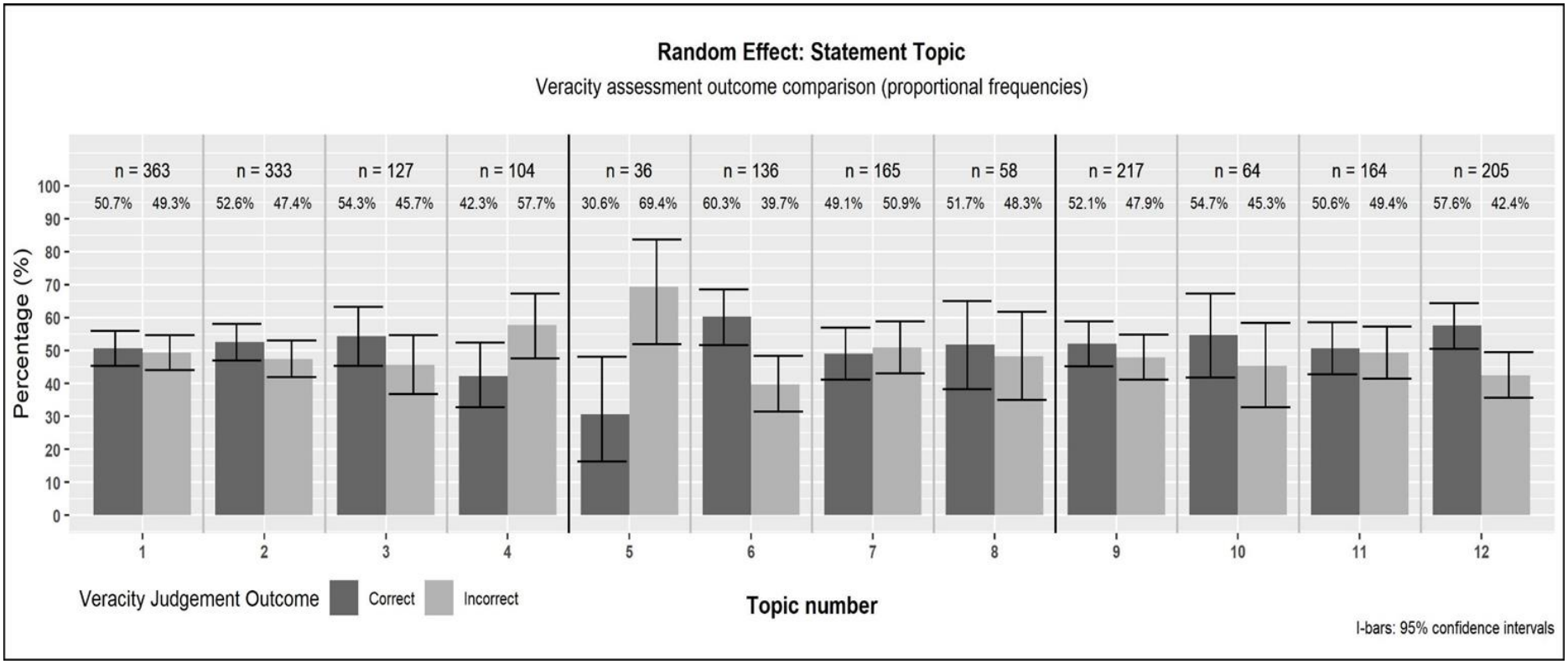

Figure 4 Outcome assessments depicted as a function of individual topic

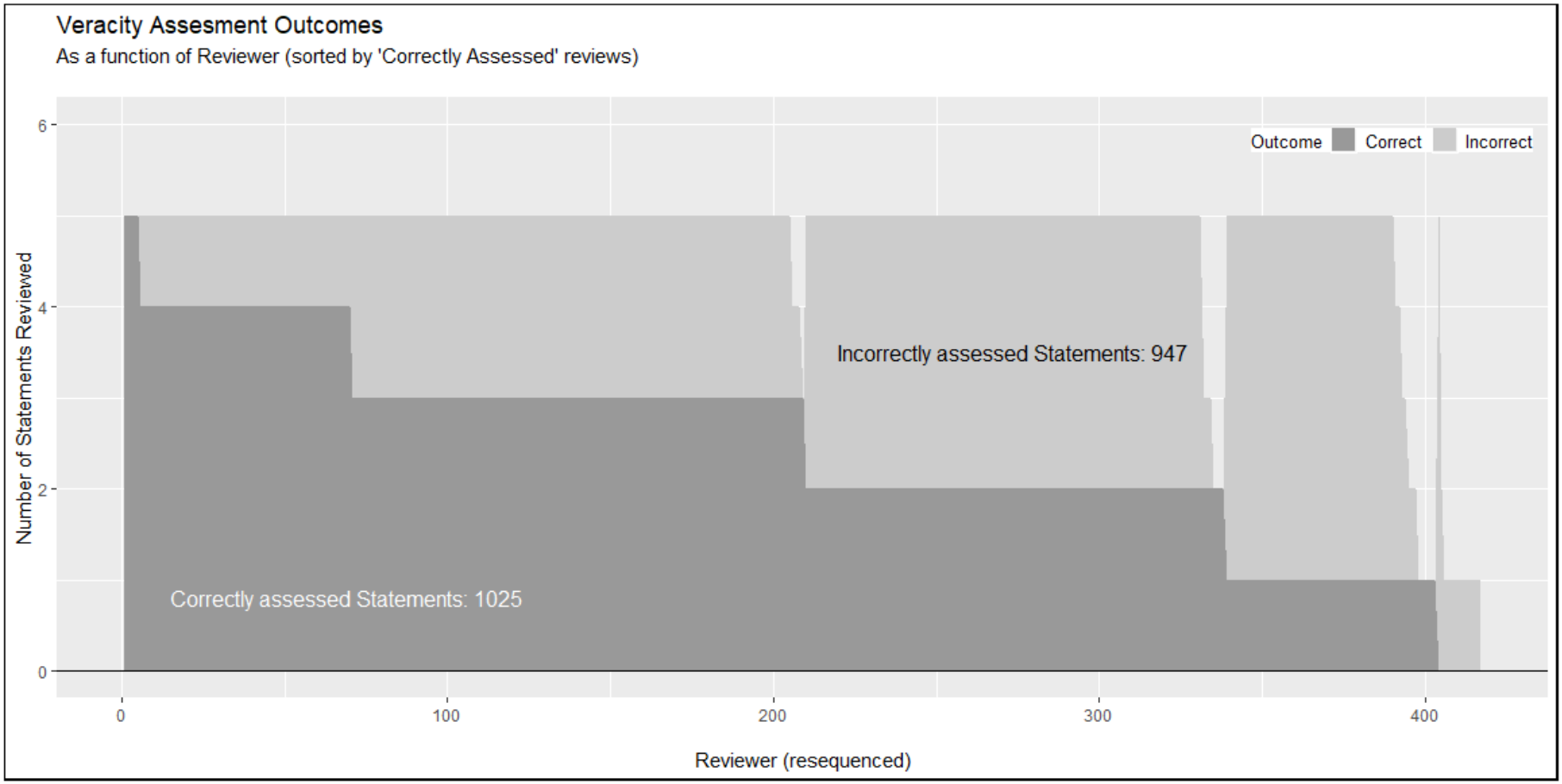

Figure 5 Outcome assessments depicted as a function of individual reviewers

Generalised Linear Mixed Model

Generalised linear mixed models (GLMM) were constructed to extrapolate the language related empirical findings of this study to the general population of the Western Cape.

Model definition. Four fixed effects and three possible random effects were identified. The four fixed (main) effects were all dichotomous categorical variables, and were (a) statement veracity (truth or lie), (b) the statement language being a first or second language of the statement provider, (c) whether the statement language was a first or second language of the statement veracity judge, and (d) whether the statement was recorded under cognitive load or not. The three random effects considered were considering the impact on the outcome of veracity accuracy when viewed against grouping by (a) the 39 individual statement providers, (b) the 12 different topics of statements, or (c) the 417 different veracity judges.

The language focussed research questions' GLMM did not consider the cognitive load main effect. Both the language focussed GLMM, as well as, the GLMM considering the cognitive load aspect only considered Statement Provider as a random effect in the end. This decision was made following an analysis of the effects that each of the three potential random effects under consideration could have on the intercept of the generalised linear model as discussed hereafter.

Figures 6 to 8, below each depict the effect that each of the three random effect considerations have on the calculated intercept of a generalised linear mixed model. The GLMM models were computed using the fixed effects statement veracity, statement language and reviewer language and then considering each of the three potential random effects individually. The plots were generated using the R package *merTools* (Knowles & Frederick, 2019).

Considering the Statement Provider as a random effect has the biggest impact (as depicted by effect range, (figure 6) on the calculated intercepts of the different GLMM models.

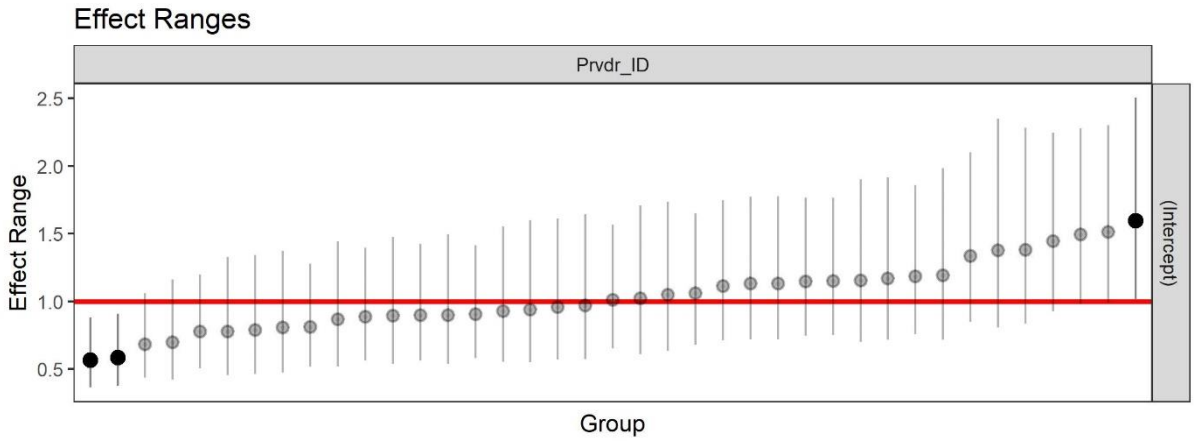

Figure 6 Effect on intercept of Random Effect: Statement Provider (n = 39)

Considering the individual topic as a random effect also indicates some impact on the model outcome (intercept) albeit smaller than considering the Statement provider, as is evidenced by comparing the vertical axis scales of figures 6 and 7.

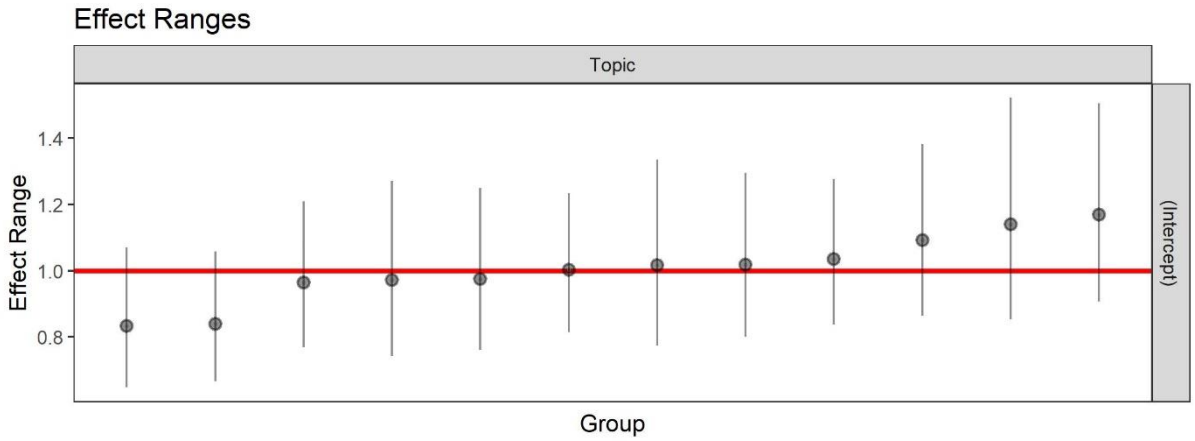

Figure 7 Effect on intercept of Random Effect: Unique Topic (n = 12)

Considering the Statement reviewer groups as random effects (figure 8, below) have very little effect on the GLMM model outcome as evidenced by the extended vertical scale required to depict variation from the no effect line.

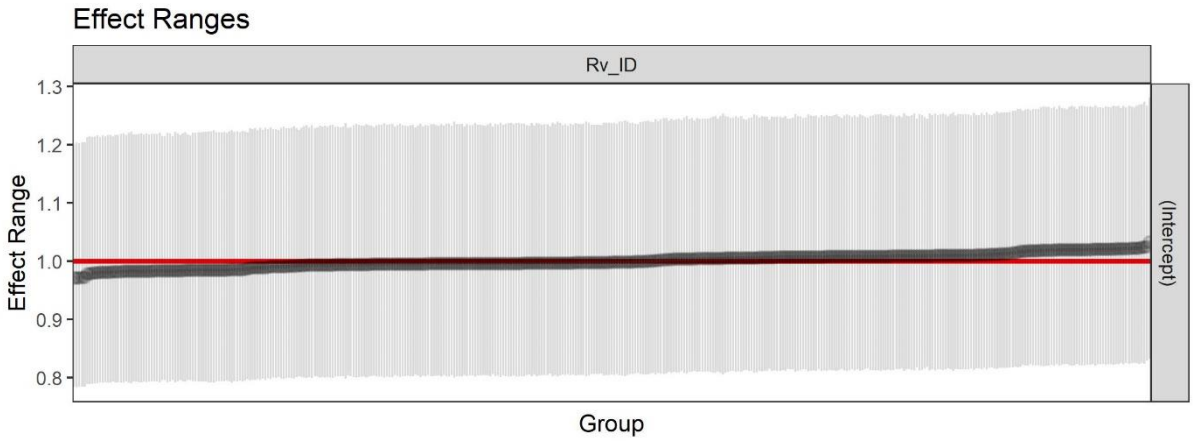

Figure 8 Effect on intercept of Random Effect: Statement Reviewer (n = 417)

Hierarchical analysis of GLMM models in which the model with Statement Provider as a random effect was compared against models with either Topic or Reviewer ID as random effects were conducted. The results presented in tables 6 and 7 below. The differences between the models using reviewers or topics as random effects compared against the statement providers were not significant. After this analysis only the statement provider was considered as a random effect in GLMM models.

Table 6
AIC values and hierarchical comparisons to assess "Topic" Random effects models

	AIC	BIC	deviance	Anova against Model 00	
				χ^2(df)	*P*
Model 00	2,291.9	2,342.2	2,273.9		
Model 05	2,306.5	2,356.8	2,288.5	0(1)	1-

Notes:

*** p < .001; ** p < .01; * p < .05

Model 00: Result ~ St. Veracity * St. L1L2 * Rv. L1L2 + (1|St.Provider)

Model 05: Result ~ St. Veracity * St. L1L2 * Rv. L1L2 + (1|Topic)

StL1 /StL2 = First/Second language of statement provider;

RvL1 / RvL2 = First/Second Language Review, i.e. statement language is 1st / 2nd language to reviewer)

Table 7
AIC values and hierarchical comparisons to assess "Reviewer" Random effects models

	AIC	BIC	deviance	Anova against Model 00	
				χ^2(df)	*P*
Model 00	2,291.9	2,342.2	2,273.9		
Model 06	2,307.5	2,357.8	2,289.5	0(1)	1-

Notes:

*** p < .001; ** p < .01; * p < .05

Model 00: Result ~ St. Veracity * St. L1L2 * Rv. L1L2 + (1|St.Provider)

Model 06: Result ~ St. Veracity * St. L1L2 * Rv. L1L2 + (1|St_Reviewer)

StL1 /StL2 = First/Second language of statement provider;

RvL1 / RvL2 = First/Second Language Review, i.e. statement language is 1st / 2nd language to reviewer)

The initial model, specifying interactions between all three main effects were contrasted with a model where the statement veracity attribute was treated as an additive effect without any interaction with the two-language attribute main effects. The comparison of the two model outputs (Table 8) shows that the initial model (specified with full interaction between variables) is preferred as a summary of the model indicates the lowest AIC score (2,291.9 vs. 2,328.7). The difference between the two models is reported as significant $\chi^2(3) = 42.85$, $p < 0.001$ and hence the full interaction model was the preferred model for this research.

Table 8
AIC values and ANOVA comparisons for different fixed effects specification model

	AIC	BIC	deviance	Anova against Model 00	
				χ^2(df)	*P*
Model 00	2,291.9	2,342.2	2,273.9	42.85(3)	< 0.001 ***
Model 04	2,328.7	2,362.2	2,316.7	-	-

Notes:

*** p < .001; ** p < .01; * p < .05

Model 00: Result ~ St. Veracity * St. L1L2 * Rv. L1L2 + (1|St.Provider)

Model 04: Result ~ St. Veracity + St. L1L2 * Rv. L1L2 + (1|St.Provider)

StL1 /StL2 = First/Second language of statement provider;

RvL1 / RvL2 = First/Second Language Review, i.e. statement language is 1st / 2nd language to reviewer)

The coefficients and significance of effects of the GLMM model, *Result ~ St. Veracity * St. L1L2 * Rv. L1L2 + (1|St.Provider),* are tabulated in table 9, below. Eight different combinations (interactions) of the three dichotomous main effect variables are possible.

Table 9
Mixed Linear Regression Coefficient Table with Statement Veracity, Statement First/Second Language and Reviewer's First/Second Language as Fixed Effects with Statement Provider as Random Effect

Variable	Coefficient	SE	*z*	*p*	95% CI
(Intercept)	0.93	0.15	6.35	< 0.001 ***	[0.65, 1.23]
Statement Language (2^{nd} | 1^{st})	0.27	0.19	1.38	0.168	[-0.11, 0.65]
Reviewer language (2^{nd} | 1^{st})	0.11	0.23	0.50	0.681	[-0.33, 0.56]
Statement Veracity (Lie | Truth)	-1.99	0.19	-10.72	< 0.001 ***	[-2.36, -1.63]
Statement language (2^{nd} | 1^{st}) x Reviewer language (2^{nd} | 1^{st})	-1.31	0.34	-3.89	< 0.001 ***	[-1.99, -0.66]
Statement language (2^{nd} | 1^{st}) x Statement Veracity (Lie | Truth)	-0.43	0.27	-1.63	0.104	[-0.95, 0.09]
Reviewer language (2^{nd} | 1^{st}) x Statement Veracity (Lie | Truth)	-0.50	0.32	-1.57	0.117	[-1.13, 0.12]
Statement Language (2^{nd} | 1^{st}) x Reviewer language (2^{nd} | 1^{st}) x Statement Veracity (Lie | Truth)	2.24	0.43	5.19	< 0.001 ***	[1.40, 3.09]

Notes: *** p < .001; ** p < .01; * p < .05

The model's intercept is at 0.93 (SE = 0.15, CI [0.65, 1.23]). Three significant effects are indicated within the GLMM model: the main effect of statement veracity (β = -1.99, SE = 0.19, CI [-2.36, -1.63], z = -10.72, p < .001), the two-way interaction between second language statements reviewed in a second language (β = -1.31, SE = 0.34, CI [-1.99, -0.66], z = -3.89, p < .001) as well as the three way interaction between all three fixed effects (β = 2.24, SE = 0.43, CI [1.40, 3.09], z = 5.19, p < .001). The remaining variables and interactions in the model were not indicated as significant.

Model validation. Hartig (2019) asserts that the use of 'standard' residual plot algorithms for GLMM models may indicate all manner of problems including non-normality, heteroscedasticity even in instances where a model is correctly specified. Hartig developed the R

module *DHARMa* (Hartig, 2019) to enable residual diagnostics of hierarchical multi-level/mixed regression models. The diagnostic output is generated by simulating models using a GLMM model definition to create interpretable scaled residuals (Hartig, 2019). Once outputs have been simulated, the DHARMa package provides utilities to plot and test for model misspecification that could result in problems such under/over dispersion, etc.

One thousand simulations using the selected GLMM model, being *Result ~ Statement Veracity * Statement First/Second Language * Reviewer First/Second language + (1| Participant_ID)* were computed. To assess whether problems are present in the model specification the user compares plots of the simulated interquartile ranges with the expected interquartile ranges. In the case where the model specification is valid, the QQ-plot of the (simulated) residuals will plot close to the "observed" value = "expected" value on a scatterplot.

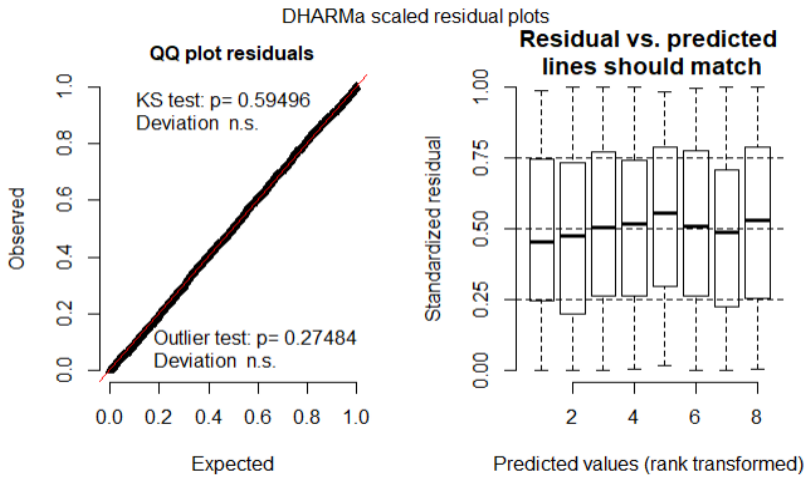

Figure 9 DHARMa scaled residuals plot

In the right-hand graph of Figure 9, the interquartile ranges of the three fixed effects and the interactions between the fixed effects are depicted as boxplots. The observation that the 0.25 to 0.75 interquartile ranges plots close to the ranges corresponding to the standardized residual values, combined with a QQ plot that shows almost no deviation of the simulated residuals from the ideal line is interpreted as indicating an appropriate model specification.

The Effect of Statement and Review language interaction on veracity judgment.

A tendency amongst veracity judges to evaluate the veracity of statements as truthful, regardless of the original statement veracity being truth or lie has been indicated in the discussion of figure 2, a figure depicting the outcomes of veracity judgment when viewed across the four main effects individually, on page 73. The GLMM model discussed in the previous section showed a significant interaction between second language statements reviewed in a second language (β = -1.31, SE = 0.34, CI [-1.99, -0.66], z = -3.89, $p < .001$) on the modelled veracity outcome. Figure 6, on page 87 depicts the veracity assessment outcomes as functions of the interaction between first and second language reviews and first and second language statements.

The percentage of correct veracity judgements for first language statements adjudicated in a first language ("first language reviews") was 51.2% CI [47.2, 55.3] and for first language statements adjudicated in a second language ("second language reviews") 54.0% CI [48.7, 59.2].

The differences in the accuracy of veracity judgement outcomes for first language statements reviewed in a first language compared to first language statements reviewed in a second language is only 2.8%. This difference is not reported as significant $\chi^2(1) = 0.58$, $p > 0.05$.

However, the percentage of correct veracity judgements for second language statements adjudicated in a first language ("first language reviews") was 49.1% CI [45.2, 53.1] and for second language statements adjudicated in a second language ("second language reviews") 56.2% CI [50.9%, 61.3%]. The differences in the accuracy of veracity judgement outcomes for second language statements reviewed in a first language compared to second language statements reviewed in a second language is 7.1% and is reported as significant, $\chi^2(1) = 4.31$, $p < 0.05$.

An overall significance test between first and second language reviews, collapsed across first and second language statements (i.e. not considering whether the statement was provided by a first or second language statement provider) is also reported as significant $\chi^2(1) = 4.26$, $p < 0.05$.

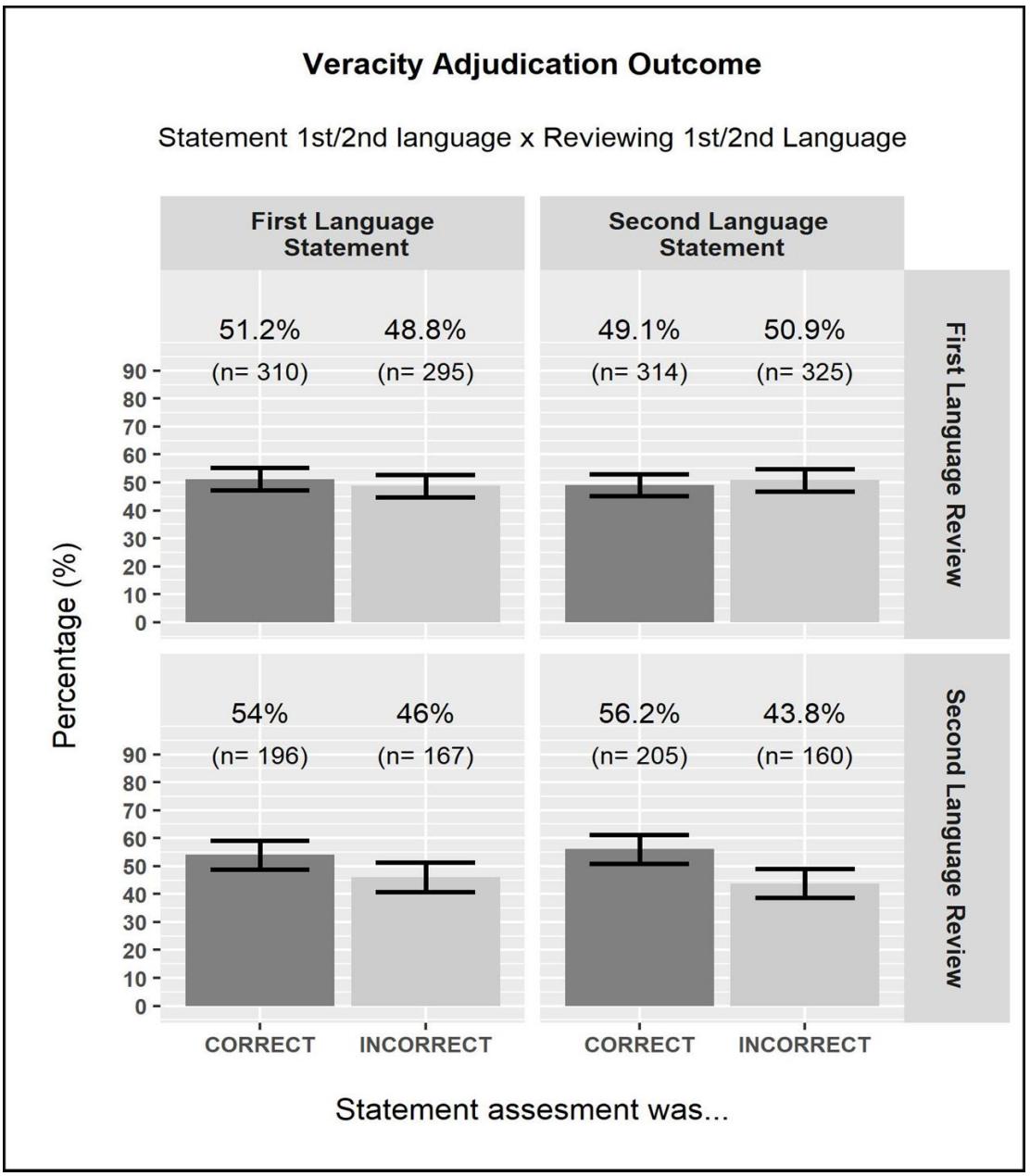

Figure 10 Veracity assessment outcome ratios against language attributes

An overall significance test between first and second language statements, collapsed across first and second language reviews (i.e. not considering whether the statement was

reviewed by a first or second language reviewer) does not report as significant; $\chi^2(1) = 0.05$ $p = 0.832$ (>0.10).

Thus, we can conclude that in this study, the consideration whether the reviewer is reviewing in the reviewer's native language or not has a bigger impact on the accuracy of the veracity assessment outcome than whether the language was provided by a first or second language statement provider.

Truth Default and the Veracity effect

The analysis is expanded to consider the impact of the original statement veracity on the veracity judgement outcome together with the four different combinations of statement and reviewer language. The outcomes of correct and incorrect adjudications for the eight possible combinations of the three dichotomous fixed effect variables are depicted in figure 11 on the following page.

The introduction of statement veracity into the grouping of language attributes to assess veracity adjudication outcomes yields non overlapping 95% confidence intervals in seven of the eight possible combinations of the three dichotomous fixed effect variables depicted. The only combination of the three dichotomous fixed effect variables that indicate overlapping confidence intervals is the second language review of second language statements combined with an untruthful (lie) statement. The statements in this group was assessed correctly 46.2%, CI [38.2, 54.3] and assessed incorrectly 53.8%, CI [45.7, 61.8].

Truth Default. All eight possible combinations (interactions) of the three dichotomous variables yielded more 'true' veracity judgements than 'untrue', i.e. more truthful statements were correctly judged 'true' than incorrectly as 'lies', and more 'untrue' statements were

incorrectly judged as 'true' than correctly judged as lies. This is in line with the original observations made in figure 2, page 73.

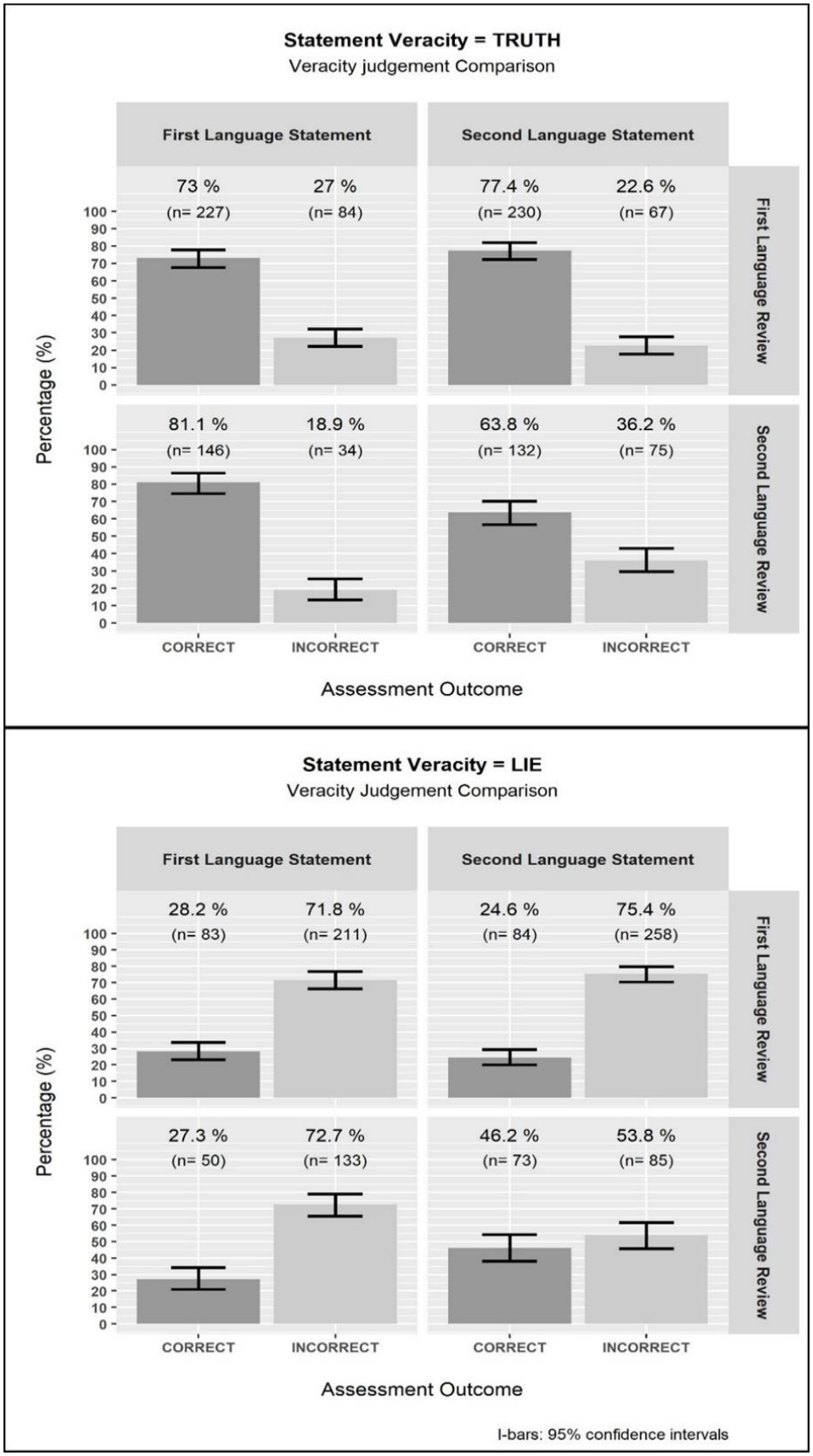

Figure 11 Veracity assessment outcomes for interacting fixed effect variables

The veracity adjudication of second language statements combined with second language reviews, for both 'truth' and 'untruth' statement veracities, indicate the smallest difference between correct and incorrect assessments. (7.6% for statement veracities a lie, 27.6 % for statement veracities being truths). This is a marked difference compared to other language combinations indicated in table 10, below.

Table 10
Proportional Differences between Truth/Lie Outcomes for Fixed Effects Interactions

Statement Veracity	Statement language	Review Language	Δ Correct, Incorrect	ΔRvL2-ΔRvl1	χ^2	*df*	*p*
Untrue	St_L1	Rv_L1	-43.6 %	-1.8%	0.0122	1	0.912
		Rv_L2	-45.4 %				
	St_L2	Rv_L1	-50.8 %	43.2%	22.503	1	< 0.001 ***
		Rv_L2	-7.6 %				
True	St_L1	Rv_L1	46.0 %	14.2%	3.6855	1	0.054
		Rv_L2	62.2 %				
	St_L2	Rv_L1	54.8 %	-27.2%	10.603	1	0.0011 **
		Rv_L2	27.6 %				

Notes: *** p < .001; ** p < .01; * p < .05
St 1st/St 2nd: First or second language statements;
Rv 1st/Rv 2nd statements adjudicated in a first or second language;
Δ Correct, Incorrect: % difference between correct and incorrect veracity judgements
ΔRvL2- ΔRvL1 Difference in outcome difference (Reviewer Language)

Only the adjudication of second language statements yielded significant differences between correct and incorrect veracity judgments (43.2% and -27,2%). Second language reviews of second language statements do report higher number of assessments being judged as lies, evidenced by the reduction of the differences between correctly and incorrectly assessed statements. Thus, whilst the interaction of second language reviews of second language statements do yield proportionally higher judgements of 'lie' than 'truth' compared to other combinations, a greater number of statements are still, albeit marginally in this category assessed as truths for this combination of attributes.

Veracity Effect. The Veracity effect holds that truthful statements are more likely to be assessed correctly as truths than deceptive statements are likely to be correctly assessed as lies (Levine et al., 1999). When considering the overall dataset, there is a significant correlation between the veracity judgment outcome and the original statement veracity of $\chi^2(1) = 383.83$; $p < 0.001$. This correlation between veracity judgment outcome and original statement veracity is depicted below and is an extract of the original descriptive plot of figure 2 on page 73. Figure 12, below, indicates that 73.9% [71.0, 76.1] of the truthful statements were correctly assessed as truths, whilst 70.3% [67.3, 73.2] of the untruthful statements were incorrectly assessed as truths.

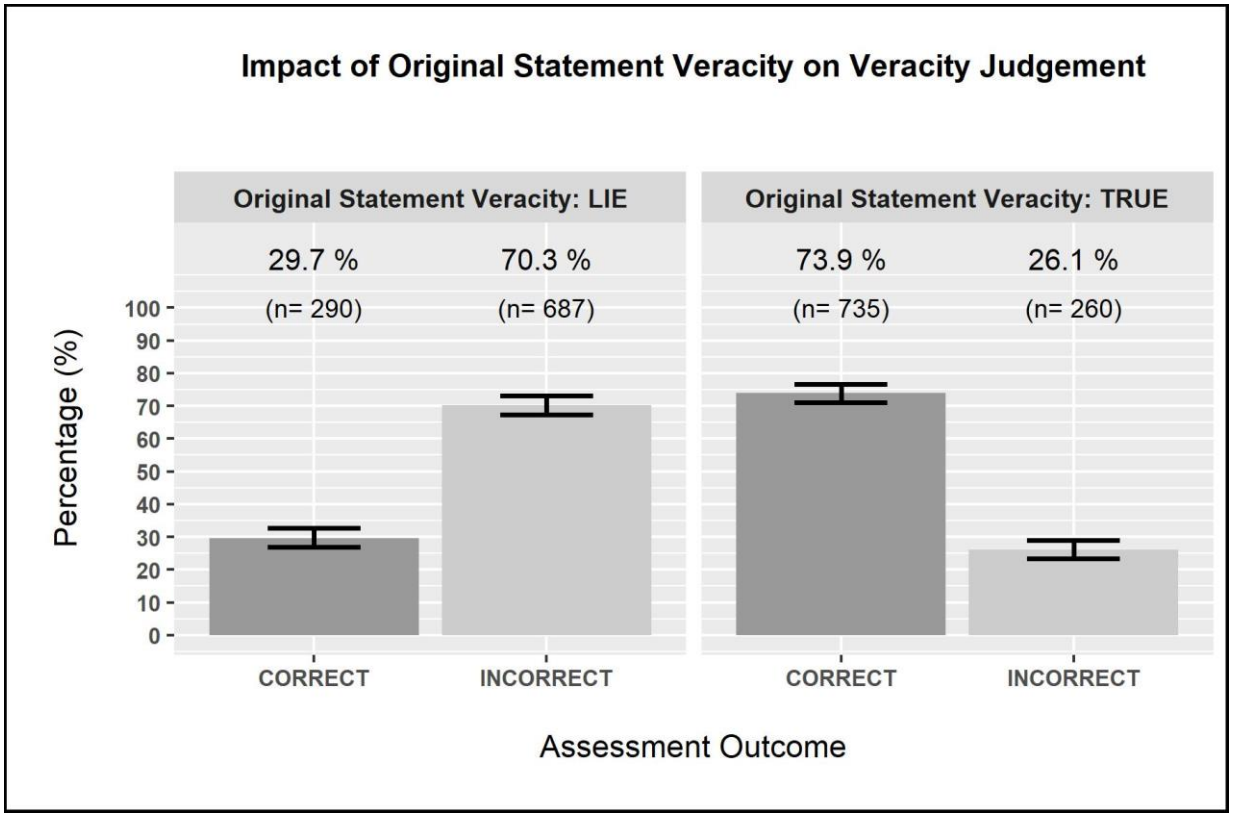

Figure 12 Impact of Original Statement Veracity

This supports the "Veracity Effect" framework of deceptive communication that holds that truthful statements are more likely to be assessed correctly as truths than deceptive statements are likely to be correctly assessed as lies.

Bias

Significance testing of correlation between reviewer's veracity assessments against the 2 x 2 language interactions for statements grouped overall by truth and lie statement veracities was done to assess the presence of bias. For the 'truth' statements, a chi-square test of the veracity judgment outcome against the four language combinations indicated $\chi^2(3) = 17.92$, $p < 0.001$ whilst the chi-square of 'lie' statements' veracity judgment outcome against the four language combinations indicated $\chi^2(3) = 25.741$; $p < 0.001$. Both correlations are reported as significant.

To better understand the contribution from each of the possible language interactions to the significant correlations, an analysis of the contributions of the Pearson residuals to the chi-square test was performed. The different interactions' contributions to the overall chi-square calculation are expressed as percentage contributions for the truth and lie statement groups and are depicted in Table 11 (truthful statements) and Table 12 (untruthful statements).

Table 11
Proportional contribution of Pearson residuals to χ2 significance for truthful statements

	StL1 x RvL1	StL1 x RvL2	StL2 x RvL1	StL2 x RvL2
Reviewer Judgement: LIE	**0.5%**	**8.1%**	**20.2%**	**45.1%**
Reviewer Judgement: TRUE	**0.2%**	**2.9%**	**7.1%**	**16.0%**

Notes:
StL1 /StL2 = First/Second language of statement provider
RvL1 / RvL2 = First/Second Language Review (Statement language is first / second language to reviewer)

Table 12
Proportional contribution of Pearson residuals to χ2 significance for untruthful statements

	StL1 x RvL1	StL1 x RvL2	StL2 x RvL1	StL2 x RvL2
Reviewer Judgement: LIE	**0.8%**	**1.3%**	**11.7%**	**56.4%**
Reviewer Judgement: TRUE	**0.3%**	**0.6%**	**5.0%**	**23.8%**

Notes:
StL1 /StL2 = First/Second language of statement provider
RvL1 / RvL2 = First/Second Language Review (Statement language is first / second language to reviewer)

Inspection of tables 11 and 12 indicate that the biggest contributions to the significant chi-square correlations reported, for both instances of truth and lie grouped statements are from the combination of second language statements and second language reviews. Not much of the contribution to the outcome is attributed to first language contributions.

Second language Statements and lie bias. To indicate a lie bias is present when statement judges review second language statements, it needs to be shown that veracity judges are more likely to judges second language statements as being untruthful regardless of the original statement veracity. The results of the study do not support the original (overall) hypotheses that a lie bias is extended to second language statements as evidenced by Figure 11 on page 88. Truthful second language statements, regardless whether reviewed in a first or second language is overall correctly assessed as truths and untrue second languages statements is overall incorrectly assessed as truths.

However, in the case of second language statements (the statement language is a second language of the statement provider), reviewed by a second language veracity judge (the statement language is a second language to the veracity judge) the analysis of the contributions of the language combinations to the significance of veracity assessment outcomes, as presented in tables 11 and 12 above, do support a lie bias in this one category. The proportional contribution of second language statements reviewed by second language veracity judges for

originally untruthful statements are indicated 32.8% higher for correct judgment of the statement as a lie, and 29.1% for incorrectly assessed truthful statements as lies.

So, whilst an overall lie bias towards second language statements is not demonstrated, in the specific case of second language statements reviewed by a second language judge a lie bias can be inferred.

First Language Statements and Truth Bias It was already established that in this study the veracity judging of statements, regardless of language attributes is more likely to be assessed as truths than lies. Figures 2, 10 and 11 presented earlier supports this conclusion. A truth bias implies that lie statements will likely be incorrectly assessed as truths, whilst truth statements will preferentially be correctly assessed as truths.

Overall, this study does indicate a preferential judging of statements as truths. However, inspection of tables 11 and 12 indicates that grouping first language statements together to assess contributions to the overall outcome significance testing are very small, specifically for the correlation between statements assessed as truths (0.3%, 0.6%, 0. 2% and 2.9%). From this observation it is concluded that the outcomes of the study are preferentially a support of the veracity effect rather than evidence of a truth bias towards first language statement reviews.

Cognitive Load and Veracity Judgement

In order to determine whether cognitive load affected the accuracy of veracity judgments, the GLMM model used was updated to include the dichotomous variable indicating the presence or absence of cognitive load.

Table 13, below summarises the hierarchical analysis between the GLMM model used to assess research question one (Model RQ1) and a model where the fixed effect variable indicating the presence of cognitive load is also considered in addition to the fixed effect variables considered in the first research question (Model RQ5). The comparison does not indicate

significant difference in the goodness of fit indicators as represented by the AIC scores: 2,291 (no cognitive load considered) and 2,302 (considering the presence of cognitive load).

Table 13
AIC values and hierarchical comparisons to assess presence of cognitive load

	AIC	BIC	deviance	Anova against no Cognitive Load	
				χ^2(df)	p
Model RQ5	2,302.4	2,397.4	2.268.4	5.45 (8)	0.707
Model RQ1	2,291.9	2,342.2	2,273.9	-	-

Notes:

Model RQ1: Result ~ St. Veracity * St. L1L2 * Rv. L1L2 + (1|St.Provider)

Model RQ5: Result ~ St. Veracity + St. L1L2 * Rv. L1L2 * St_CogLoad + (1|St.Provider)

StL1 /StL2 = Statement in first/second language of statement provider;

RvL1 / RvL2 = Statement is reviewed in a first/second language of adjudicator;

St_CogLoad = Indicator whether cognitive load was present or not when statement was provided

The summary of the updated GLMM model's coefficients follows in Table 12 on the next page. Previously, as tabulated in Table 9, three significant effects were indicated within the original GLMM model used to assess research question one: the effect of statement veracity (β = -1.99, SE = 0.19, CI [-2.36, -1.63], z = -10.72, p < .001), the two-way interaction between second language statements reviewed in a second language (β = -1.31, SE = 0.34, CI [-1.99, -0.66], z = -3.89, p < .001) as well as the three way interaction between all three fixed effects (β = 2.24, SE = 0.43, CI [1.40, 3.09], z = 5.19, p < .001). In the updated GLMM model, when considering the presence of cognitive load as well, four significant effects are indicated: the effect of statement veracity (β = -2.20, SE = 0.27, CI [-2.74, -1.68], z = -8.15, p < .001), the two-way interaction between second language statements reviewed in a second language (β = -0.93, SE = 0.48, CI [-1.88, 0.00], z = -1.93, p < .05), the three way interaction between the three fixed effects without cognitive load (β = 1.69, SE = 0.60, CI [0/51, 2.88], z = 2.81, p < .01); as well as the interaction between statement first/second language and cognitive load (β = 0.78, SE = 0.39, CI [0.01, 1.54], z = 2.00, p < .05).

Table 14
Mixed Linear Regression Coefficient Table with Statement Veracity, Statement First/Second Language, Reviewer's First/Second Language and presence of Cognitive Load as Fixed Effects with Statement Provider as Random Effect

Variable	Coefficient	SE	*z*	*p*	95% CI
(Intercept)	1.13	0.21	5.30	< 0.001 ***	[0.72, 1.56]
Statement Language (2^{nd} | 1^{st})	-0.14	0.28	-0.51	0.613	[-0.70, 0.41]
Reviewer language (2^{nd} | 1^{st})	0.02	0.33	0.07	0.941	[-0.62, 0.69]
Statement Veracity (Lie | Truth)	-2.20	0.27	-8.15	< 0.001 ***	[-2.74, -1.68]
Cognitive Load (Yes| No)	-0.38	0.30	-1.28	0.200	[-0.96, 0.20]
Statement language (2^{nd} | 1^{st}) x Reviewer language (2^{nd} | 1^{st})	-0.93	0.48	-1.93	0.053	[-1.88, 0.00]
Statement language (2^{nd} | 1^{st}) x Statement Veracity (Lie | Truth)	-0.02	0.38	-0.06	0.955	[-0.76, 0.72]
Reviewer language (2^{nd} | 1^{st}) x Statement Veracity (Lie | Truth)	-0.23	0.44	-0.53	0.598	[-1.12, 0.63]
Statement language (2^{nd} | 1^{st}) x Cognitive Load (Yes| No)	0.78	0.39	2.00	0.046 *	[0.01, 1.54]
Reviewer language (2^{nd} | 1^{st}) x Cognitive Load (Yes| No)	0.17	0.46	0.38	0.073	[-0.72, 1.07]
Statement Veracity (Lie| Truth) x Cognitive Load (Yes| No)	0.40	0.37	1.09	0.275	[-0.32, 1.13]
Statement Language (2^{nd} | 1^{st}) x Reviewer language (2^{nd} | 1^{st}) x Statement Veracity (Lie | Truth)	1.69	0.60	2.81	0.005 **	[0.51, 2.88]
Statement Language (2^{nd} | 1^{st}) x Reviewer language (2^{nd} | 1^{st}) x Cognitive Load (Yes| No]	-0.75	0.66	-1.14	0.252	[-2.03, 0.54]
Statement Language (2^{nd} | 1^{st}) x Statement Veracity (Lie | Truth) x Cognitive Load (Yes| No]	-0.80	0.53	-1.50	0.132	[-1.84, 0.24]
Reviewer Language (2^{nd} | 1^{st}) x Statement Veracity (Lie | Truth) x Cognitive Load (Yes| No]	-0.55	0.64	-0.86	0.391	[-1.81, 0.70]
Statement Language (2^{nd} | 1^{st}) x Reviewer language (2^{nd} | 1^{st}) x Statement Veracity (Lie | Truth) x Cognitive Load (Yes| No]	1.10	0.86	1.28	0.206	[-0.59, 2.80]

Note: *** p < .001; ** p < .01; * p < .05

Table 14, above, shows that whilst the same effects were reported as the more significant compared to the other effects in the updated GLMM model, overall the significances were weaker once the cognitive load variable is introduced in the model. The interaction between statement language and the presence of cognitive load was additionally indicated as weakly significant, $p < 0.05$. Figure 12 depicts the extent to which the presence of cognitive load impacts the veracity adjudication of statements

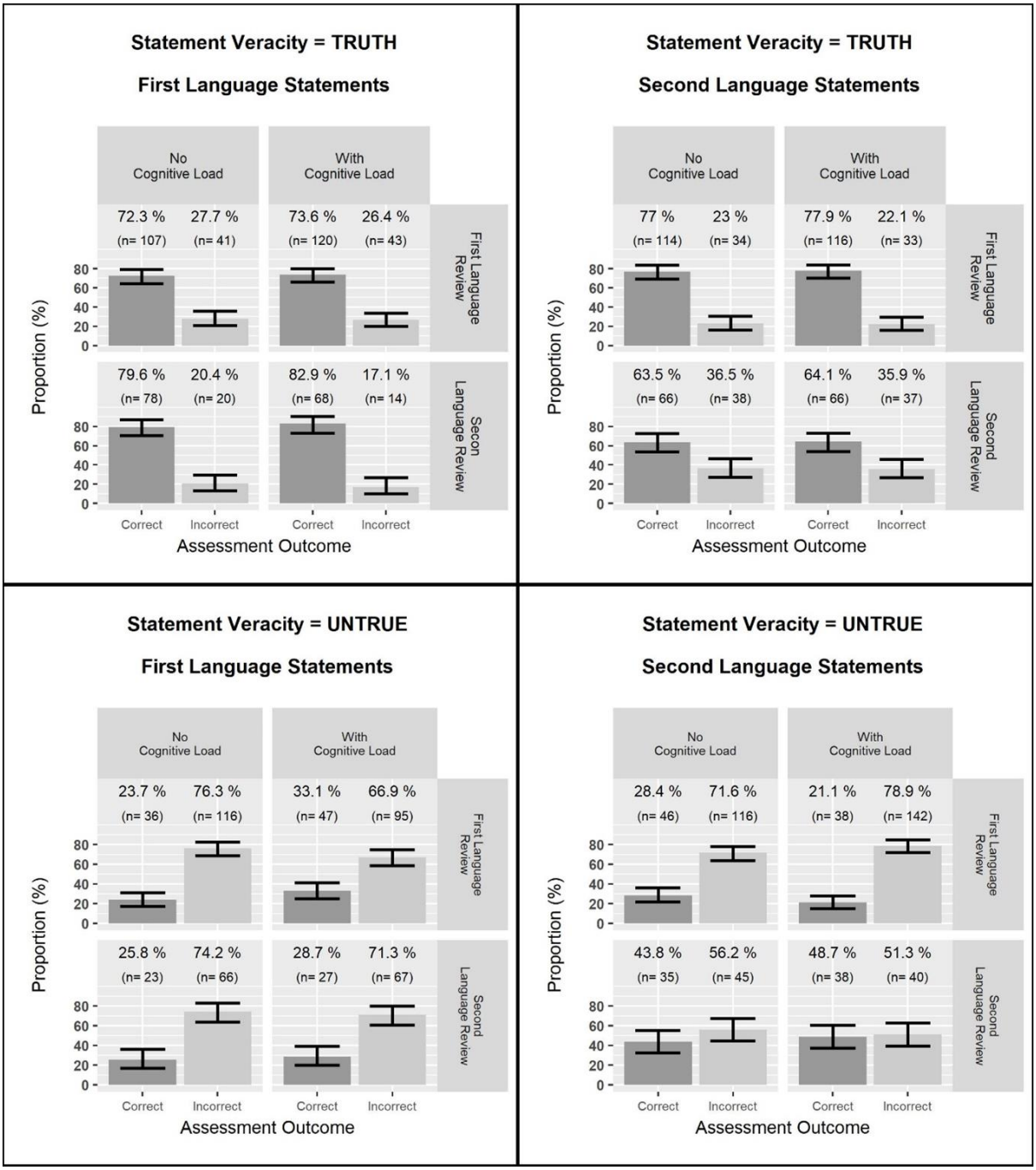

Figure 13 The Effect of Cognitive Load on the veracity adjudication of statements

Table 15
Impact on veracity judgement outcome with and without the presence of cognitive load

Statement Veracity	Statement language	Review Language	Cognitive Load	Δ Correct, Incorrect	ΔCL (Δ)	χ^2	*df*	*p*
Truth	St 1st	Rv 1st	Yes	47.2%	2.6%	0.02	1	0.893
			No	44.6%				
		Rv 2nd	Yes	65.8%	6.6%	0.14	1	0.750
			No	59.2%				
	St 2nd	Rv 1st	Yes	55.8%	27,6%	0.00	1	0.975
			No	28.2%				
		Rv 2nd	Yes	28.2%	1.2%	0.00	1	1.000
			No	27.0%				
Untruth	St 1st	Rv 1st	Yes	9.4%	43.2%	2.76	1	0.096
			No	-33.8%				
		Rv 2nd	Yes	-42.6%	5.8%	0.73	1	0.786
			No	-48.4%				
	St 2nd	Rv 1st	Yes	-57.8%	14.6%	2.06	1	0.150
			No	-43.2%				
		Rv 2nd	Yes	-2.6%	9.8%	0.22	1	0.640
			No	-12.4%				

Notes: *** p < .001; ** p < .01; * p < .05
St 1st/St 2nd: First or second language statements;
Rv 1st/Rv 2nd statements adjudicated in a first or second language;
CLY: Statement provided under cognitive load;
CLN: No secondary task (cognitive load) during statement provision.
Δ Correct, Incorrect: % difference between correct and incorrect veracity judgements
ΔCL (Δ): Difference in difference of outcomes due to cognitive load

Figure 13, and Table 15, above summarises the differences observed between the statement veracity adjudication and the presence or absence of cognitive load. Introducing cognitive load when true statements are provided does not have a big impact on the veracity adjudication of statements. The difference (expressed as percentage) between truth statements with cognitive load present against the condition with no cognitive load present ranges from 0.6% to 3.3%. However, in the case of untrue statements (attempts to deceive) the correct outcomes range from -7.3% to 9.4%. The presence of cognitive load does improve the number of

correct judgements of veracity of statements for first language reviews of first language statements (9.4%) and second language review of second language statements (5.4%). The presence of cognitive load does influence the percentage of correct veracity adjudication but results in opposing movements in outcomes for first and second language statements. None of the differences between correct and incorrect adjudications are reported as significant, hence it is concluded that in this study, the presence of cognitive load during statement generation did not have any significant impact on the veracity adjudication outcome.

NLP (LIWC2015) vs Human Evaluators

One hundred and two (n = 102) English statements, 51 true (n = 51) and 51 untrue (n = 51) were processed through LIWC2015 and the resultant linguistic profiles compared against published deceptive linguistic frameworks. The comparisons were made to assess whether NLP assessments of the transcribed English statements generated in this study were more accurate than with the human veracity judges.

Word Count One of the observations made in the existing literature is that a word count difference (statement length difference) is often observed between truthful and deceptive statements. There were no significant differences in the word count of statements comparing truthful and untruthful statements in this study. The word count descriptive statistics for the English statements are presented in Table 16, below.

Table 16
Word Count Comparison of truthful vs untruthful English statements

Veracity	*n*	mean	SD	χ^2	*Df*	*p*	Cohen's d	
							d	CI
Truth	51	359.94	98.20	83	84	>0.05	0.28	[-0.11, 0.68]
Untruth	51	357.53	88.86					

Note:
n: the number of statements for each veracity category

As participants were required to provide statements limited to a three-minute duration whether telling truth or untruth, the design of this study potentially excluded using this as an indicative criterion assisting with categorising statements as truth or lie as requiring a three-minute statement in each case 'forces' statements to be similar in length.

Summary Variables and Word Categories Appendix 4 contains several figures showing the complete summary of all LIWC2015 summary variables and word categories provided for in LIWC2015 except for punctuation categories. Punctuation categories were excluded from analysis as the analyses were done on transcribed statements and not originally written statements where the use of punctuation from individual statement providers could be beneficial.

Published Linguistic Profiles associated with Deception Comparisons of the linguistic profile differences between truthful (n = 51) and untruthful (n = 51) statements using a Natural Language Processing (NLP) deceptive statement linguistic profile is presented in Table 17 on page 100 and depicted in Figure 14 on page 101. A similar comparison between truthful (n = 51) and untruthful (n = 51) statements using a Reality Monitoring deceptive statement linguistic profile is presented in Table 19 on page 102 and graphically depicted in Figure 15, page 103.

For both linguistic frameworks (Reality Monitoring and Natural Language Processing) generalised linear models (GLM) were defined and the significance of the logistics regression coefficients assessed to determine whether the study's observations can be extrapolated successfully to that of a general population. The Natural Language Processing framework, which is based off of Pennebaker et al., (2003) understanding of what a deceptive statement profile will consist of coefficient table is reported in table 18, page 101 and the Reality Monitoring processing framework's GLM in table 20, page 103.

.

Table 17
Word category comparisons within a Natural Processing framework

NLP framework	Example words	Untruthful		Truthful		Significance of difference			Cohen's d	
		M	SD	M	SD	χ^2	*df*	*p*	d	CI
First person singular	*I, me, my*	1.85	1.37	2.19	1.57	90.67	81	>0.05	0.23	[-0.16, 0.62]
Third person	*he, she, them*	2.78	2.13	3.17	2.59	98.00	89	>0.05	0.16	[-0.22, 0.56]
Negative emotions	*hate, enemy*	1.63	1.38	1.39	1.30	84.50	82	>0.05	-0.18	[-0.57, 0.21]
Negate	*no. not. never*	2.65	1.63	2.51	1.22	88.00	87	>0.05	-0.10	[-0.50, 0.29]
Differ	*hasn't, but, else*	4.71	1.79	4.60	1.59	88.67	88	>0.05	-0.06	[-0.46, 0.33]
Motion words	*go, walk, run*	1.95	1.22	1.73	0.94	84.67	84	>0.05	-0.21	[-0.60, 0.19]

Notes
The original NP framework used a word category "Exclusive words" that is no longer in the 2015 dictionary. Replaced with "Negations" and "differentiations"
The third person word category was computed from the sum of the third person singular and third person plural categories of the LIWC 2015 dictionary

Table 18
Linear Regression Coefficient Table for with NLP word categories as Fixed Effects

Word category variables	Coefficient	SE	*z*	*p*	95% CI
(Intercept)	0.406	0.84	0.47	> .1	[-1.26, 2.09]
I	0.175	0.14	1.21	> .1	[-0.11, 0.47]
Third person	0.133	0.09	1.44	> .1	[-0.05, 0.32]
Negate	-0.161	0.19	-0.83	> .1	[-0.55, 0.22]
negemo	-0.192	0.16	-1.18	> .1	[-0.53, 0.12]
Differ	0.007	0.14	0.05	> .1	[0.28,0.29]
motion	-0.260	0.20	-1.32	> .1	[0.66, 0.12]

*Notes: *** p < .001; ** p < .01; * p < .05*

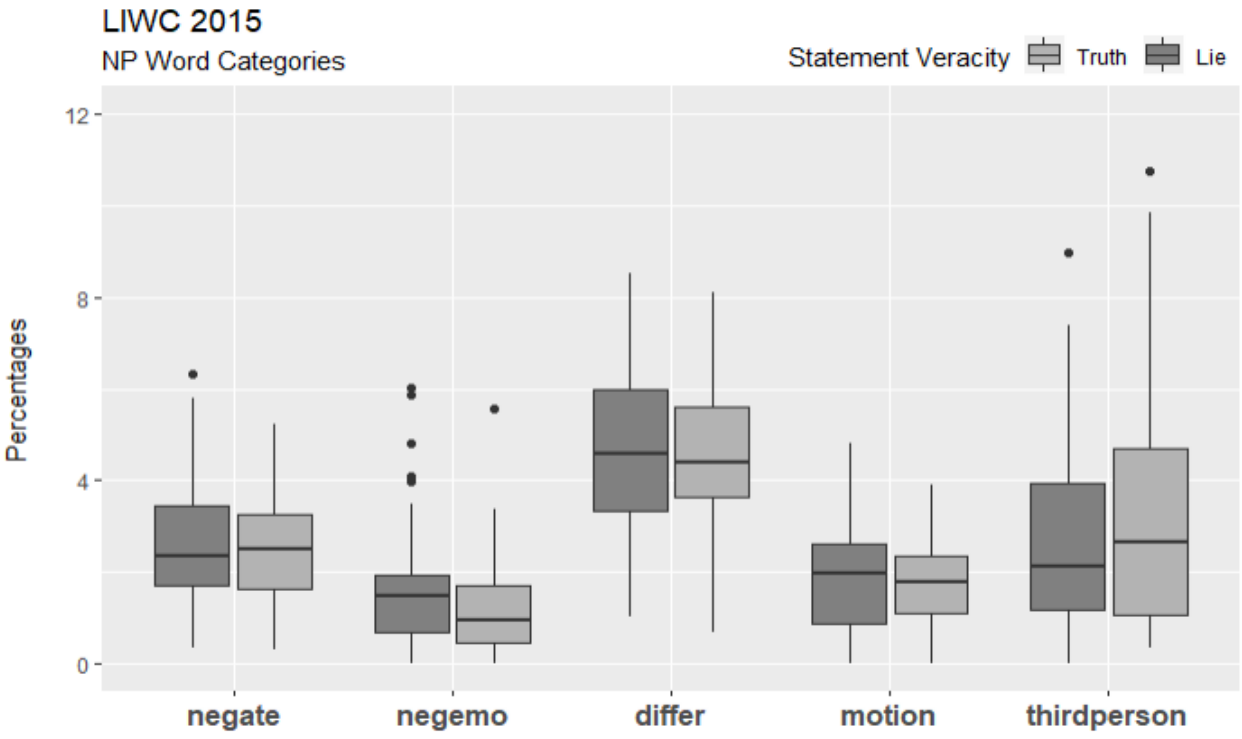

Figure 14. Linguistic Profile within a Natural Language Processing Framework

The Natural Language Processing framework logistic regression model using this study's observations of the variables associated with the NLP framework was not significant and was not considered any further.

Table 19
Word Category comparisons within a Reality Monitoring framework

Reality Monitoring framework Word Categories	Example words	Untruthful		Truthful		Significance of difference			Cohen's d	
		M	SD	M	SD	χ^2	*df*	*p*	d	CI
Perceptual Processes	*see, feel, hear*	1.50	1.18	1.50	0.94	82.42	77	>0.05	0.0059	[-0.3869, 0.3987]
Relativity (spatial)	*go, down, over*	9.83	3.30	11.36	2.79	100.00	94	>0.05	0.4988	[0.0999, 0.8978]
Time Orientations (Temporal)	*today, is, now*	17.46	3.56	17.79	2.56	97.33	91	>0.05	0.1070	[-0.2862, 0.5001]
Affective	*happy, love, hate*	3.99	1.82	3.96	1.61	86.00	86	>0.05	-0.2160	[0.4144, 0.3713]
Cognitive Processes	*think, perhaps*	16.49	3.32	16.47	1.01	91.33	92	>0.05	-0.0068	[-0.3997, 0.3861]

Note

The Time Orientation word category was computed from the sum of the 'past focus', 'present focus' and 'future focus' word categories of the LIWC2015 dictionary.

Table 20
Linear Regression Coefficient Table for Reality Monitoring word categories as fixed effects.

Word category variables	Coefficient	SE	*z*	*p*	CI
(Intercept)	-5.13	2.531	-2.027	< .05	[-10.337, -0.331]
affect	0.154	0.139	1.104	> .1	[-0.117, 0.435]
Cognitive Processing	0.073	0.075	0.972	> .1	[-0.072, 0.225]
Perception)	-0.039	0.197	-1.098	> .1	[-0.433, 0.348]
Time Orientation	0.059	0.069	0.847	> .1	[-0.076, 0.199]
Relative	0.219	0.081	2.689	< .01	[0.066, 0.389]

*Notes: *** p < .001; ** p < .01; * p < .05*

The Reality Monitoring framework for deceptive language detection's model (GLM) using the observations of this study's intercept is -5.13 (SE = 2.53, CI [-10.34, -0.33]) Within the model, only the effect of the spatial word category ("relativ") is significant (β = 0.22, SE = 0.08, CI [0.07, 0.39], z = 2.69, p < .01). The effect can be considered as "small"

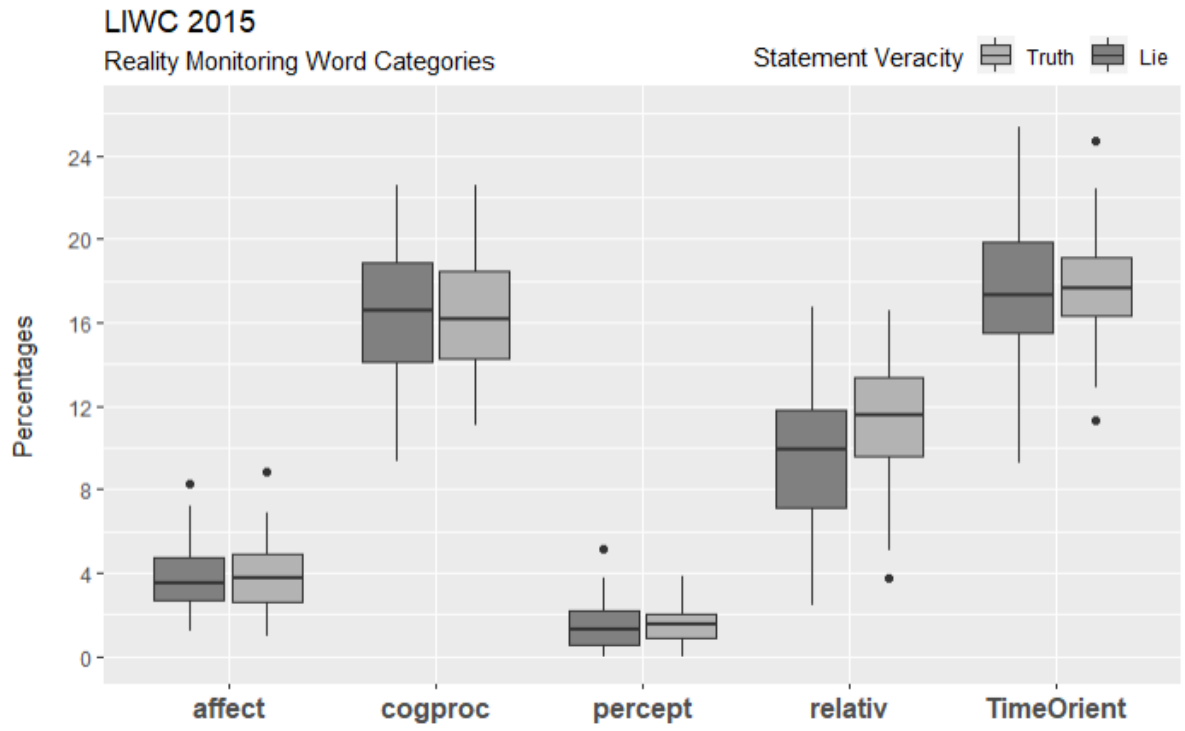

Figure 15 Linguistic Profile of deceptive communication within a Reality Monitoring Framework

In order to assess the effectiveness of the model to correctly assess the veracity judgement outcome of the statements analysed by LIWC2015 using the Reality monitoring

framework, models were fitted using 71 randomly selected statements (70% of 102 statements) and the fitted model used every time to predict the outcome of the remaining 31 statements (30% of 102 statements) . This was repeated 1,000 times, in a simulation run where the selection of the training of model fitting statements randomly varied between each run. The outcome of the 1,000 simulations conducted is depicted in Figure 16.

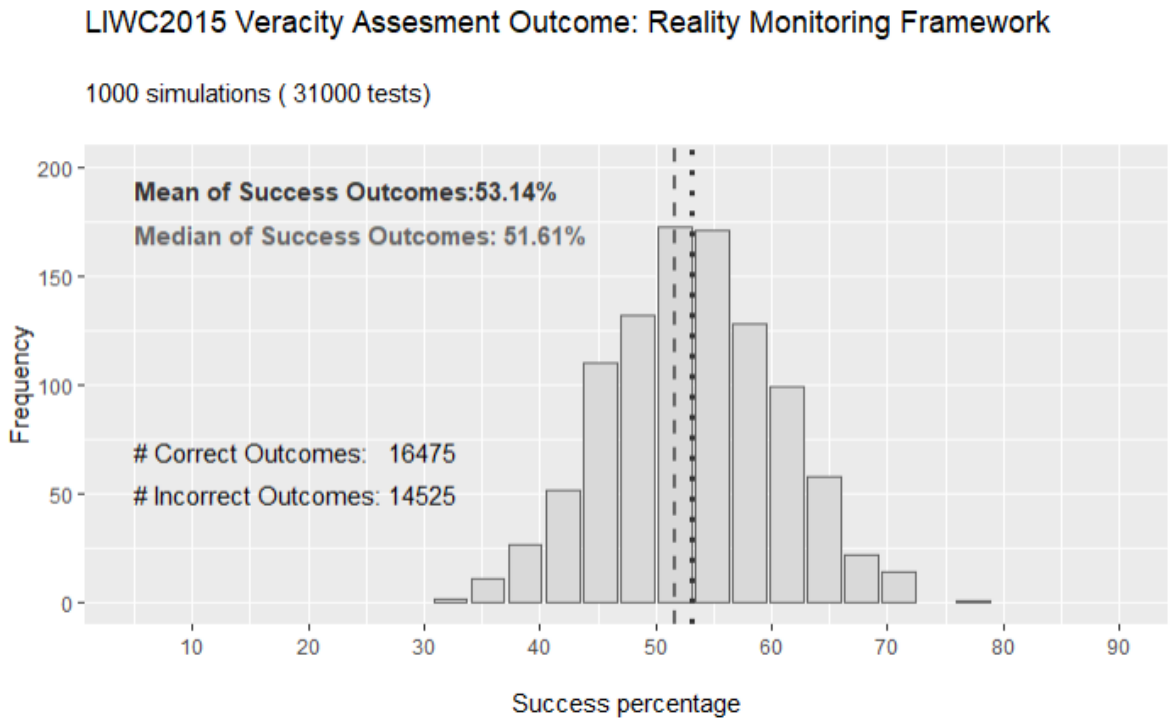

Figure 16 Reality Monitoring deceptive language framework simulation

Post-hoc Analysis

Each statement in the phase 2 questionnaires asked, in addition to opining on whether the statement under review is truthful or not, a number of questions pertaining to the perception of characteristics or attributes the reviewer has formed of the statement provider and the quality of the statement. The descriptive results of these ancillary questions were:

Language Proficiency. Figure 17 depicts the perceptions of the statement judges of the language proficiency of the speaker after reviewing each statement, expressed as a view of whether the statement language was a native or non-native language of the statement provider.

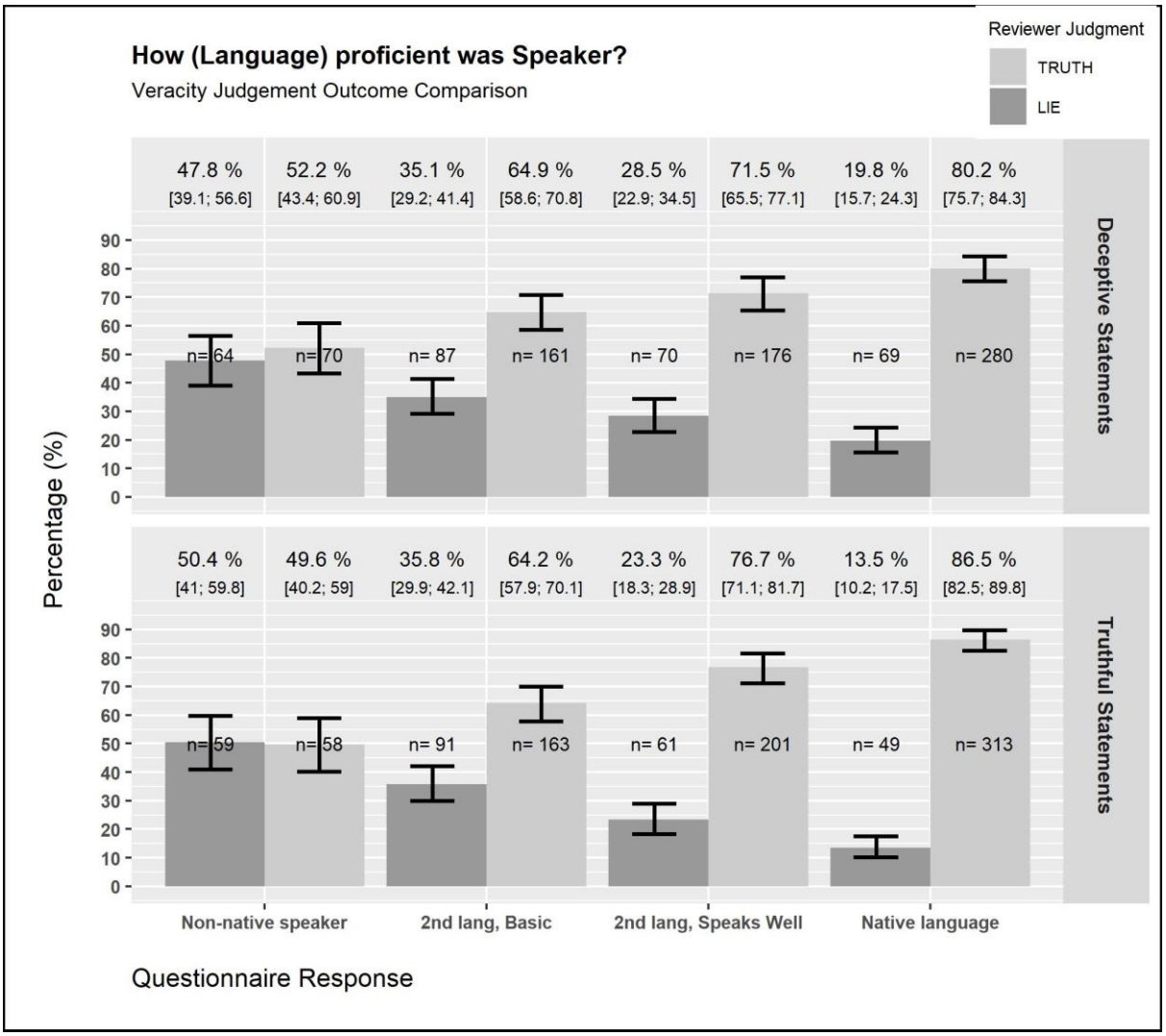

Figure 17 Phase 2 Participant Responses: Language Proficiency of Speaker

For those speakers, judged to be non-native speakers, the assessment of whether the statement was a truth, or a lie was no different to that of a chance outcome. In the case of deceptive statements ('lies') 52.2% [43.4%; 60.9%] of the statements were incorrectly assessed as truths against 47.8% [39.1%; 56.6%], a difference of only 4.4% ($\Delta = 4.4\%$). In the case of truthful statements, the difference between correct and incorrect veracity judgment outcomes is negligible ($\Delta=0.8\%$). For both truthful and deceptive statements, the confidence intervals

associated with the outcome of the veracity assessment outcomes overlap substantially. In keeping with the observations in figure 2, page 73, for the other three categories of language proficiency perception, statements are preferentially assessed as truths regardless of whether the original statement was truthful or deceptive. Figure 15, page 105 shows that the less proficient the statement judge perceived the speaker to be, the more chance-like the accuracy of the outcomes are, and the more proficient the judge perceived the speaker to be the more likely the judge would assess the veracity of the statement as truthful, regardless of the original statement veracity.

Statement Clarity. Figure 18, page 107 depicts the experience of the reviewers in judging each statement, expressed as a view of how easy it was to follow the argument(s) put forward by the speaker.

A similar trend than observed in figure 15, presents in figure 16 as well. The easier the statement was to follow, the more likely the judges evaluated the statements as truthful, regardless of the original statement veracity. For the statements assessed as not easy to follow, in the case of originally deceptive statements there is no differences between correct and incorrect assessments ($\Delta=0$). There is, however, some difference indicated between the correct/incorrect assessments of originally truthful statements for statements difficult to follow ($\Delta=10.2\%$)

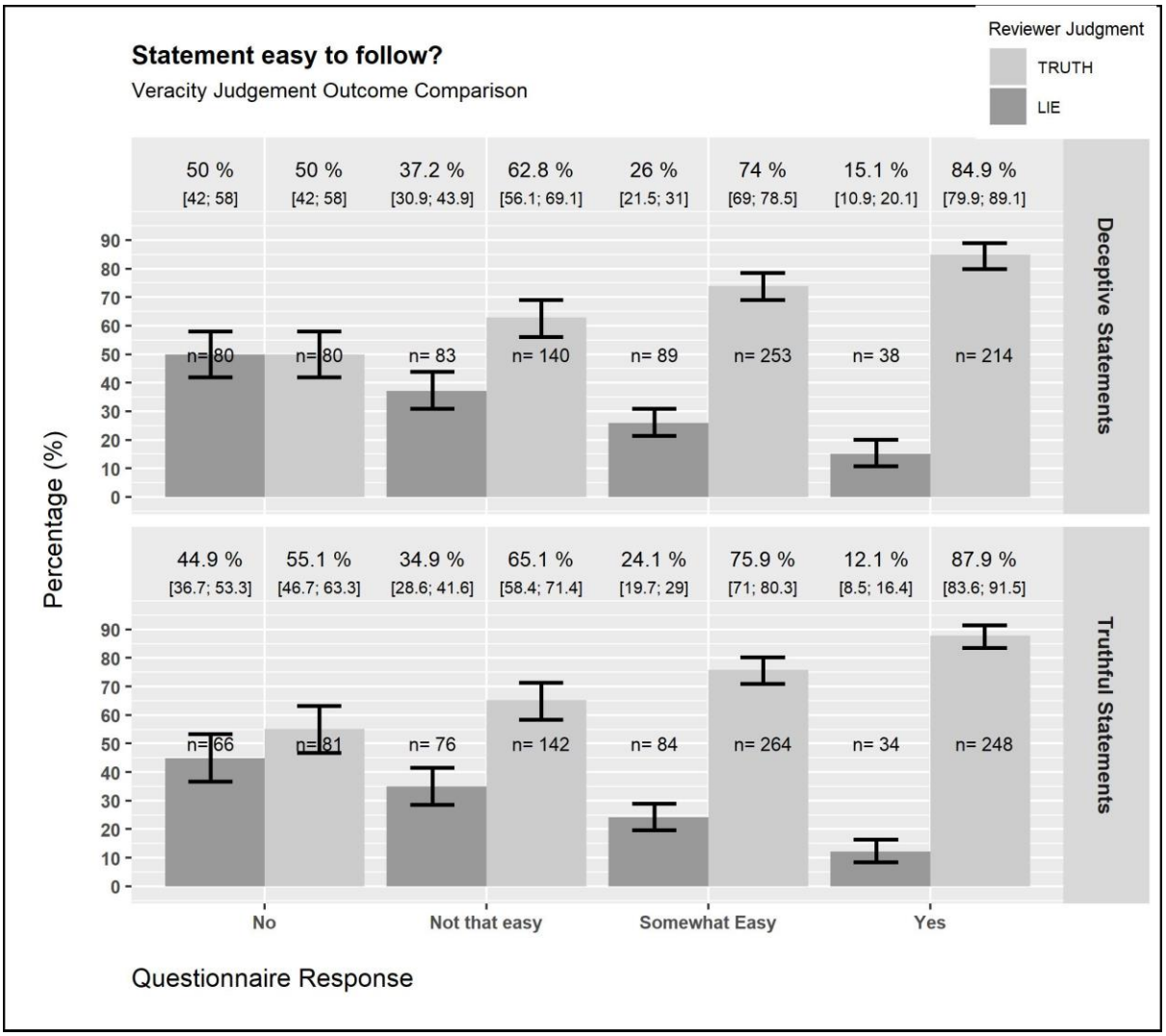

Figure 18 Phase 2 Participant Responses: How easy was the statement to follow?

Statement Repetition. Figure 19 page 108 depicts the opinions of the statement judges in judging the amount of repetition in each statement.

It shows judges perceived most of the statements as being repetitive in varying degrees. Only 138 of the total 1,972 statements assessed were characterised as having no repetition ("None"). Six hundred and forty-one (n = 641) statements were assessed as highly repetitive ("lots"), seven hundred and eighty-one (n = 781) as statements with some repetition and four hundred and twenty-two (n = 422) with little repetition. Evaluating the percentage difference between correct and incorrect veracity judgements in figure 17 for the different perceptions on repetition indicates a similar observation in figures 14 and 15, i.e. the 'better' a statement is received (as expressed through the amount of repetition in the view of the judge in this instance)

the more likely the veracity would be judged as truthful regardless of the original statement veracity. This observation of preferentially judging statements to be truthful also holds for statements assessed as being very repetitive ("lots of repetition")

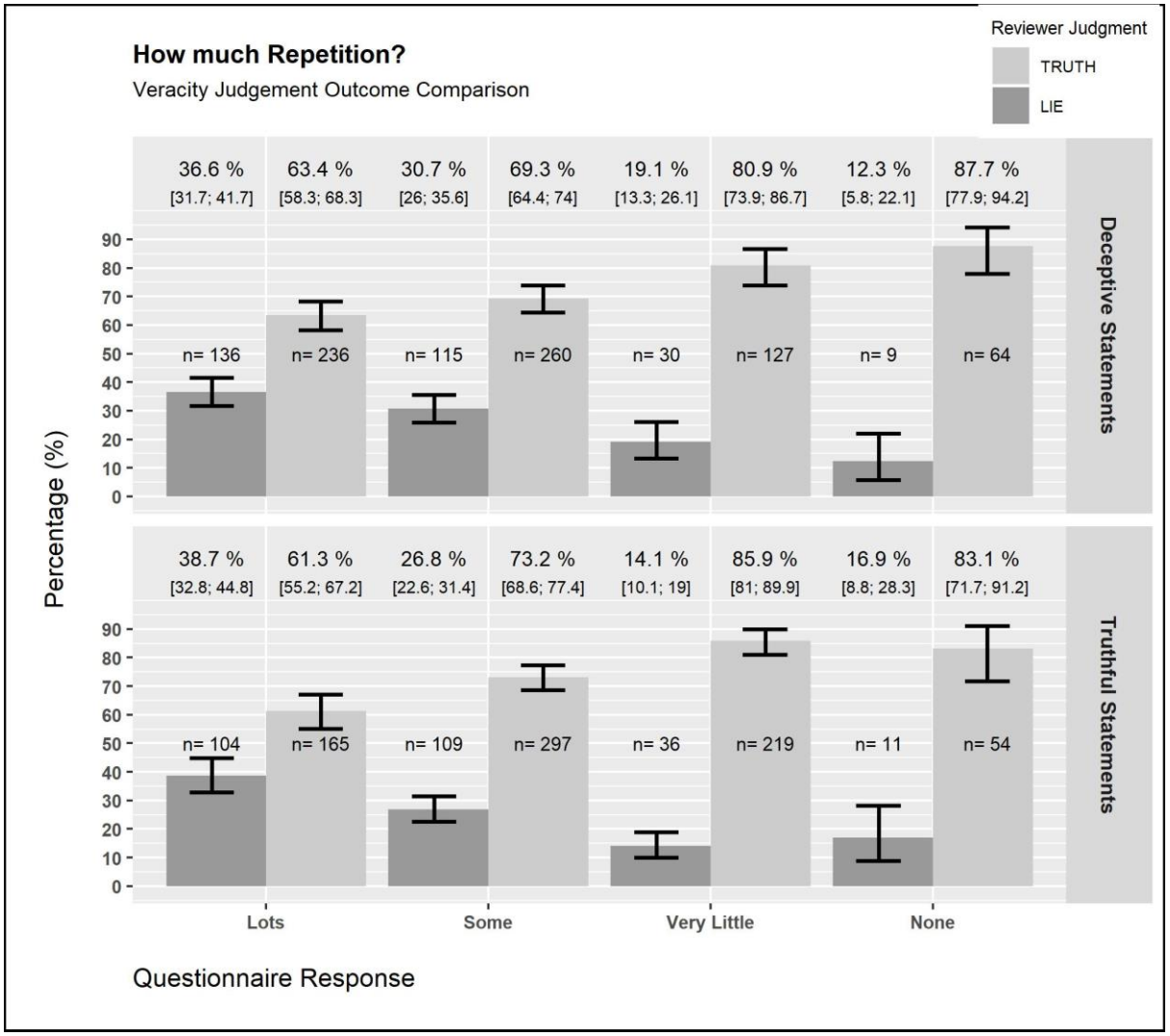

Figure 19 Phase 2 Participant Responses: How much repetition occurred in Statement?

How informed was the speaker? Figure 20, page 109 depicts the opinions formed by the veracity judges on how informed the speakers were about a topic.

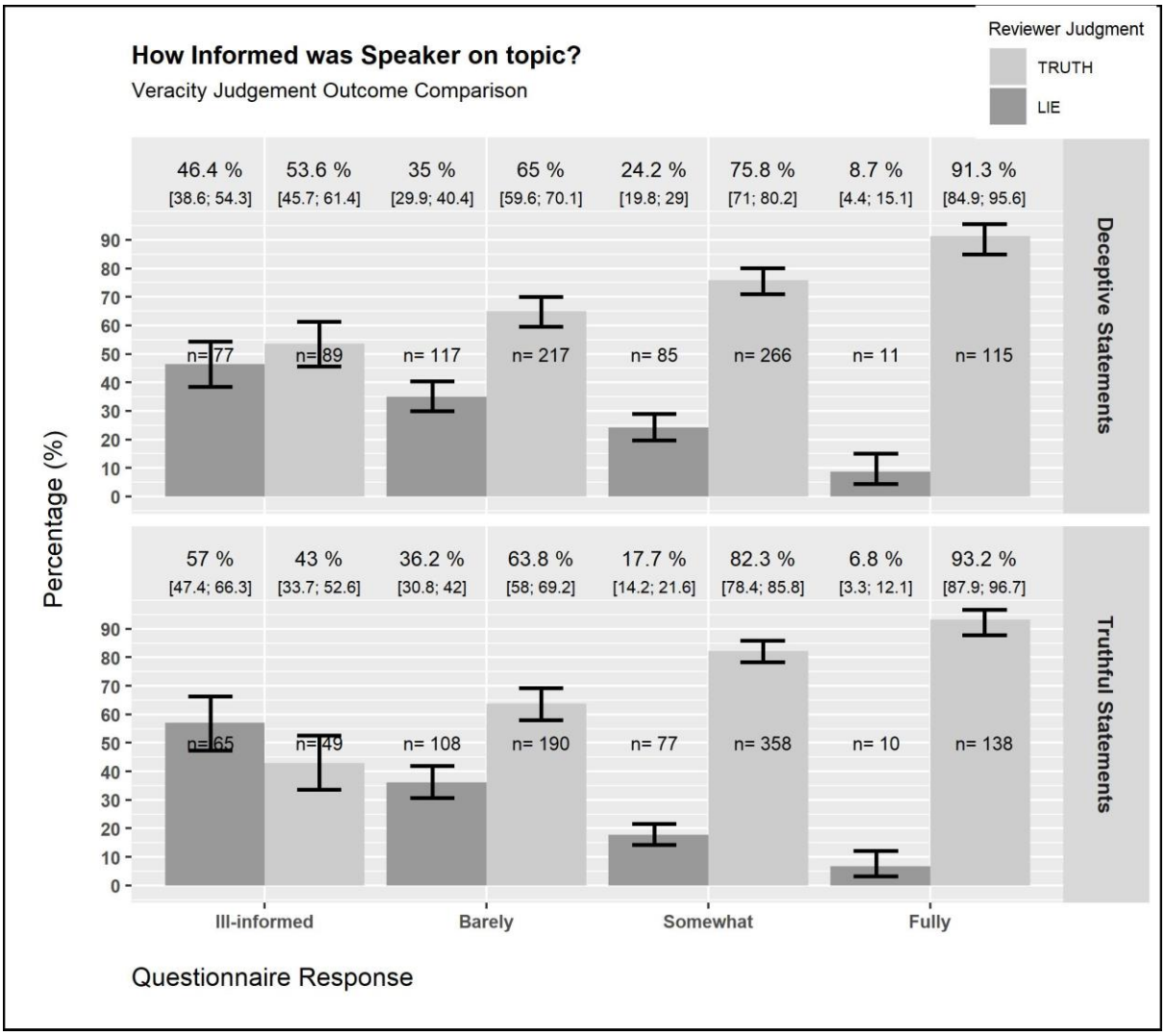

Figure 20 Phase 2 Participant Responses: How Informed was the speaker on the topic?

In instances where the veracity judge assessed the speaker as ill-informed, the accuracy of the veracity assessments suffered. In the case of truthful statements 57% [47.4; 66.3] of statements were incorrectly assessed as lies and in the case of deceptive statements 53.6% [45.7; 61.4] was incorrectly assessed as truths. This is the first instance where a category of statements were preferentially judged as untruths, Nevertheless, the overall trend observed in the earlier judges' response graphs are also present in this view with a preferentially increase in judging statements to be true regardless of original statement veracity as judges perceived speakers more

positively, in this case to be better informed (with the exception as indicated of views of ill-informed responses of originally truthful statements).

Agreement with the speaker? Figure 21, page 110 depicts the extent to which a veracity judge agreed with the opinion of the speaker providing a statement.

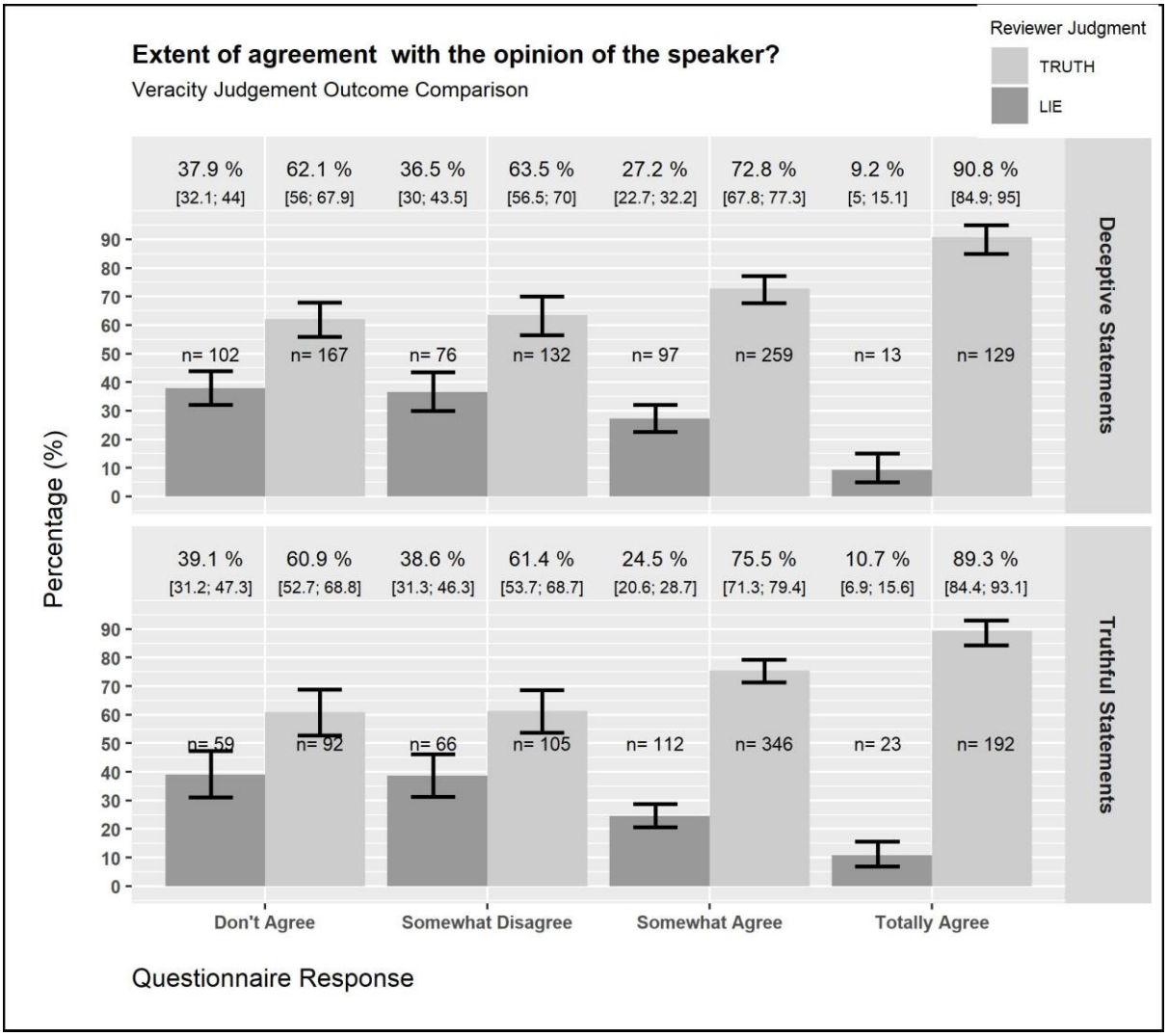

Figure 21 Phase 2 Participant Responses: To what extent do you agree with the speaker?

There is little differentiation in the preference to assess the veracity as truthful regardless of original statement veracity across the 4 categories when expressed as a percentage between whether the original statement veracity was truthful or deceptive, e.g. 90.8% [84.9, 95.0] in the case of deceptive statements assessed incorrectly as truths, vs. 89.3% [84.4; 93.1] in the case of

truthful statements correctly assessed as truths (Δ% = 1.5%). The biggest difference is observed for the category 'Somewhat agree' with (Δ% = 2.7%.)

Chapter 5 Discussion

This chapter discusses the results of this study within the research frameworks and literature reported on in preceding chapters. It starts with a brief summary of the overall findings and then goes into a more in-depth discussion of the study's findings.

An overall lie bias towards statements provided in a second language of the statement provider could not empirically be demonstrated in this study. However, a further subset of the second language statements into whether the assessment was done by a first or second language statement veracity judge hinted at a lie bias. Results of the study were consistent with extant literature in terms of an overall truth bias, but the results were more marginal when restricting this assessment to the first language statements as required by the second research hypothesis. The assessment of the overall dataset supports the premise behind the Veracity Effect framework that holds that truthful statements are more likely to be assessed correctly as truths than deceptive statements are likely to be correctly assessed as lies as suggested by Levine and colleagues (1999).

In contrast to what was expected the influence of cognitive load was not significant when assessed on the overall data set. However, when considering the smaller subset of the data representing cognitive load placed on second language statement providers an observable effect is noticeable. Finally, this study found that the software programme LIWC2015 marginally outperformed human evaluators in making veracity assessments.

This study found, in keeping with studies reported in the literature, that the veracity assessment outcomes overall was still close to a chance level outcome (Blandón-Gitlin et al., 2014; Bond & DePaulo, 2006; DePaulo et al., 2003; van Vuuren, 2017).

This study presented the veracity judges with transcribed of audio-recorded statements. The statement judges had no information about the statement provider(s), and particularly

relevant to this study no direct knowledge of whether the statements were provided by native-language (first-language) speaker or non-native language (second language) speakers.

Most deception research allows for reviewers to use various modalities to make veracity assessments. Reviewers often make use of an audio-visual modality or an audio modality. In these instances, veracity judges have access to more cues potentially such as subjective verbal and non-verbal cues to aid in their veracity assessments. It has been found that at times these inferred cues are not based on actual cues associated with deception which then misleads the veracity judge and resulting in incorrect conclusions to the veracity of a statement (Carlucci et al., 2013; Vrij et al., 1999). This thesis restricted the communication between the statement provider and the statement veracity judge to the written modality only. This was done to restrict the ability of participants resorting to their own perceived cues to deception. Despite this manipulation, veracity assessment outcomes were still comparable to that of a chance outcome as is typically found in deception research (Bond, & DePaulo, 2006; DePaulo et al., 2003).

Investigating the Presence of Bias

This study postulated in hypothesis 1 that a lie bias would be present when statements, provided in a second language, are assessed for veracity. The converse was formulated for the second hypothesis that a truth bias would be present when first language statements are reviewed. The results of the study do not support the hypothesis that a lie bias is extended to second language statements overall. Truthful second language statements, regardless of whether reviewed in a first or second language, are overall correctly assessed as truths and untrue second languages statements are overall incorrectly assessed as truths.

In the case of first language statements and the presence of a truth bias, the study found that an overall truth bias was extended to all the statements, regardless of whether the statement was provided in a first or a second language. The study found that deceptive statements, whether

provided in a first and second language, were more likely to be incorrectly assessed as truthful whilst truthful statements were more likely be correctly assessed as truthful by the statement veracity judges.

The outcomes of the investigations into bias in this empirical study is not aligned with the findings of literature that a lie bias is likely to be afforded to those speaking in their second language (Castillo, 2011; Cheng & Broadhurst, Da Silva & Leach, 2013; Evans & Michael, 2014; Meissner & Kassin, 2002). The only observation made regarding second language statements and the veracity judgment outcomes thereof, is that second language reviews of second language statements indicated a higher percentage of assessments being judged as lies compared to first language reviews of second language statements. Despite this narrowing of the difference between judging statements preferentially as lies as opposed to truths, a greater number of statements are still, albeit marginally assessed as truths in this category. So, whilst an overall lie bias towards second language statements was not demonstrated, in the specific case of second language statements reviewed by a second language judge the overall observation of a truth judgment being extended to all statements is moderated in the case where second language statements are reviewed by second language reviewers.

A future area of research should consider whether the results of this study is a result of the reviewer not being presented with any direct evidence of the language proficiency of the statement provider in respect to the statement language. The question can be asked whether knowledge of whether the statement language was provided in a first or second language would affect the veracity judges.

This study then concludes that, as suggested by Levine (2014), that not only do people enter (verbal) conversation adopting a truth-default state but that this truth-default state is also adopted when written statements are reviewed. This study finds that when written statements are

reviewed, with no additional input to observe or characterise the provider of the statements, reviewers are still truth biased.

This conclusion could be considered within Interpersonal Deception Theory and Truth-Default Theory frameworks. These frameworks hold that certain cognitive biases, such as visual and demeanour biases are present in veracity judging which influence veracity assessments, (Burgoon et al., 2008). Visual bias speaks to the reliance on non-verbal behaviour exhibited by speakers and the demeanour bias speaks to the communication style employed by the speaker (Burgoon et al., 2008). In this empirical study it was not possible for these biases to influence the veracity assessments as only written statements were being used to make veracity assessments. But this overall finding that a truth bias was still present despite these biases not being in play, suggests that this truth bias, which results from the truth-default state, is still in play when it comes to reviewing written statements.

However, there is one nuance to this conclusion that requires further attention. The study did find that when considering whether statement veracity judges are reviewing statements in their native language or not has a bigger impact on the outcomes of the veracity assessments than whether the language was provided by a first or second language statement provider.

No studies of second language statements, reviewed by a second language veracity judge was found in the literature review for this study. The international literature tends to limit the presence of a lie bias towards second language speakers, when they speak in their non-native tongue (Castillo, 2011; Cheng & Broadhurst, Da Silva & Leach, 2013; Evans & Michael, 2014; Meissner & Kassin, 2002). No distinction is made in literature accessed for this study regarding a review of a second language statement in a second language. To illustrate through means of a scenario, the question would be what factors would influence the veracity assessment when a first language isiXhosa speaker provides an English statement (i.e. a second language statement) and this statement is reviewed by an Afrikaans speaking first language reviewer (i.e. a second

language English reviewer)? This empirical study found that in this scenario the difference between correct and incorrect is moderated closer to chance than to the overall prevalence of an assessment of truth (aligned overall with the Truth Default Theory), This is an avenue of further enquiry to determine why this should be the case only for reviews conducted by second language reviewers.

This study can therefore not conclude that a lie bias is generally exhibited towards second language speakers, as it is only within a smaller subset of the second language statements that some modification of the outcomes has been observed. Therefore, the findings of this study do not support the first hypothesis.

The second line of enquiry pertaining to bias in this study was an investigating whether a truth bias was afforded to first language statements from first language reviewers (Hypothesis 2). The finding of this study that reviewers were generally truth biased (regardless of whether the statements being reviewed was in a first or second language) is in keeping with the assumptions as proposed by both Interpersonal Deception Theory and Truth-Default Theory – both theories argue that the receivers of the communication (in this case the reviewers) will be truth biased (Burgoon & Buller, 2014; Levine, Park, & McCornack, 1999; Levine, 2014). Studies argued that unless receivers of messages have a reason or experience an event that elicits a suspicion that deceit may be present, they will engage in the conversation assuming that the exchange is an honest on (Levine et al., 2010; Serota et al., 2010).

Given the design of the study, with the statement generation and statement judging being decoupled and done in isolation of each other, the opportunity for receivers of the messages (the statement veracity judges) to assess the actual process of statement generation was not available to reviewers. Results indicate that overall a truth bias was afforded to all statements and that there are not any significant differences between assessing either first or second language statements preferentially as truths. This contrasts with what Snelling (2013) and Solodukin

(2015), who found the greater one's proficiency with the language i.e. was the language use at a first language level, the more likely it was deemed a truth.

This study did find that the original statement veracity, in keeping with the veracity effect framework influenced the veracity judgment outcomes. More truthful statements were judged to be true than deceptive statements correctly assessed as lies. This is in keeping with existing international literature in terms of the veracity effect. As the veracity effect posits that truths will show greater correct classification as compared to lies if one were to separate classification between truths and lies (Levine et al., 1999).

To provide a more nuanced understanding of this overall truth bias, several post-hoc tests were conducted in this study. The post-hoc tests revealed that on average reviewers spent 31 minutes on 56 questions (~ 6 minutes per statement). It is proposed that this has contributed to the overall truth bias rating because it is indicative of heuristic processing being utilised to make veracity assessments. Heuristic processing is a mechanism through which decisions are made. This type of cognitive processing results in quick and rapid decision making and consumes little-to-no cognitive effort (Chaiken, Liberman, & Eagley, 1989; Evans, 2007). Heuristic processing often biases the judgements being made, as compared to the slower, more cognitively demanding analytic processing (Buller, Burgoon, Afifi, White, & Buslig, 1997; Burgoon, Buller, Afifi, White, & Hamel, 2008; Fiedler, & Walka, 1993). It is likely that participants made use of heuristic processing as there was no real consequences to them identifying the statement as a deceptive one. Robust research on how people make decisions when it comes to deception argues that people have a very strong tendency to take people and by extension their statements at face value (Fein, & Hilton, 1994) and that this is only negated by the elicitation of suspicion (Gilbert, 1991). Without this elicitation of suspicion that the person may have an ulterior motive, quick and biased information processing may be active.

Therefore, it is speculated that participants utilised heuristic processing in this study, rather than analytical processing, as there was no consequence to themselves or providers of the statements which they deemed as being deceptive and they had relatively short pieces of text to analyse. According to a study conducted by Masip, Garrido, and Herrero (2009) people tended to use heuristic processing when they were analysing or rating brief communications. It is argued that this occurred in this study, as all participants had to review relatively short statements.

Finally, this truth bias can be explained by how familiar the reviewer was with topic of the statement. Stiff, Miller, Sleight, Mongeau, Garlick and Rogan (1989) argued that as familiarity with the topic increased, the more likely it was the heuristic processing would be used when making veracity assessments. In this project, familiarity was proxied by the post-hoc test looking at how much the reviewer agreed with the sentiment expressed by the speaker. The more a reviewer agreed with the sentiments expressed, the more likely the statement would be assessed as truthful. It is proposed that this is an example of confirmation bias that was present in some instances as people often seek out information to confirm their believes. If the reviewer believes that the statement is a truth because it is speaking to their belief system, they will actively seek out confirmatory information when making their veracity assessments.

This overall truth bias could also potentially be explained from moral and social psychological perspectives. This line of reasoning argues that deception and lying is morally wrong, and this in turn feeds the principle of veracity. According to Bok (1999) the truth is always preferential barring certain special circumstances, and lying often requires a justification, where the truth requires no justification. Therefore, unless given reason to do so, people will not actively seek out the presence of deceit (Levine, Kim, & Hamel., 2010). Thus, people will typically act in accordance with the veracity principle, and many assume that others will as well. Hence, without a direct reason to suspect that the deceit is present, people will presume that others have no reason to lie, and therefore will not suspect the present of deceit (Levine et al.,

2010; Gilbert, 1991). In this particular study reviewers had no reason to actively suspect that the statement providers were instructed to lie, therefore, without the knowledge that the statement providers were asked to lie, they would have no reason to suspect deceit was present, and therefore, they would not think to question the veracity of the statement as anything other than honest.

Overall this study found that statement reviewers, regardless of; (a) whether statements were truthful or untruthful; (b) whether provided in a first or a second language; (c) whether the provided statement was under cognitive load or not, or (d) whether a statement was reviewed in a first or a second language, preferentially assessed the veracity of the statement as truthful. Overall, in the case of untruthful statements, 70.3% of the statements were incorrectly adjudicated to be truthful, and in the case of truthful statements, 73.9% was correctly assessed as truthful. The overall study results agree with the truth bias concept as postulated in both Interpersonal Deception Theory and Truth-Default Theory.

Impact of language proficiency on veracity adjudication. This study expected that the veracity assessment will differ dependent on whether the provided statement was a first or a second language statement. For this study native versus non-native language were used as indicators of language proficiency.

This study concludes that veracity assessment judgement outcomes are not affected by whether the statements were provided in a first or a second language if it the review was conducted in the native language of the reviewer. This study, counter to what was expected found that the judgement of statement veracities was overall more successful for statement veracity adjudicated by a second language judge compared to statements adjudicated by a first language judge. This was the case for both first and second language statements reviewed by a second language judge. The veracity judgment outcomes for first language reviews of both first and second language statements was not much different than a chance outcome. For first

language statements judged by first language reviewers 51.2% was correctly judged. For first language reviews of second language statements, 49.1% was correctly judged. The results were different for second language reviews. The percentage of second language reviews of first language statements correctly assessed was 54% and for second language reviews of second language statements 56.2%. In the case of first language statements adjudicated by a second language judge, the improvement in accuracy compared to first language adjudicators was not significant, only 2.8%. However, in the case of second language statements adjudicated by a second language judge, the improvement in accuracy compared to first language adjudicators was significant at 7.1%.

To provide more nuanced insights into this finding, a post-hoc assessment of reviewer's perceptions of language proficiency was conducted. This post hoc assessment found that when reviewers perceived the language abilities of the statements to be that of a non-native speaker, the accuracy of correct versus incorrect veracity assessments were similar to a chance outcome. This was the case for both deceptive and truthful statements. When reviewers perceived the statement to have come from a native speaker there was a marked difference between the veracity assessments in favour of judging the statements as true (in keeping with the truth default theory). Accepting for this argument that the perception of whether the statement was provided as a first or second language statement as proxy for language proficiency, would be in keeping with international studies by Snelling (2013) and Solodukin (2015), who found that the greater the language proficiency of the speakers, the greater the likelihood of it being assessed as a truth.

Impact of Statement Veracity on Veracity Judgement.

The fourth hypothesis of this study speculated that the original veracity of the statement would influence veracity assessments. The veracity effect was popularised by Levine and colleagues (1999), according to the veracity effect, if one separates overall detection rates by truth and lie, different picture emerges than the chance outcome as report by the bulk of

deception studies (Bond, & DePaulo, 2006; DePaulo et al., 2003). According to this effect, truths tended to be more correctly assessed as compared to lies (Levine et al, 1999). In keeping with Levine and colleagues (1999) the results of this study support the presence of the veracity effect. As can be seen in figure 2 page 73 in the case where original veracity was truthful, truths were more correctly identified compared to lie. In the case where original veracity was a lie, then overall original statement veracity did not improve accuracy. The exception to this observation was in the case of second language reviews of second language statements, where the difference between correct and incorrect assessments in the case of lies was $\Delta = 7.6\%$. Potential reasons offered for why the veracity effect exerts such an influence is because of heuristics. Levine et al., (1999) found that the more familiar the reviewer was the information being reviewed, the stronger the veracity effect, as was suggested by Stiff et al., (1989). In terms of this study, it is postulated that the veracity effect was present, not necessarily only due to how familiar (as measured by the extent of agreement with the speakers opinion) but also because, when reviewing statements, if the participants made use of heuristic processing they would not have to be methodical or analytical when it came to their assessments.

The impact of Cognitive Load on Veracity Assessment

The fifth hypothesis of the study was that participants would be better able to differentiate between truth-telling and lie-telling when cognitive load was imposed during statement generation. Past literature on the topic has shown that the introduction of cognitive load makes it harder to get away with deceit, and that the introduction improved detection accuracy (Blandón-Gitlin et al., 2014; Mann et al., 2002; Vrij, 2000; Walczyk et al., 2013). Contrary to this expectation, the accuracy of participants ability to discriminate between truthful and deceptive statements did not improve to above chance level when statements they were evaluating had been made under cognitive load. Half of the participants generating statements were placed under cognitive load. The participants selected for providing statements under

cognitive load, were required to tap a finger in time with a metronome whilst giving their respective statements (both in the first and second language). As indicated in figure 2, the introduction of cognitive load did not improve the discriminability between truthful and deceitful statements at all, with only a 0.7% improvement in the percentage of correct veracity assessment outcomes achieved with the introduction of cognitive load. This is in contrast to several findings (Hartwig et al., 2006; Vrij, & Mann, 2006; Vrij et al., 2007; Vrij et al., 2015a). In the studies reported, the cognitive load included story-telling or reporting in reverse chronological order or asking unexpected questions during the statement generation. This is likely because of the greater focus required from the participants to maintain the continuation of deceitful communication. It is speculated by the primary researcher that the amount of cognitive load induced on the participants of the reported studies was greater than the cognitive load imposed in the current study. It is proposed that this is the reason for the contrary outcomes observed in this study when comparing the accuracy of veracity judgments of cognitive load statement against the veracity judgement outcomes of statements provided without cognitive load imposed. It is important that a cognitive load task be chosen that taxes the sender (statement provider's) sufficiently to make it difficult to engage in deceitful communication, but not too taxing to make statements incoherent.

A different picture emerges from this study when the impact of cognitive load is assessed against the fixed effect attributes of first and second languages. When considering second language, untruthful statements, the introduction of cognitive load yielded unexpected outcomes in the veracity judgement of the statements.

One would expect that there would be a better chance of discriminability between truthful and untruthful statements achieved with the introduction of cognitive load on second language statements, given that communicating in a non-native language (second language) is already inherently cognitively taxing as reported in published studies (Acka & Elkilic, 2011; Andila,

2003; Da Silva, 2011; Da Silva & Leach, 2003; Perani & Abutalebi, 2005; Ullman, 2001). This inherent cognitive load in communicating in a second language whilst being deceptive stems from the premise that lying is a cognitively taxing activity on its own, as the deceiver must supress the truth, construct a credible story and continuously monitor that they are being perceived as credible (Evans & Michael, 2014). Add onto this already cognitively taxing task the fact that a sender now has to communicate in a second language, the availability or access to working memory is further reduced (Service et al., 2002). This reduced access to working memory makes it more difficult to construct and maintain a credible story whilst being untruthful.

This study found the converse, that there was a reduction in the percentage of correct outcomes for second language, untruthful statements under cognitive load reviewed by first language reviewers. In the case of second language, untruthful statements under cognitive load reviewed by second language reviewers the number of correct outcomes increased for the statements assessed by second language reviewers.

Human Evaluators and Computer Programme Detection Rates

LIWC2015 a software programme that analyses text and reports statistics of word category use. In recent years, linguistic profiles using the reported LIWC word categories have been developed and used to successfully differentiate between truths and lies. This study investigated whether a computer programme, in this case LIWC2015 (Linguistic Inquiry and word Count, 2015 dictionary), would outperform human evaluators at correctly assessing truthful from untruthful statements (Hypothesis 6). Two published linguistic profiles where analysed using LIWC2015 generated profiles to assess whether one would perform better at differentiating between truthful and untruthful statements than the other. The first of these linguistic profiles used word categories based on a natural language framework (NP profile) and is based on past literature in which LIWC was successfully used to differentiate between truth and lie (Newman

et al., 2003; Vrij, 2000). The second linguistic profile assessed was based on word categories used in a reality monitoring framework as proposed in 2005 by Bond and Lee. The Reality Monitoring framework performed better in correctly assessing the statement veracity compared to the NP framework which was based on past LIWC studies. It was expected that the veracity judgments using LIWC2015 generated language profiles analysed within published deceptive language profiles would perform significantly better at distinguishing between truthful and deceptive statements. However, in this study, although outperforming the human evaluators (50.4%), it was only slightly better than chance (53%).

Although this is encouraging, that a reality monitoring framework can be coded into LIWC2015 and outperform human evaluators, it needs further investigation. One of the major criticisms that exists when using reality monitoring categories to detect deception, is the lack of consistency among empirical studies. For example, Sporer (1997) included the following reality monitoring categories; clarity, sensory experiences, emotions and feelings, spatial and time information, realism and story re-constructability and cognitive operations in order to discriminate between truths and lies. Whereas, in Vrij, Akehurst, Soukara, & Bull (2004) only used the following reality monitoring categories; spatial and temporal details, cognitive operations, auditory and visual details. Porter and Yuille (1996) found that reality monitoring could not successfully differentiate between truth and lie if the following categories were used; self-references, verbal hedges, pauses and word count.

It should also be noted that past studies using LIWC to detect deceit, has often found that deceitful statements would contain fewer total number of words (Newman et al., 2003). However, this study could not consider this as a discriminating factor as all participants providing statements were instructed to provide three-minute long statements. This requirement of similar durations for verbal statements, whether truthful or untruthful resulted in similar length

statements. This is demonstrated in comparing the median word counts for truth statements (n = 51) is 345 with lie statements (n = 51) is 340.

The main results discussed only summarises the analysis performed against the two mentioned deceptive linguistic profiles. Appendixes V and VI contains the full LIWC generated word categories and language profiles for all 102 English statements.

Study Limitations and Recommendations for Future Research

The design of the study called for the statement generation and statement veracity assessment phases to be consecutive. This does not allow for any interaction between the statement provider and the statement veracity adjudicator, and therefore the simultaneous encoding and decoding happening between the two parties could not be assessed. As a result of this, the statement reviewer was never in a position to judge the sender of message and evaluate both verbal and non-verbal cues associated with deceit. This study only made use of undergraduate students to judge the veracity of a statement and as such the age distribution of participants were quite narrow and hence care should be taken to extrapolate the findings of this study to the broader population.

This study was limited to written (transcribed) statements. The study did not assess whether the outcomes would have been any different if the human evaluators listened to the statement providers, allowing the evaluators to form an opinion of the statement provider. It is proposed that a future study should account for this by comparing assessments of verbal statements against the transcribed statements.

The outcome of the veracity assessments of statements provided under cognitive load is not in keeping with published studies. A future avenue of research should address the interplay between the cognitive load inherently induced by speaking a second language and an additional task imposed on the statement provider.

Conclusion

This research was undertaken in order to determine if certain biases (namely a lie and truth bias) were extended to people when they gave statements in both their native and non-native tongue. Additionally, this study was done to determine if cognitive load facilitated the detection of deception in second language statements, and finally to see if a computer programme namely LIWC2015 outperformed human evaluators when it came to correct veracity assessment outcomes.

This study was conducted in two phases, a statement generation phase (Phase 1) and then a statement review phase (Phase 2). The study investigated whether biases are exhibited by statement evaluators toward second language statements. This area of enquiry is appropriate in South Africa, where despite only 9.6% of the population identifies with English as a first language, English is the accepted lingua franca. The initial results obtained by analysing the veracity judgement accuracy as a function of whether the statements were provided in first or second languages of the statement providers, did not indicate a clear differentiation between correct and incorrect assessments of the veracity assessment judges. In order to understand the results presented in the study, it was required to broaden the area of investigation to also consider the correspondence between the statement language and the first language of the veracity judge.

Overall, this study found that a truth bias was extended when reviewing written statements, confirming that people have a truth-default state as posited by Levine (2014) in his Truth-Default Theory. An overall lie bias towards statements provided in a second language of the statement provider could not empirically be demonstrated in this study. Results of the study were consistent with extant literature in terms of an overall truth bias, but the results were more marginal when restricting this assessment to the first language statements as required by the second research hypothesis. The assessment of the overall dataset supports the premise behind

the veracity effect framework that holds that truthful statements are more likely to be assessed correctly as truths than deceptive statements are likely to be correctly assessed as lies.

In contrast to what was expected the influence of cognitive load was not significant when assessed on the overall data set. However, when considering the smaller subset of the data representing cognitive load placed on second language statement providers. the effect of cognitive load is observed. Finally, this study found that the software programme LIWC2015 marginally outperformed human evaluators in making veracity assessments.

This study found, in keeping with studies reported in the literature, that the veracity assessment outcomes overall was still close to a chance level outcome

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 https://doi.org/10.1162/jocn.2007.19.2.287

Akca, C., & Elkilic, G. (2011). The effects of age and production mode on the bilingual performance of Turkish learners of English in a serial recall task. *International Journal of Psychology, 46*, 9–17. 10.1080/00207594.2010.503762

Akehurst, L., Arnhold, A., Figueiredo, I., Turtle, S., & Leach, A-M. (2018). Investigating deception in second language speakers: Interviewee and assessor perspectives. *Legal and Criminological Psychology*, 23, 230-251. https://doi.org/10.1111/lcrp.12127

Akehurst, L., Köhnken, G., & Höfer, E. (2001). Content credibility of accounts derived from live and video presentations. *Legal and Criminological Psychology, 6*, 65-83. doi:10.1348/135532501168208

Alpers, G. W., Winzelberg, A., Classen, C., Roberts. H., Parvati, Koopman, C. … Taylor, B. (2005). Evaluation of computerized text analysis in an internet breast cancer support group. *Computers in Human Behavior*, *21*, 361-37. doi:10.1016/j.chb.2004.02.008

Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology*¸ *52*(1), 27-58. https://doi.org/10.1146/annurev.psych.52.1.27

Anderson, D. E., DePaulo, B. M., Ansfield, M. E., Tickle, J. J., & Green, E. (1999). Beliefs about the cues to deception: Mindless stereotypes or untapped wisdom. *Journal of Nonverbal Behavior*, *23*(1), 67-69. https://doi.org/10.1023/a:1021387326192

Anolli, L., & Ciceri, R. (1997). The voice of deception: vocal strategies of naïve and able liars. *Journal of Nonverbal Behavior, 21*(4). 259-284. https://doi.org/10.1023/a:1024916214403

Ardila, A. (2003). Language representation and working memory with bilinguals. *Journal of Communication Disorders*, *36*, 233–240.doi: 10.1016/S0021-9924(03)00022-4

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. doi: 10.1017/50142716410000044

Bachenko, J., Fitzpatrick, E., & Schonwetter, M. (2008). Verification and implementation of language-based indicators in civil and criminal narratives. *In Proceedings of the 22nd International Conference on Computational Linguistics (COLING)* (pp. 41-48). Stroudsburg, PA: Association for Computational Linguistics.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01.

Bargh, J.A., & Chartrand, T.L. (1999). The unbearable automaticity of being. *American Psychologist, 54*(7), 462–479. https://doi.org/10.1037/0003-066x.54.7.462

Ben-Shakhar, G. (2002). A critical review of the control questions test (CQT). In M. Kleiner (Ed.) *Handbook of Polygraph Testing* (pp 103 - 126). Academic Press.

Ben-Shakhar, G & Dolev, H. (1996). Psychophysiological Detection Through the Guilty Knowledge Test: Effects of Mental Countermeasures. *Journal of Applied Psychology, 81*(3), 273-281.

Ben-Shakhar, G., & Elaad, E. (2002). The Guilty Knowledge Test as an Application of Psychophysiology. In M. Kleiner (Ed.) *Handbook of Polygraph Testing* (pp 87 -102). Academic Press

Ben-Shakhar, G., Bar-Hillel, M., & Kemnitzer, M. (2002). Trial by Polygraph: Reconsidering the Use of the Guilty Knowledge Technique in Court. *Law and Human Behavior, 26*(5), 527-541. doi: 0147-7307/02/1000-0527/1

Blandón-Gitlin, I., Fenn, E., Masip, J & Yoo, A. H. (2014). Cognitive-load approaches to detect deception: searching for cognitive mechanisms. *Trends in Cognitive Science¸ 18*(9), 441-444. http://dx.doi.org/10.1016/j.tics.2014.05.004

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

Bond, C. F., & Lee, A. Y. (2005). Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology, 19*, 313-329. https://doi.org/10.1002/acp.1087

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgements. *Personality and Social Psychology Review, 10*(3), 214–234 doi:10.1207/s15327957pspr1003_2

Bond, C. F., Jr., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin, 134*(4), 477–492. doi:10.1037/0033-2909.134.4.477

Bond, C. F., Levine, T. R., & Hartwig, M. (2014). New findings in non-verbal lie detection. In P. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting Deception* (pp. 37–58). Chichester, UK: John Wiley & Sons, Ltd.

Burgoon, J. K., Blair, J. P., Qin, T., & Nunamaker Jr, J. F. (2003). Detecting deception through linguistic analysis. *Proceedings of the Symposium on Intelligence and Security Informatics*. New York: Springer-Verlag

Burgoon, J. K. (2015). Rejoinder to Levine, Clare et al.'s comparison of the Park-Levine probability model versus Interpersonal deception theory: Application to deception detection. *Human Communication Research, 41*(3). doi:10.1111/hcre.12065

Burgoon, J. K. (2018). Predicting veracity from linguistic indicators. *Journal of Language and Social Psychology*, 1-29. doi:10.1177/026197X18784119.

Burgoon, J. K., Buller, D. B., Ebesu, A., & Rockwell, P. (1994). Interpersonal deception: V. Accuracy in deception detection. *Communication Monographs*, *61*, 303 – 325. https://doi.org/10.1080/03637759409376340

Burgoon. J. K., Buller, D. B, Afifi, W., White, C. H., & Hamel, L. (2008). *Cognitive heuristics and the stability of receiver judgements in detecting deception*. Paper presented at the 41st Hawaii International Conference on System Sciences, Kona.

Burgoon, J. K., Blair, J. P., & Strom, R. E. (2008). Cognitive biases and nonverbal cue availability in detecting deception. *Human Communication Research, 34*(4), 572–599. doi:10.1111/j.1468- 131 2958.2008.00333.x

Burgoon, J. K., & Buller, D. B. (2014). Interpersonal deception theory. In D. O. Braithwaite & P. Schrodt (Eds.), *Engaging Theories in Interpersonal Communication: Multiple Perspectives* (2nd ed.). Sage Publications.

Burgoon, J. K., Buller, D. B., Guerrero, L. K., Afifi, W., & Feldman, C. (1996). Interpersonal deception: XII. Information management underlying deceptive and truthful messages. *Communication Monographs, 63*(1), 50-69. https://doi.org/10.1080/03637759609376374

Burgoon, J. K., Mayew, W. J., Girboney, J. S., Elkins, A. C., Moffit, K., Dorn, B., Byrd, M., & Spitzley, L. (2015). Which spoken language markers identify deception in high-stakes settings? Evidence from earnings conference calls. *Journal of Language and Social Psychology, 34*, 1-35

Burgoon, J. K., & Qin, T. (2006). The dynamic nature of deceptive verbal communication. *Journal of Language and Social Psychology*, *25*, 76-96.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, *6*(3), 203–242. doi:10.1111/j.1468-2885.1996.tb00127.x

Buller, D. B., Burgoon, J. K., Buslig, A., & Roiger, J. (1996). Testing interpersonal deception theory: The language of interpersonal deception. *Communication Theory*, *6*, 268-289. https://doi.org/10.1111/j.1468-2885.1996.tb00129.x

Buller, D. B., Burgoon, J. K., Afifi, W., White, C. H., & Buslig, A. L. S. (1997). Judging message credibility: The impact of heuristics, cognitive busyness, and strategic behavioral adaptation, *International Communication Association*, Montreal.

Buller, D. B., & Burgoon, J. K. (2008). Interpersonal deception theory. In L. Baxter & D. O. Braithwaite (Eds.), *Engaging theories in interpersonal communication: Multiple perspectives*. London: Sage Publications

Burgoon, J. K., & Floyd, K. (2000). Testing for the motivation impairment effect during deception and truthful interaction. *Western Journal of Communication*, *64*(3), 243-267. https://doi.org/10.1080/10570310009374675

Castillo, P. A., Tyson, G., & Mallard, D. (2014). An investigation of accuracy and bias in cross cultural lie detection. *Applied Psychology in Criminal Justice*, *10*, 66–82

Carlucci, M. E., Compo, N. S., & Zimmerman, L. (2013). Lie detection during high-stakes truths and lies. *Legal and Criminological Psycholology*, *18*, 314-323.

Caldwell-Harris, C. L., & Ayçiçeği-Dinn, A. (2009). Emotion and lying in a non-native language. *International Journal of Psychophysiology*, *71*(3), 193–204. https://doi.org/10.1016/j.ijpsycho.2008.09.006

Cladwell-Harris, C. (2014). Emotionality between a native and foreign language; theoretical implications. *Frontiers in Psychology*, *5* (1055). DOI: 10.3389/fpsyg.2014.01055

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212-252). New York: Guilford Press

Cheng, K. H. W., & Broadhurst, R. (2005). The detection of deception: The effects of first and second language on lie detection ability. *Psychiatry, Psychology and Law, 12*(1), 107–118. doi:10.1375/pplt.2005.12.1.107

Christ, S. E., van Essen., D. C., Watson, J. M., Brubaker, L. E., & McDermott, K. B. (2009). The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analysis. *Cerebral Cortex ,19*. 1557-1566 https://doi.org/10.1093/cercor/bhn189

Caldwell-Harris, C. L., & Ayçiçeği-Dinn, A. (2009). Emotion and lying in a non-native language. *International Journal of Psychophysiology*, 71, 193–204. https://doi.org/10.1016/j.ijpsycho.2008.09.006

Caso, L., Gnisci, A., Vrij, A., & Mann, S. (2005). Processes underlying deception: An empirical analysis of truth and lies when manipulating the stakes. *Journal of Investigative Psychology and Offender Profiling*, 2, 195-202. https://doi.org/10.1002/jip.32

Davatzikos, C. R., Ruparel. K., Fan, Y., Shen, D., Acharyya, M. … Langleben, D. (2005). Classifying spatial patters of brain activity with machine learning methods: Application to lie detection. *NeuroImage, 28*(3)¸663-668. https://doi.org/10.1016/j.neuroimage.2005.08.009

Da Silva. C., S. (2011). *Detecting Deception in Second-Language Speakers* (master's thesis). University of Ontario Institute of Technology.

Da Silva, C. S., & Leach, A. M. (2013). Detecting deception in second-language speakers. *Legal and Criminological Psychology, 18*(1), 115–127. doi:10.1111/j.2044-8333.2011.02030.x

Deeb, H., Vrij, A., Hope, L., Mann, S., Leal, S., Granhag, P. A., Strömall, L. A. (2017). The Devil's Advocate approach: An interview technique for assessing consistency among deceptive and truth-telling pairs of suspects. *Legal and Criminological Psychology, 23*(1), 31 – 57. https://doi-org.ezproxy.uct.ac.za/10.1111/lcrp.12114

DePaulo, B. M. (1994). Spotting lies: Can humans do better? *Current directions in psychological science, 3*(3), 83-86. https://doi.org/10.1111/1467-8721.ep10770433

DePaulo, B. M., Kashy, D. A., Kirkendon, S. E., Wyer. M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*(5), 979-995. https://doi.org/10.1037//0022-3514.70.5.979

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*(1), 74–112.doi:10.1037//0033-2909.129.1.74

DePaulo, B. M., Kirkendol, S. E., Tang, J., & O'Brien, T. P. (1988). The motivational impairment effect in communication of deception: Replication and extensions. *Journal of Nonverbal Behaviour,12*(3), 177-202. https://doi.org/10.1007/bf00987487

Dewaele, J.-M. (2004). The emotional force of swearwords and taboo words in the speech of multilinguals. *Journal of Multilingual and Multicultural Development, 25,* 204 –222. https://doi.org/10.1093/oxfordhb/9780198808190.013.15

Dewaele, J.-M. (2008). The emotional weight of I love you in multilinguals' languages. *Journal of Pragmatics, 40*(10), 1753–1780. https://doi.org/10.1016/j.pragma.2008.03.002

Dowle, M., & Srinivasan, A. (2019). data.table: Extension of `data.frame`. R package version 1.12.2. https://CRAN.R-project.org/package=data.table

Duñabeitia, J. A., & Costa, A. (2015). Lying in a native and foreign language. *Psychonomic Bulletin & Review, 22,* 1124–1129. http://dx.doi .org/10.3758/s13423-014-0781-4

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*(3), 439-462. https://doi.org/10.1017/s0142716410000068

Dunbar, N. E., Jensen, M. L., Bessabarova, E., Burgoon, J. K., Bernard, D. R., Roberston, K.J., … Eckstien, J. M. (2014). Empowered by persuasive deception: The effects of power and deception on interactional dominance, credibility, and decision-making. *Communication Research, 41*(6), 852-876. https://doi.org/10.1177/0093650212447099

Dragojevic, M., & Giles, H. (2016). I don't like you because you're hard to understand: The role of process fluency in the language attitudes process. *Human Communication Research, 42*, 396–420. https://doi.org/10.1111/hcre.12079

Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology, 71,* 5-24. doi: 10.1037/0022-3514.71.1.5

Ekman, P. (1992). Are there basic emotions? *Psychological Review, 99*(3), 550-553. https://doi.org/10.1037//0033-295x.99.3.550

Ekman, P. (2001). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. New York, NY: W. W. Norton & Company, Inc

Ekman, P. (2009). *Telling lies: Cues to deceit in the marketplace, politics, and marriage*. New York, NY: W. W. Norton.

Ekman, P., & Friesen, W, V. (1969). Nonverbal leakage and clues to deception. *Psychiatry, 32*, 88-102. https://doi.org/10.1080/00332747.1969.11023575

Ekman, P., & Friesen, W, V. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology, 29*(3), 288 – 298. https://doi.org/10.1037/h0036006

Evans, J. R., Meissner, C. A., Michael, S. W., & Brandon, S. E. (2013). Validating a new assessment method for deception detection: Introducing a Psychologically Based Credibility Assessment Tool. *Journal of Applied Research in Memory and Cognition, 2*, 33-41. doi: 10.1016/j/jarmac.2013.02.002.

Evans, J. R., & Michael, S. W. (2014). Detecting deception in non-native English speakers. *Applied Cognitive Psychology, 28*(2), 226–237. doi:10.1002/acp.2990

Evans, J. R., Pimental, P. S., Pena, M. M., & Michael, S. W. (2017). The ability to detect false statements as a function of the type of statement and the language proficiency of the statement provider. *Psychology, Public Policy and Law, 23*(3), 290-300. https://doi.org/10.1037/law0000127

Frank, M. G., & Eckman, P. (1997). The ability to detect deceit generalizes across different types of high-stakes lies. *Journal of Personality and Social Psychology, 72*(6), 1429-1439. ttps://doi.org/10.1037//0022-3514.72.6.1429

Frank, M. G., Feeley. T. H., Paolantonio, N., & Servoss, T. J. (2004). Individual and small group accuracy in judging truthful and deceptive communication. *Group Decision and Negotiation, 13*(1), 45-59. https://doi.org/10.1023/b:grup.0000011945.85141.af

Fein, S., & Hilton, J. L. (1994). Judging others in the shadow of suspicion. *Motivation and Emotion, 18*(2), 167-198. https://doi.org/10.1007/bf02249398

Fiedler, K., & Walka, I. (1993). Training lie detectors to use nonverbal cues instead of global heuristics. *Human Communication Research, 20*(2), 199-223. https://doi.org/10.1111/j.1468-2958.1993.tb00321.x

Fuller, C. M., Biros, D. P., & Wilson, R. L. (2009). Decision support for determining veracity via linguistic-based cues. *Decision Support Systems, 46*(3), 695-703. https://doi.org/10.1016/j.dss.2008.11.001

Fuller, C. M., Biros, D. P., Burgoon, J. K., & Nunamaker, J. F. (2013). An examination and validation of linguistic constructs for studying high-stakes deception. *Group Decision and Negotiation*, *22*(1), 117-134. https://doi.org/10.1007/s10726-012-9300-z

Forgas, J. P. Baumeister, R. F. & Tice, D. M. (2009). The psychology of self-regulation: an introductory review. In J. P. Forgas, R. F. Baumeister, & D. M. Tice (Eds). *Psychology of Self-Regulation: Cognitive, Affective and Motivational Processes (The Sydney Symposium of Social Psychology 11)* (pp.1-17). Psychology Press, New York.

Ganis, G., Kosslyn, S., Stose, S., Thompson, W., & Yurgelun-Todd, D. (2003). Neural correlates of different types of deception: An fmri investigation. *Cerebral Cortex*, *13* (8), 830-836. https://doi.org/10.1093/cercor/13.8.830

Gilbert, D. (1991). How mental systems believe. *American Psychologist*, *46*(2), 107-119. https://doi.org/10.1037//0003-066x.46.2.107

Giolla, Ask, Granhag., & Karisson. (2019). Can reality monitoring criteria distinguish between true and false intentions? *Journal of Applied Research in Memory and Cognition*, *8*(1), 92-97.

Gollan, T. H., & Acenas, L. A. (2004). What is a TOT? Cognate and translation effects on tip-of-the-states in Spanish-English and Tagalog-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 246–269. 10.1037/0278-7393.30.1.246

Granhag, P.A. & Hartwig, M. (2008). A new theoretical perspective on deception detection: On the psychology of instrumental mind-reading. *Psychology, Crime & Law*, *14*(3), 189- 200. https://doi.org/10.1080/10683160701645181

Gregersen, T. S. (2005). Nonverbal cues: Clues to the detection of foreign language anxiety. *Foreign Language Annuals*, *38*, 388-400. doi:10.1111/j.1944- 9720.2005.tb02225.x

Hancock, J. T., Curry, L. E., Goorha. S., & Woodworth, M. (2008). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, *45*(1), 1-23. https://doi.org/10.1080/01638530701739181

Hancock, J. T., Woodworth, M. T., & Goorha, S. (2010). See no evil: The effect of communication medium and motivation on deception detection. *Group Decision and Negotiation*, *19*(3), 327-343. https://doi.org/10.1007/s10726-009-9169-7

Hansen, K., & Dovidio, J. F. (2016). Social dominance orientation non-native accents and hiring recommendations. *Cultural Diversity and Ethnic Minority Psychology*, *22*(4), 544–551. https://doi.org/10.1037/cdp0000101

Hartig, F. (2019). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.2.4. https://CRAN.R-project.org/package=DHARMa

Hauch, V., Blandón-Gitlin, I., Masip, J., Sporer, S. L. (2015). Are computers effective lie detectors? A meta-analysis cues of deception. *Personality and Social Psychology Review*, *19*(4), 307-342. https://doi.org/10.1177/1088868314556539

Hartwig, M., & Bond, C., J., Jr. (2011). Why do lie-catchers fail? A lens model meta-analysis human lie judgement. *Psychological Bulletin, 137*(4), 643-659. https://doi.org/10.1037/a0023589

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Vrij, A. (2005). Detecting deception via strategic disclosure of evidence. *Law and Human Behavior*, *29*(4), 469–484. https://doi.org/10.1007/s10979-005-5521-x

Hartwig, M., Granhag, P.A., Strömwall, L.A., & Kronkvist, O. (2006). Strategic Use of Evidence during Police Interviews: When training to detect deception works. *Law and Human Behavior, 30*(5), 603-619. https://doi.org/10.1007/s10979-006-9053-9

Hartwig, M., Granhag, P. A., & Luke, T. (2014). Strategic Use of Evidence During Investigative Interviews: The State of the Science. In D. C. Raskin, C. R. Honts & J. C Kircher (Eds). *Credibility Assessment Scientific Research and Applications.* Elsevier Inc.

Harvey, A. C., Vrij, A., Nahari, G., & Ludwig, K. (2016). Applying the verifiability approach to insurance claims settings: Exploring the effect of the information protocol. *Legal and Criminological Psychology*. https://doi.org/10.1111/lcrp.12092

Hernández-Castañeda, A., & Calvo, H. (2017). Deceptive text detection using continuous semantic space models. *Intelligent Data Analysis*, *21*, 679-695. doi: 10.3233/IDA-170882.

Hollien, H. (1990). *The acoustics of crime: The new science of forensic phonetics*. New York: Plenum.

Iacono, W. G., & Lykken, D. T. (1997). The validity of the lie detector: Two surveys of scientific opinion. *Journal of Applied Psychology, 82*(3), 426-433.doi: 10.1037/0021-9010.82.3.426

Johnson, M. K., & Rye, C. L. (1981). Reality Monitoring. *Psychological Review, 88*(1), 67-85. https://doi.org/10.1037//0033-295x.88.1.67

Joffe, R. D. (1992). *Criteria-Based Content Analysis: An experimental investigation with children.* Unpublished doctoral dissertation, Department of Psychology, University of British Columbia

Jupe, L. M., Leal, S., Vrij, A., & Nahari, G. (2017). Applying the Verifiability Approach in an international airport setting. *Psychology, Crime & Law, 23*(8), 812–825

Jupe, L. M., Vrij, A., Leal, S., & Nahari, G. (2018). Are you for real? Exploring language use and unexpected process questions within the detection of identity deception. *Applied Cognitive Psychology*, *32*, 622–634. https://doi.org/10.1002/acp.3446

Kassin, S. M., Meissner, C. A., & Norwick, R. J. (2005). "I'd know a false confession if I saw one": A comparative study of college students and police investigators. *Law and Human Behavior 29,* 211-227.

Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science, 23*(6), 661-۸.

Khokhlova, I. N. (2015). Global English: Africanization of the English language. *Review of European Studies*, 7(6), 201–208. doi:10.5539/res.v7n6p201

Knowles, J. E., & Frederick, C. (2019). merTools: Tools for Analyzing Mixed Effect Regression Models. R package version 0.5.0. https://CRAN.R-project.org/package=merTools

Kolkman, M. C. (2012). Deception detection: importance of modality. *16th Twente Student Conference on IT*. Enschede, The Netherlands.

Köhnken, G. (2004). Statement Validity Analysis and the 'detection of the truth' In P. A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp 41-64). Cambridge University Press.

Kormos, J., & Denes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *Systems*, *32*(2), 145-164. https://doi.org/10.1016/j.system.2004.01.001

Leach, A-M., Snellings, R. L., & Gazaille, M. (2017). Observers language proficiencies and the detection of non-native speaker's deception. *Applied Cognitive Psychology, 31*(2), 247-257. doi: 10.1002/acp.3322

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 Psychologica.* doi: 10.1016/j.actpsy.2010.03.005

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology, 46,* 1093-1096. doi:10.1016/j.jsep2010.05.025

Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect." *Communication Monographs, 66*(2), 125–144. doi:10.1080/03637759909376468

Levine, T. R., Kim, R. K., & Hamel, L. M. (2010). People lie for a reason: An experimental test of the principle of veracity. *Communication Research Reports, 27*(4), 271-285. https://doi.org/10.1080/08824096.2010.496334

Levine, T. R. (2014). Truth-Default Theory (TDT): A Theory of Human Deception and Deception Detection. *Journal of Language and Social Psychology*, *33*(4), 378-392. https://doi.org/10.1177/0261927x14535916

Lüdecke, D. (2018). _sjPlot: Data Visualization for Statistics in Social Science_. doi: 10.5281/zenodo.1308157 (URL: http://doi.org/10.5281/zenodo.1308157), R package version 2.6.2, <URL:https://CRAN.R-project.org/package=sjPlot>.

Luke, T. R., Hartwig, M., Brimbal, L., Chan, G., Jordan. S., Joseph, E., & Osborne, J. (2012) Interviewing to Elicit Cues to Deception: Improving Strategic Use of Evidence with General-To-Specific Framing of Evidence. *Journal of Police and Criminal Psychology, 28*(1), 54-62. https://doi.org/10.1007/s11896-012-9113-7

Mairesse. F., Walker, M. A., Mehl, M., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research, 30*(1),547-500. https://doi.org/10.1613/jair.2349

Makowski, D. (2018). The Psycho Package: An Efficient and Publishing-Oriented Workflow for Psychological Science. *Journal of Open Source Software*, *3*(22), 470. Available from https://github.com/neuropsychology/psycho.R

Mann., S., Vrij., A., & Bull., R. (2002). Suspects, Lies and Videotapes: An Analysis of Authentic High-Stake Liars. *Law and Human Behavior, 26*(3), 365-367. https://doi.org/10.1023/a:1015332606792

Masip, J., Garrido., E, & Herrero, C. (2004). Defining Deception. *Anales de Psicologia, 20*(1), 147 – 171.

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. doi: 10.1080/10683160410001726356

Masip, J., Garrido, E., & Herrero, C. 2009. Heuristic versus systematic processing of information in detecting deception: Questioning the truth bias. *Psychological Reports*.

McCornack, S. A. (1992). Information manipulation theory. *Communication Monographs, 59*(1), 1-16. https://doi.org/10.1080/03637759209376245

McCornack, S., Morrison, K., Paik, J. E., Wisner, A. M., & Zhu, X. (2014). Information manipulation theory 2: A propositional theory of deceptive discourse production. *Journal of Language and Social Psychology, 33*, 348-377. doi:10.1177/0261927X14534656

Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgements of truth and deception. *Law and Human Behavior, 5*, 469-480. doi:10.1023/A:1020278620751

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*, 57-74. doi: 10.1017/S0140525X10000968

Mihalcea, R., & Strapparava, C. (2009). The lie detector: Explorations in the automatic recognition of deceptive language. In *Proceedings of the Association for Computational Linguistics (ACL-IJCNLP 2009)*, Singapore, 309-312.

Mitchell, R.W. (1986). A framework for discussing deception. In R.W. Mitchell and N.S. Mogdil (eds.), Deception: perspectives on human and nonhuman deceit. State University of New York Press, Albany.

Nahari, G., Vrij, A., & Fisher, R. P. (2014a). Exploiting liars' verbal strategies by examining the verifiability of details. *Legal and Criminological Psychology, 19*, 227-239. https:// doi.org/10.1111/j.2044-8333.2012.02069.x.

Nahari, G., Vrij, A., & Fisher, R. P. (2014b). The verifiability approach: Countermeasures facilitate its ability to discriminate between truths and lies. *Applied Cognitive Psychology, 28*, 122 - 128. https://doi.org/10.1002/acp.2974

Nahari, G. (2018). The Applicability of the Verifiability Approach to the Real World. In J. P. Rosenfeld (ed.), *Detecting Concealed Information and Deception: Recent Developments*. Academic Press.

Nahari G. (2019) Verifiability Approach: Applications in Different Judgmental Settings. In: Docan-Morgan T. (eds) *The Palgrave Handbook of Deceptive Communication*. Palgrave Macmillan, Cham.

Newmann, 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*(5), 665 - 675. https://doi.org/10.1177/0146167203029005010

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175-220.

Nortje, A., & Tredoux, C. G. (2019). How good are we at detecting deception? A review of current techniques and theories. *South African Journal of Psychology*, 1-14. doi: 10.1177/0081246318822953

Oberlader, V. A., Naefgen, C., Koppehele-Gossel, J., Quinten. L., Banse, R. … Schimdt, A.F. (2016). Validity of content-based techniques to distinguish true and fabricated statements: A meta-analysis. *Law and human behavior, 40*(4), 440-457. https://doi.org/10.1037/lhb0000193

O'Sullivan, M., Ekman, P., & Friesen, W. V. (1988). The effect of comparisons on detecting deceit. *Journal of Nonverbal Behavior, 12*(3), 203–215. https://doi.org/10.1007/bf00987488

O'Sullivan, M., & Ekman, P. (2004). The wizards of deception detection. In P. A. Granhag, & L. A. Strömwall (Eds.), *Deception detection in forensic contexts* (pp. 269–286). Cambridge, England: Cambridge University Press.

Ott, M., Cardie, C., & Hancock, J. T. (2013). Negative deceptive opinion spam. In L. Vanderwende, H. Daumé III, & K. Kirchoff (Eds.), *Proceedings of the 2013 conference of the North America chapter of the association for computerized linguistics: Human language* (pp.497-501). Atlanta, Georgia: Association for Computational Linguistics.

Palmatier, J. J., & Rovner, L. (2015). Credibility assessment: Preliminary Process Theory, the polygraph process, and construct validity. *International Journal of Psychophysiology, 95*(1), 3-15. ttps://doi.org/10.1016/j.ijpsycho.2014.06.001

Parker, A. D., & Brown, J. (2000). Detection of deception: Statement validity analysis as a means of determining truthfulness or falsity of rape allegations. *Legal and Criminological Psychology, 5*, 237-259. doi: 10.1348/135532500168119

Patterson, T. D. (2009). The Effect of Cognitive Load on Deception (Unpublished doctoral thesis). Florida International University, Miami, Florida

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

Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. Current Opinion in Neurobiology, 15, 202-206. doi: 10.1016/j.conb.2005.03.007

Pérez-Rosas, V., & Mihalcea, R. (2014). Cross-cultural deception detection, In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, (440–445). Baltimore, Maryland, USA.

Petty, R. E., Wegener, D. T., & Fabrigar, L. R. (1997). Attitudes and attitude change. *Annual Review of Psychology*, *48*, 609-647. doi: 10.1146/annurev.psych.48.1.609

Picornell, I. (2013). *Cues to deception in a textual narrative context* (published doctoral thesis). Department of Psychology: Aston University.

Porter, S., & ten Brinke, L. (2008). Reading between the Lies: Identifying Concealed and Falsified Emotions in Universal Facial Expressions. *Psychological Sciences, 19*(5), 508-14. https://doi.org/10.1111/j.1467-9280.2008.02116.x

Porter, S., & Yuille, J. C. (1996). The language of deceit: an investigation of the verbal cues to deception in the interrogation context. *Law and Human Behavior, 20*(4), 442 – 458. https://doi.org/10.1007/bf01498980

Porter, S., ten Brinke, L., & Wallace, B. (2012). Secrets and lies: Involuntary leakage in deceptive facial expressions as a function of emotional intensity. *Journal of Nonverbal Behavior*, *36*(1), 23-37. https://doi.org/10.1007/s10919-011-0120-7

Qin., T., Burgoon, J. K., & Nunamaker, J. F., Jr. (2004). An exploratory study on promising cues in deception detection and application of decision trees. In *Proceedings of the 37thHawaii International Conference of Systems Sciences.* Retrieved from https//www. researchgate.net/publication/221184315_An_Exploratory_Study_on_Promising_Cues_in_ Deception_Detection_and_Application_of_Decision_Trees

Raskin, D. C. & Honts, C. R. (2002). The Comparison Question Test. In M. Kleiner (Ed.) *Handbook of Polygraph Testing* (pp 1-49). Academic Press. Remland, M. S. (2009). *Nonverbal communication in everyday life*. Houghton Mifflin Harcourt (HMH).

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Rude, S. S., Gortner, E. M., & Pennebaker, J. W. (2004). Language use of depressed and depression-vulnerable college students. *Cognition and Emotion, 18*(8), 1121 -1133. https://doi.org/10.1080/02699930441000030

Serota, K. B., & Levine, T. R., (2014). A Few Prolific Liars: Variation in the Prevalence of Lying. *Journal of Language and Social Psychology*, 1-20. doi: 10.1177/0261927x14528804

Service, E., Simola, M., Metsanheimo, O., & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology*, *14*, 383–408. 10.1080/09541440143000140

Snelling, R. (2013). *The effect of language proficiency on second-language lie detection*. University of Ontario Institute of Technology

Solodukhin, E. (2015). *The effect of target language proficiency on deception detection among undergraduate students*. University of Ontario Institute of Technology.

SouthAfrica-info.com. (2018). What languages are spoken in South Africa's nine provinces? Retrieved from: https://southafrica-info.com/infographics/animation-languages-south-africas-provinces/

Sporer, S. L. (1997). The less travelled road to truth: verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology, 11*, 373-397. https://doi.org/10.1002/(sici)1099-0720(199710)11:5<373::aid-acp461>3.0.co;2-0

Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology, 20*(4), 421-446. doi: 10.1002/acp.1190.

Sporer, S. L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: a meta- analytic synthesis. *Psychology, Public Policy and Law, 13*(1), 1–34. https://doi.org/10.1037/1076-8971.13.1.1

Steller, M., & Köhnken, G. (1989). Criteria-based statement analysis. In D. C. Raskin (Ed.). *Psychological methods in criminal investigation and evidence* (pp. 217-245). New York, USA: Springer

Stiff, J. B., Miller, G. R., Sleight, C., Mongeau, P., Rogan, R., & Garlick, R. Explanations for Visual Cue Primacy in Judgments of Honesty and Deceit. *Journal of Personality and Social Psychology 56*(4), 555-564.

Street, C. N. H., & Richardson, D. C. (2015). Lies, damn lies, and expectations: How base rates inform lie-truth judgements. *Applied Cognitive Psychology, 29*(1), 149–155. doi:10.1002/acp.3085.

Ströfer, S., Noordzij, M, J., & Ufkes, E, G., & Giebels, E. (2015). Deceptive Intentions: Can Cues to Deception Be Measured before a Lie is Even Stated. *PLos ONE, 10(5)* doi:10.1371/ journal.pone.0125237.

Strömwall, L. A., Granhag, P. A., & Hartwig, M. (2004). Practitioners' beliefs about deception. In P. A. Granhag, & L. A. Strömwall (Eds.), *The Detection of Deception in Forensic Contexts* (pp. 15–40). Cambridge: Cambridge University Press.

Strömwall, L. A., Bengtsson, L., Leander, L., & Granhag, P. A. (2004). Assessing children's statements: The impact of a repeated experience on CBCA and RM ratings. *Applied Cognitive Psychology, 18*, 653-668. doi: 10.1002/acp.1021

Suchotzki, K., & Gamer, M. (2018). The Language of Lies: Behavioral and Autonomic Costs of Lying in Native Compared to a Foreign Language. *Journal of Experimental Psychology: General, 147*(5), 734-746. https://doi.org/10.1037/xge0000437

ten Brinke, L., Porter, S., & Baker, A. (2012). Darwin the detective: Observable facial muscle contractions reveal emotional high-stake lies. *Evolution and Human behaviour, 33*(4), 411-416. https://doi.org/10.1016/j.evolhumbehav.2011.12.003

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29, 24-54. doi: 0.1177/0261927x09351676

Undeutsch, U. (1989). The development of statement validity analysis. In U. Undeutsch (Ed.) *Credibility assessment* (pp.101-121). Dordrecht, Netherlands: Kluwer.

Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgements of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 219–230. https://doi.org/10.1037/0278-7393.33.1.219

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognitive, 4*, 105-122. doi: 10.1017/S1366728901000220.

van Vuuren, H. H. J. (2017). Second language bias and accuracy of deception judgments. (published Masters thesis). Department of Psychology. University of South Africa.

Vrij, A., Semin, G. R., & Bull, R. (1996). Insight into behavior displayed during deception. *Human Communication Research, 22*(4), 544-562. https://doi.org/10.1111/j.1468-2958.1996.tb00378.x

Vrij, A., & Baxter, M. (1999). Accuracy and confidence in detecting truths and lies in elaborations and denials: Truth bias, lie bias and individual differences. *Expert Evidence, 7*(1), 25–36. https://doi.org/10.1023/a:1008932402565

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

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities.* Chichester: England: Wiley.

Vrij, A., Edward, K., Roberts, K. P., Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, *24*(4), 239-263. https://doi.org/10.1023/a:1006610329284

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA scores. *Law and Human Behavior, 26*, 261-283. doi: 10.1023/A:1015313120905

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004). Let me inform you how to tell a convincing story: CBCA and reality monitoring scores as a function of age, coaching, and deception. *Canadian Journal of Behavioural Science, 36*, 113-126. doi:10.1037/h0087222

Vrij, A. (2005). Criteria-based content analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law, 11*, 3-41. doi:10.1037/1076-8971.11.1.3

Vrij, A., & Mann, S. (2006). Criteria-Based Content Analysis: An empirical test of its underlying processes. *Journal of Psychology, Crime, & Law, 12*(4), 337-349, doi: 10.1080/10683160500129007.

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by Manipulating cognitive load. *Trends in Cognitive Sciences, 10*(4), 141-142. https://doi.org/10.1016/j.tics.2006.02.003

Vrij, A., Mann, S., Kristen, S., & Fisher., R. (2007). Cues to Deception and ability to detect lies as a function of police interview styles. *Law and Human Behavior, 31*(5), 499-518. https://doi.org/10.1007/s10979-006-9066-4

Vrij, A., Mann, S., 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(3),* 253–265. doi:10.1007/s10979-007-9103-y.

Vrij, A., Granhag, P. A., & Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest, 11*(3), 89–121. doi:10.1177/1529100610390861.

Vrij, A., Leal, S., Mann, S., & Fisher, R. (2012). Imposing cognitive load to elicit cues to deceit: Inducing the reverse order technique naturally. *Psychology, Crime & Law, 18*(6), 579–594. doi:10.1080/1068316X.2010.515987.

Vrij, A., & Ganis, G. (2014). Theories in deception and lie detection. In Raskin, D.C., Honts, C.R & Kircher, J.C. (Eds.), *Credibility Assessment: Scientific Research and Applications* (pp. 301–374). Academic Press

Vrij, A., Fisher, R. P., & Blank, H. (2015a). A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology*, 1–21. doi:10.1111/lcrp.12088.

Vrij, A., Taylor, P., & Picornell, I. (2015b). Verbal Lie Detection. In G. Oxburgh, T. Myklebust., T. Grant & R. Milne (Eds.). *Communication in Investigative and Legal Context: Integrated Approaches from Forensic Psychology, Linguistic and Law Enforcement* (pp. 259 – 287). John Wiley and Sons, Ltd.

Vrij, A. (2018). Deception and truth detection when analysing verbal and nonverbal cues. *Applied Cognitive Psychology*, 1-8

Wang, Y, Xue, G., Chen, C., Cue, F., & Dong, Q. (2007). Neural bases of asymmetric language switching in second-language learners: An ER-fMRI study. *NeuroImage, 35,* 862-870

Wagenaar, W. A., van Koppen, P. J., & Crombag, H. M. (1993). *Anchored narratives: the psychology of criminal evidence.* St Martin's Press, New York.

Walker, W. R., Skowronski, J. J., & Thompson, C. P. (2003). Life is pleasant – and memory helps to keep it that way! *Review of General Psychology, 7*(2), 203-210. https://doi.org/10.1037/1089-2680.7.2.203

Walker, W. R., Vogl, R. J., & Thompson, C. P. (1999). Autobiographical memory: unpleasantness fades faster than pleasantness over time *Applied Cognitive Psychology*, *11*(5). 399-413. https://doi.org/10.1002/(SICI)1099-0720(199710)11:5<399::AID-ACP462>3.0.CO;2-E

Walczyk, J. J., Roper, K., 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. https://doi.org/10.1002/acp.914

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 for lie detection. *Personnel Psychology, 58*(1), 141–170. https://doi.org/10.1111/j.1744-6570.2005.00484.x

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*(1), 33–49. https://doi.org/10.1007/s10869-009-9090-8

Walczyk, J. J, Griffith, D. A., Yates, R., Visconte, S. R., Simoneaux, B., & Harris, L. L. (2012). Lie detection by inducing cognitive load: eye movements and other cues to the false answers of false answers of "witnesses" to crimes. *Criminal Justice and Behavior*. *39*(7), 887–909. https://doi.org/10.1177/0093854812437014

Walczyk, J. J., Igou, F. P., Dixon, A. P., & Tcholakian, T. (2013). Advancing lie detection by inducing cognitive load: a review of relevant theories and techniques guided by lesson from polygraph-based approaches. *Frontiers in psychology*, 1(3) doi: 10.3389/fpsyg.2013.00014.

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

Williams, S. M., Talwar, V., Lindsay, R. C. L, Bala, N, & Lee, K. (2014). Is the truth in your words? Distinguishing children's deceptive and truthful statements. *Journal of Criminology*. https://doi.org/10.1155/2014/547519

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Woodworth, M., Hancock, J. & Goorha, S. (2005). The Motivational Enhancement Effect: implications for our chosen modes of communication in the 21st century. *HICSS* 38.

Zapamiuk, J., Yuille, J. C., & Taylor, S. (1995). Assessing the credibility of true and false statements. *International Journal of Law and Psychiatry*, *18*(3), 343-352. https://doi.org/10.1016/0160-2527(95)00016-b

Zhou, L., Burgoon, J. K., Nunamaker, Jr, J. F., & Twitchell, D. (2004). Automating linguistic based cues for detecting deception in text-based asynchronous computer-mediated communications. *Group Decision and Negotiation*, *13*(1), 81-106. https://doi.org/10.1023/b:grup.0000011944.62889.6f

Zimmerman, L. A., Veinott, E. S. Meissner, C. M., Fallon, C., & Mueller, S. T. (2010). Field Training and Testing of Interview and Elicitation Methods (Final Report prepared under Contract W91CRB-09-C-0082 for US ARMY REDCOM Acquisition Center, Aberdeen Proving Ground, MD). Fairborn, OH: Klein Associates Division of Applied Research Associates.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, 14. 1-59. New York, NY: Academic Press.

Zuckerman, M, & Driver, R.E. (2014). Telling lies: Verbal and Non-Verbal Correlates of Deception. In Siegman, A. W. and Feldstein, S (eds). *Multichannel Interrogations of Non-verbal Behavior* (pp. 129-149), S. Psychology Press.

Appendix I: Consent Form 1a

Dear Student,

Thank you for expressing interest to be part of in this study. This study seeks to determine whether statements made by first and second language speakers are treated differently by receivers of the communication.

In order to do the analysis, we will ask you to specify your home (or first) language. Please note that you will be voice recorded in parts of the study. Furthermore, please note that your agreement to partake in the study is completely voluntary and that you may withdraw from the study at any given time, without any repercussions.

The only information we require from you for this experiment is your student number for the data analysis part of this research project. Please also take note that all information gathered during this research project will be confidential in nature, and that your student number will not be made available to the public or used in any part of the final report. Please also note that the recordings of the second phase of the data collection will also be confidential in nature and not made available to the general public

There are no foreseeable risks with this experience, and you should not experience any psychological distress nor stress and no harm should befall you. Should you wish to leave the study early, please inform the researcher and your data will be disregarded. You need not provide a reason, you simply need to ask for the data provided until your point of departure to be disregarded, and the data shall be disregarded and destroyed in the appropriate manner.

Consent

I hereby confirm that this phase of the experiment in which I voluntary partake has been explained to me and that I hereby give my informed consent to partake in the study. I understand

that my participation is voluntary and that I may withdraw my participation at any stage, without repercussions or penalty.

Finally, I hereby confirm that I speak English at a first/second language level.

Student Number: _____

Date: _____

Signature: _____

Appendix II: Consent Form 1b

Dear Student,

Thank you for expressing your continued interest to be part of in this study. As explained before this study seeks to determine whether statements made by first and second language speakers are treated differently by receivers of the communication. In order to fully assess this question, we will need to ask whether there are natural biases exhibited by reviewers when they assess statements made in languages other than their own.

In order to do this part of the analysis, we will ask you to provide a total of 4 3-minute statements on two of the topics you have responded to in phase 1a. Please note that for two of these statements we will ask you to be untruthful – to express a contrary view to that indicated during phase 1a. In essence we will require 2 truthful and two untruthful statements from you for phase 1b. We need the untruthful statements to assess whether it is possible to differentiate between truths and lies under different statement providing conditions as well as different statement receiving conditions. Please note that you will be voice recorded in parts of the study. Furthermore, please note that your agreement to partake in the study is completely voluntary and that you may withdraw from the study at any given time, without any repercussions.

As before, please note that all information gathered to date and going forward should you consent is confidential. At no time will personal identification information be associated with any of the statements. Please be informed that the 4 will be transcribed for subsequent analysis of review panels during phase 2. No personal identification information of the statement providers will be available to the reviewers of the statements.

There are no foreseeable risks with this experience, and you should not experience any psychological distress nor stress and no harm should befall you. However, the researcher acknowledges that you are being asked to be deceptive and this may cause you some

psychological distress or discomfort. Should this be of concern for you at this stage it is recommended that you withdraw. If you elect to continue and you develop some concerns or stress because of your participation you can contact student wellness at the following number (021) 650 1020/17. Should you wish to leave the study early, please inform the researcher and your data will be disregarded. You need not provide a reason, you simply need to ask for the data provided until your point of departure to be disregarded, and the data shall be disregarded and destroyed in the appropriate manner.

Consent

I hereby confirm that this phase (phase 1b) of the experiment in which I voluntary partake has been explained to me and that I hereby give my informed consent to partake in the study. I understand that my participation is voluntary and that I may withdraw my participation at any stage, without repercussions or penalty.

Finally, I hereby confirm that I speak English at a first/second language level.

Student Number: _____

Date: _____

Signature: _____

Appendix III: Consent Form 2

Dear Student,

Thank you for expressing interest and to partake in this study. This study seeks to investigate the notion of lie bias and second language as well as the influence of cognitive load on detecting deception. Please note that you during this stage you will be asked to make veracity judgements on pre-recorded and transcribed statements.

The only information we require from you for this experiment is your student number for the data analysis part of this research project. Please also take note that all information gathered during this research project will be confidential in nature, and that your student number will not be made available to the public. Please also note that the recordings of the second phase of the data collection will also be confidential in nature.

There are no foreseeable risks with this experience, and you should not experience any psychological distress nor stress and no harm should befall you. However, the researcher acknowledges that you are being made to be deceptive and this may cause you some psychological distress or discomfort. If this does happen to you can contact student wellness at the following number (021) 650 1020/17. Should you wish to leave the study early, please note the researcher may still keep and use the data that you provided till your voluntary withdrawal, however, if you explicitly ask for your data to be disregarded, the researcher shall do so. You need not provide a reason, you simply need to ask for the data provided until your point of departure to be disregarded, and the data shall be disregarded and destroyed in the appropriate manner.

I hereby confirm that the experiment has been explained to me and that I hereby give my voluntary consent to partake in the study. I understand that my participation is voluntary and that

I may withdraw my participation at any stage, without penalty. Finally, I hereby confirm that I speak English at a first/second language level.

Student Number: _____

Date: _____

Signature: _____

Appendix IV: List of Topics

1) The mother of an unborn child has the right to abort without the father's consent.
2) It is the government's job to provide free education.
3) Homosexual people do not have the right to get married.
4) Military service should be mandatory for all 18-year-old male South Africans.
5) Until South Africa addresses sport inequality, it should be banned from hosting World Cup sports events.
6) South African citizens who are working overseas should pay the difference in the tax rates if it differs from the tax they would have been paying in South Africa, to support the economy.
7) The Death Penalty is an appropriate punishment for crimes such as murder and rape.
8) School children should not be allowed to attend schools that are not in the community in which they live.
9) Assisted suicide should be allowed for the terminally ill if they request it, and if they are found to be fully cognisant of the fact that they are asking a physician to assist in ending their life, being aware that they are essentially asking to die?
10) The legal age for the consumption of alcohol should be increased to 21 years.
11) The speed limit on national roads should be reduced to 100 km/h to reduce the number of road accidents and to be more fuel efficient.
12) No school assembly should be opened with prayer.
13) English and Afrikaans need to be removed from the National Anthem of South Africa.
14) The consumption of marijuana should be legalised in South Africa.
15) The consumption of psychedelic drugs should be legalised at festivals, in South Africa.

Appendix V: LIWC word category summaries

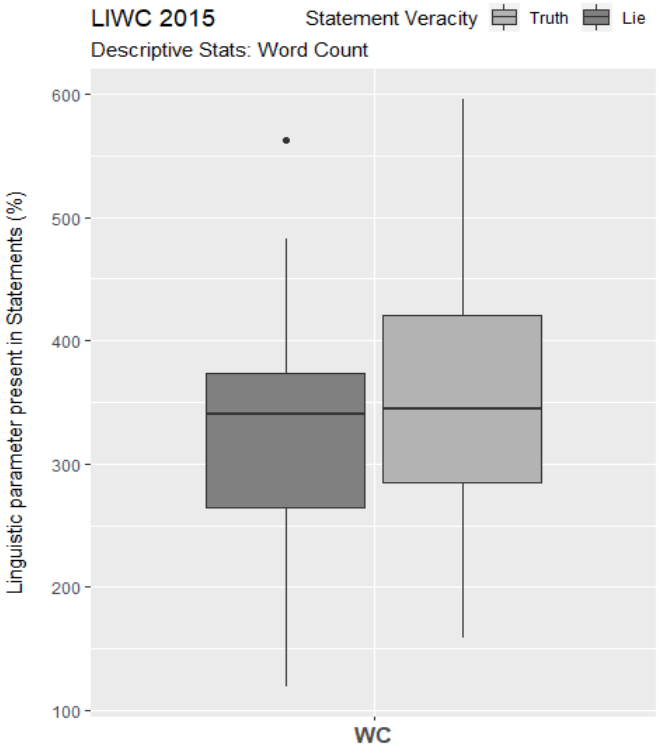

Figure 23 LIWC2015 Summary: Word Count

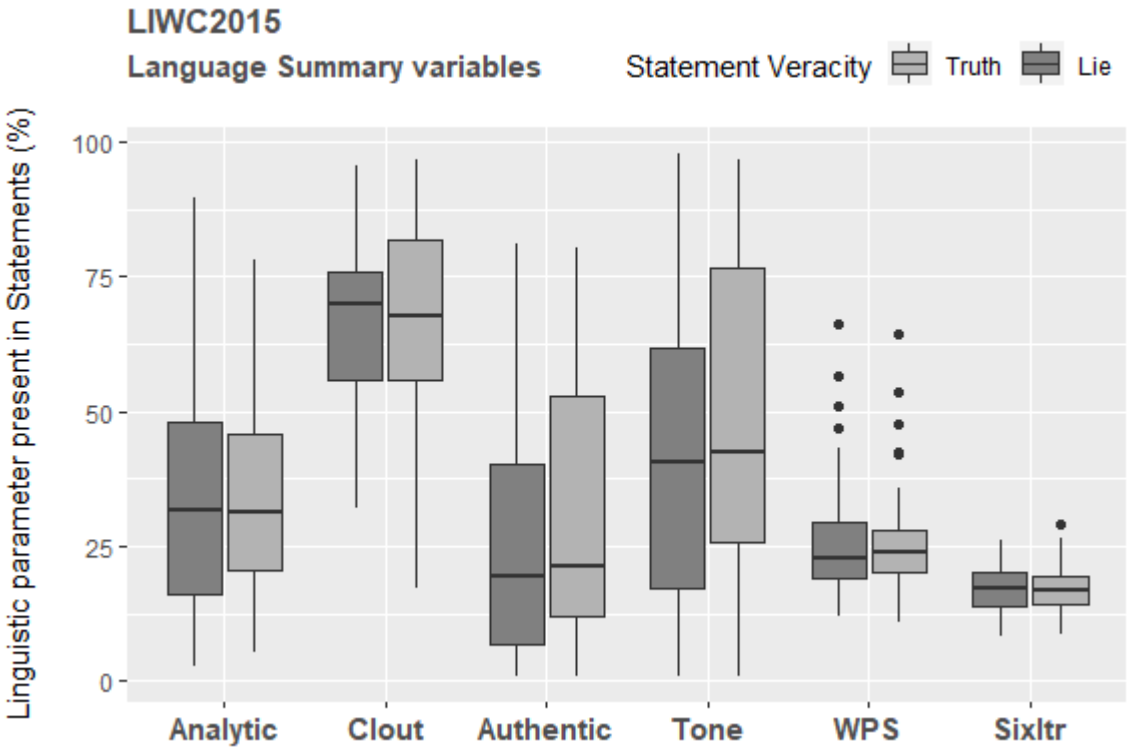

Figure 22 LIWC2015 Summary: Language Attribute (grouped variables) Summaries

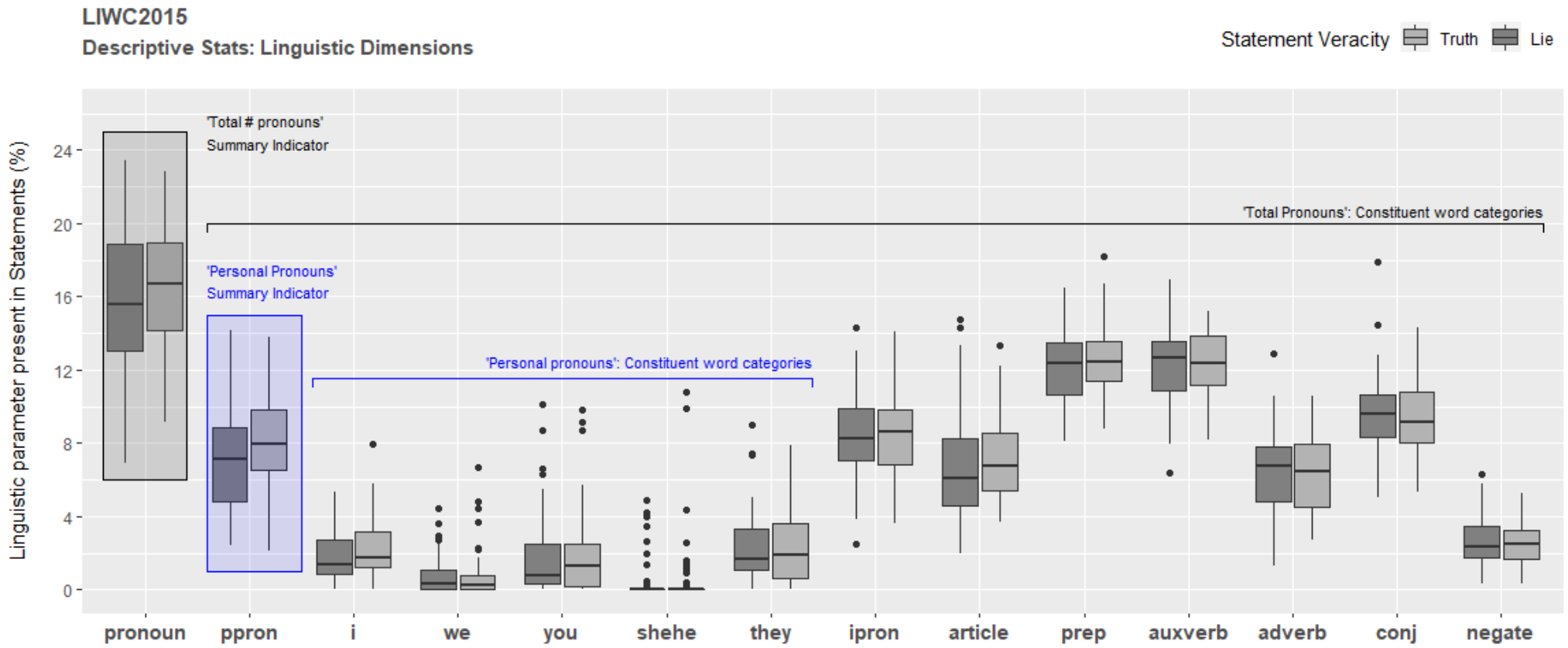

Figure 24 LIWC2015 Summary: Linguistic Dimensions

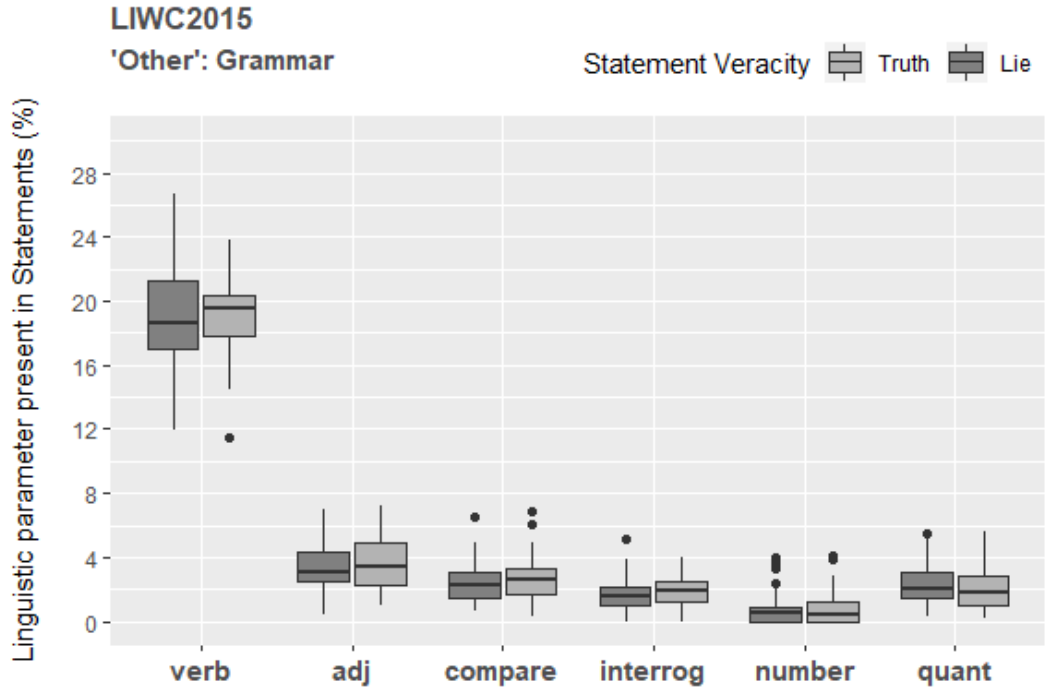

Figure 26 LIWC2015 Summary: Grammar categories

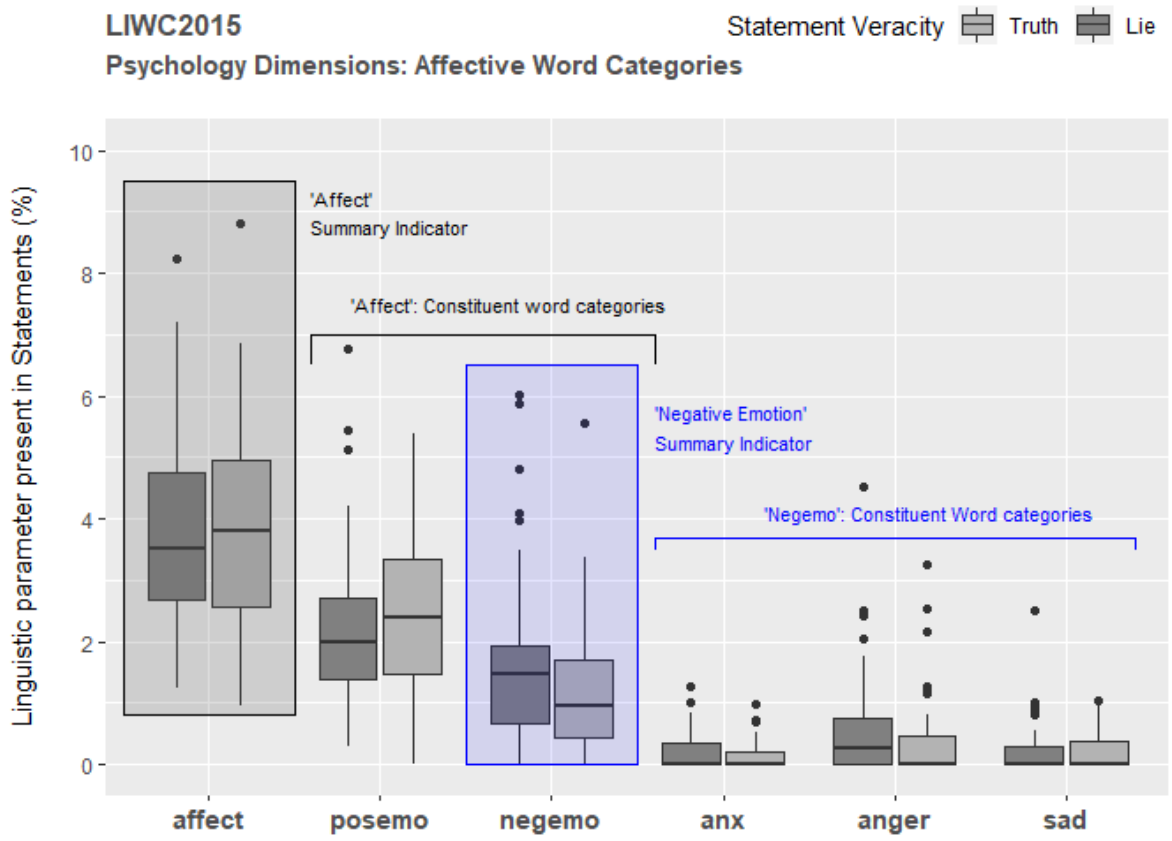

Figure 25 LIWC2015 Summary: Affective Word Categories

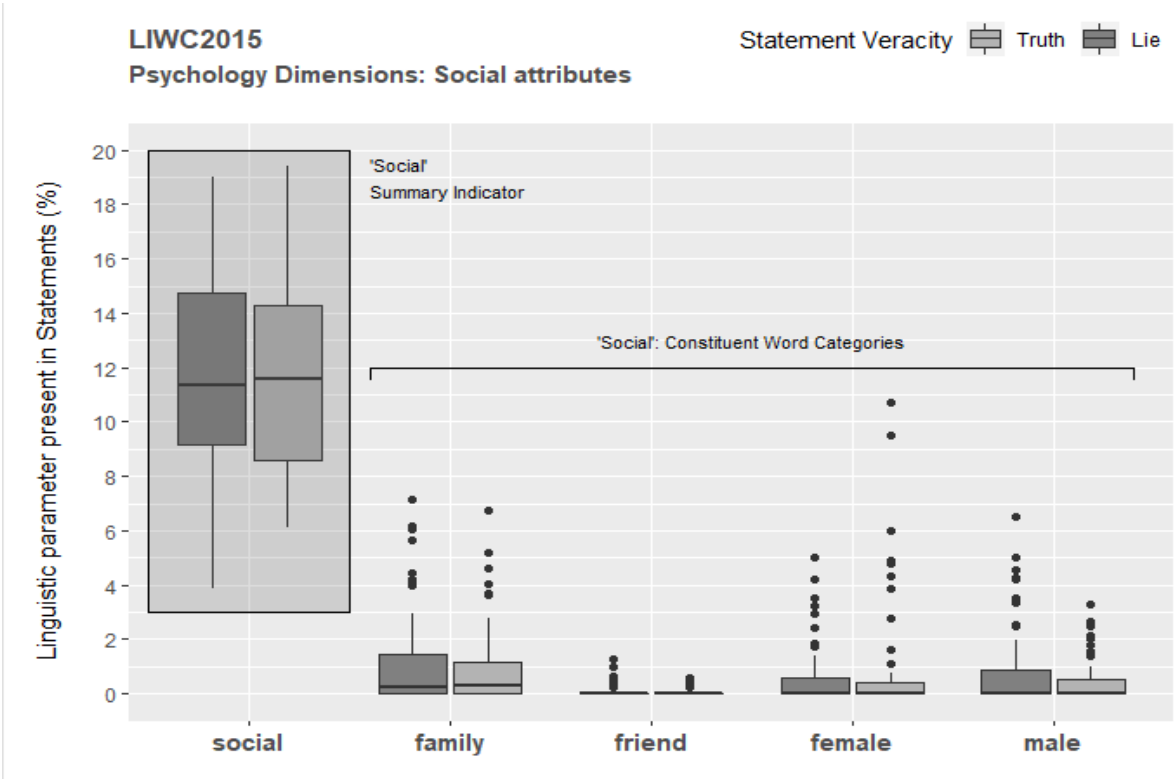

Figure 28 LIWC2015 Summary: Social Attribute Word Categories

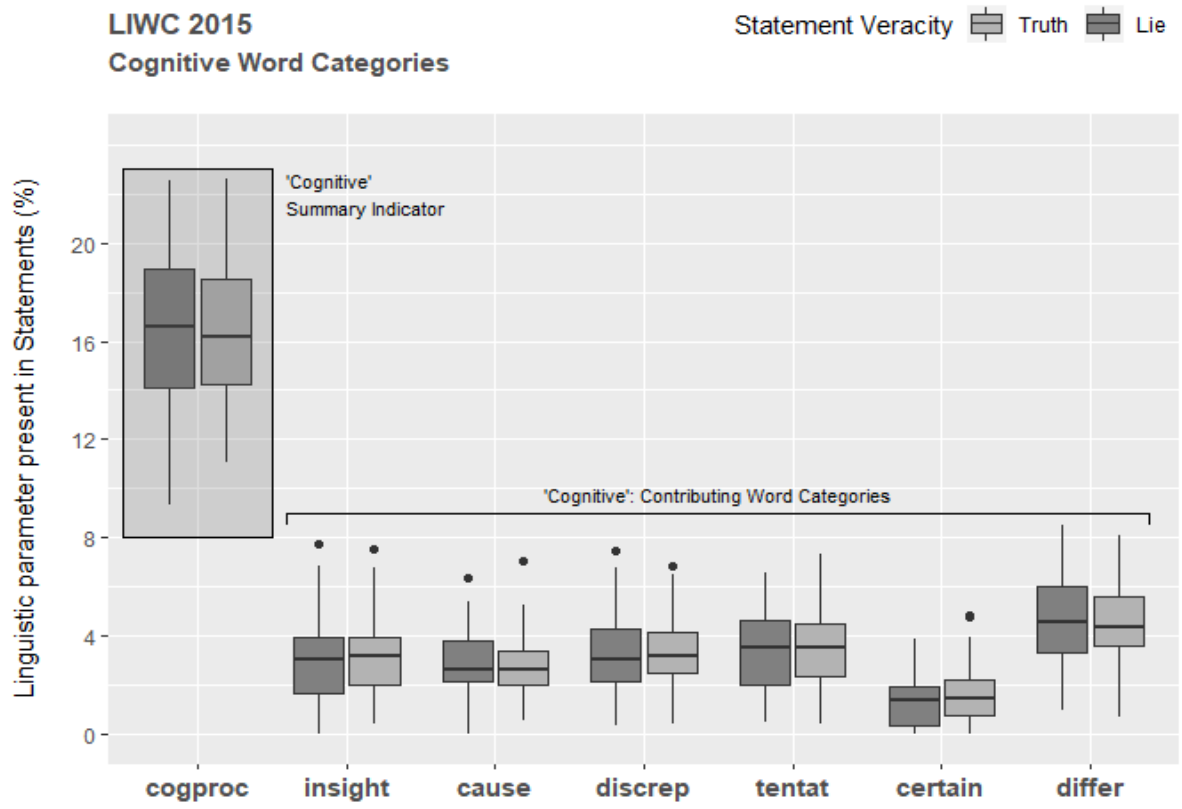

Figure 27 LIWC2015 Summary: Cognitive Word Categories

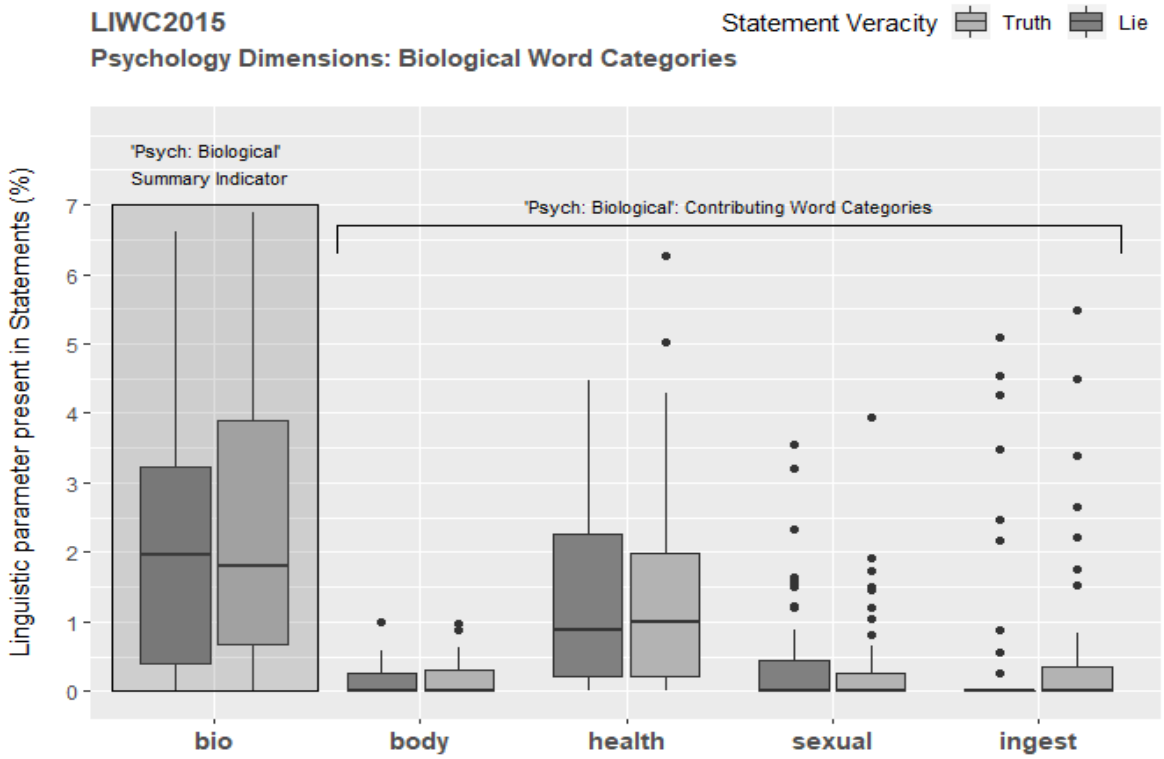

Figure 30 LIWC2015 Summary: Biological Word Categories

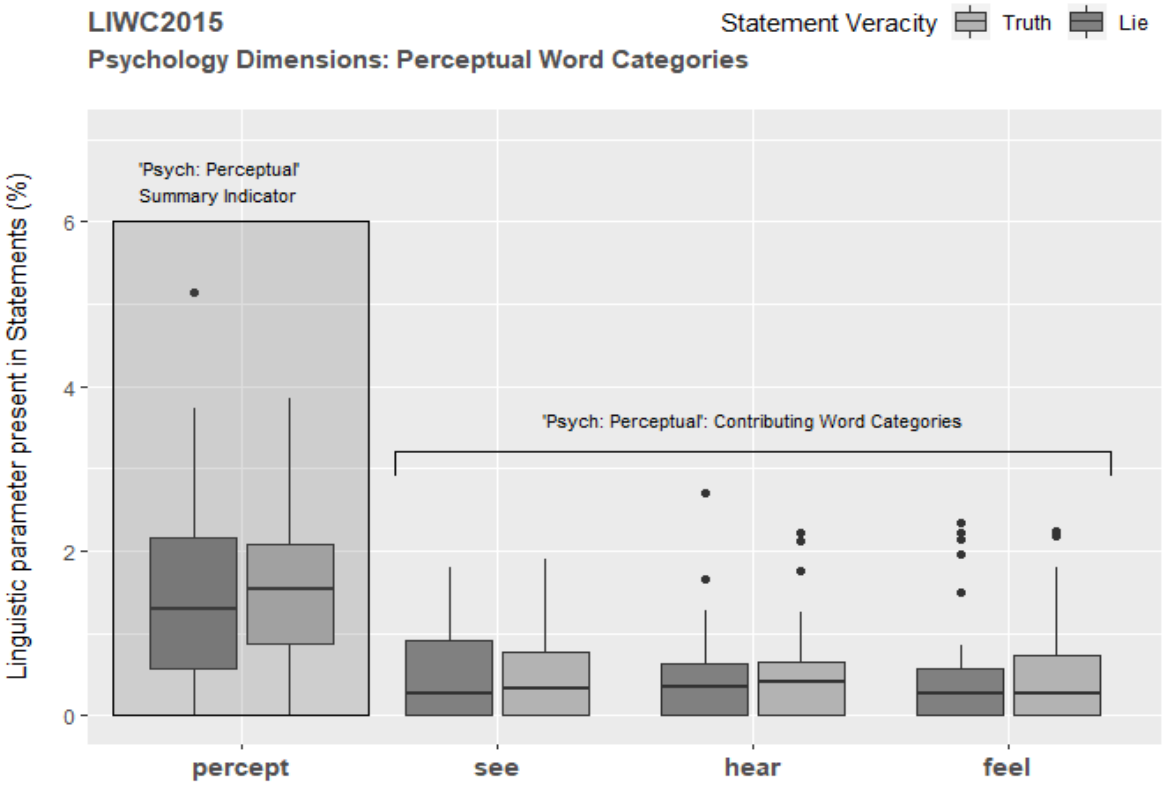

Figure 29 LIWC2015 Summary: Perceptual Word Categories

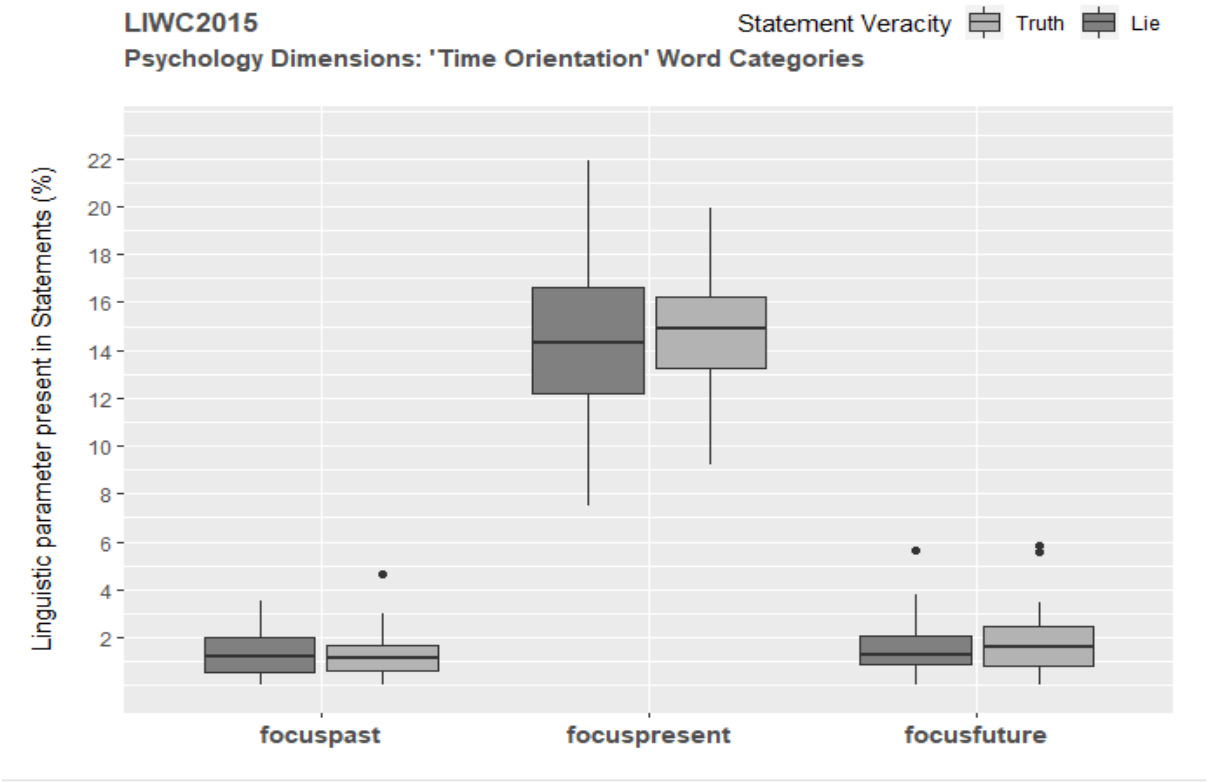

Figure 32 LIWC2015 Summary: 'Time Orientation' Word Categories

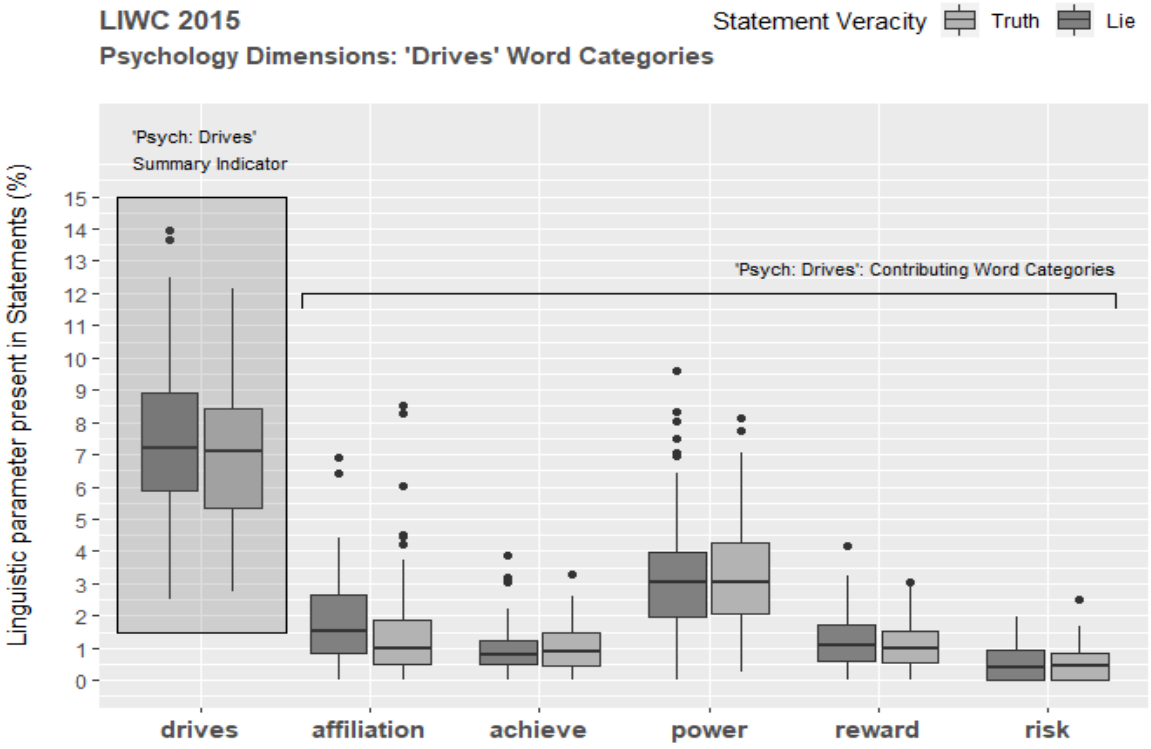

Figure 31 LIWC2015 Summary: 'Drives' Word Categories

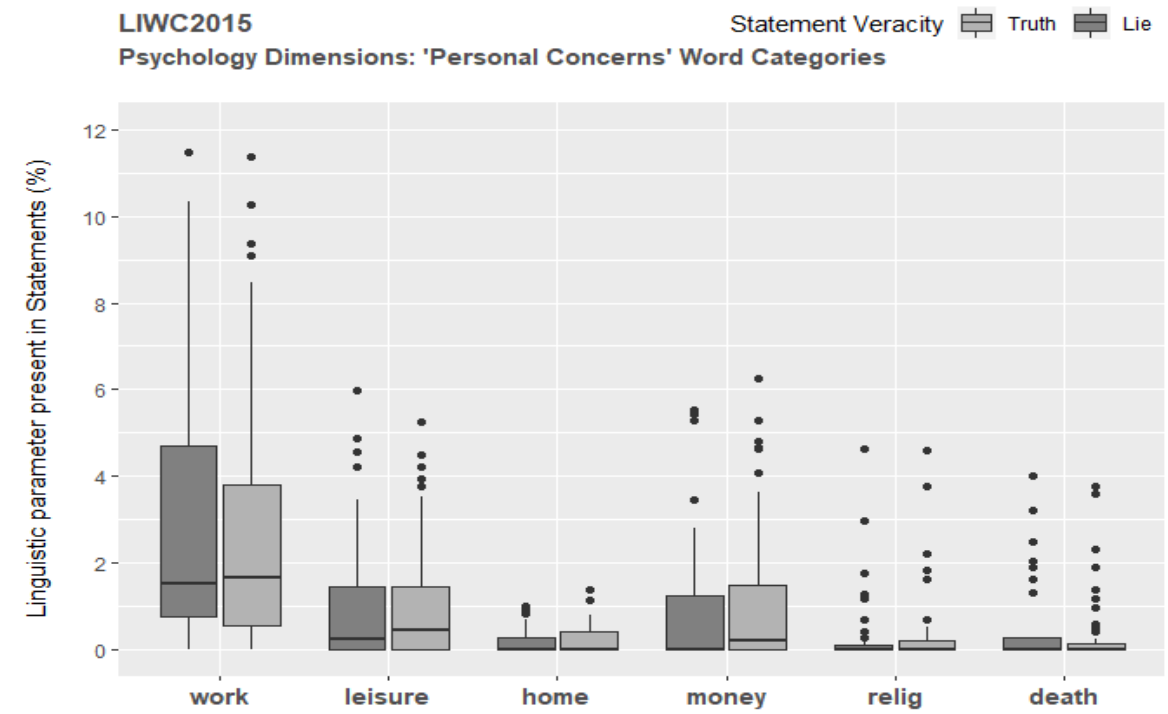

Figure 34 LIWC2015 Summary: 'Personal Concerns' Word Categories

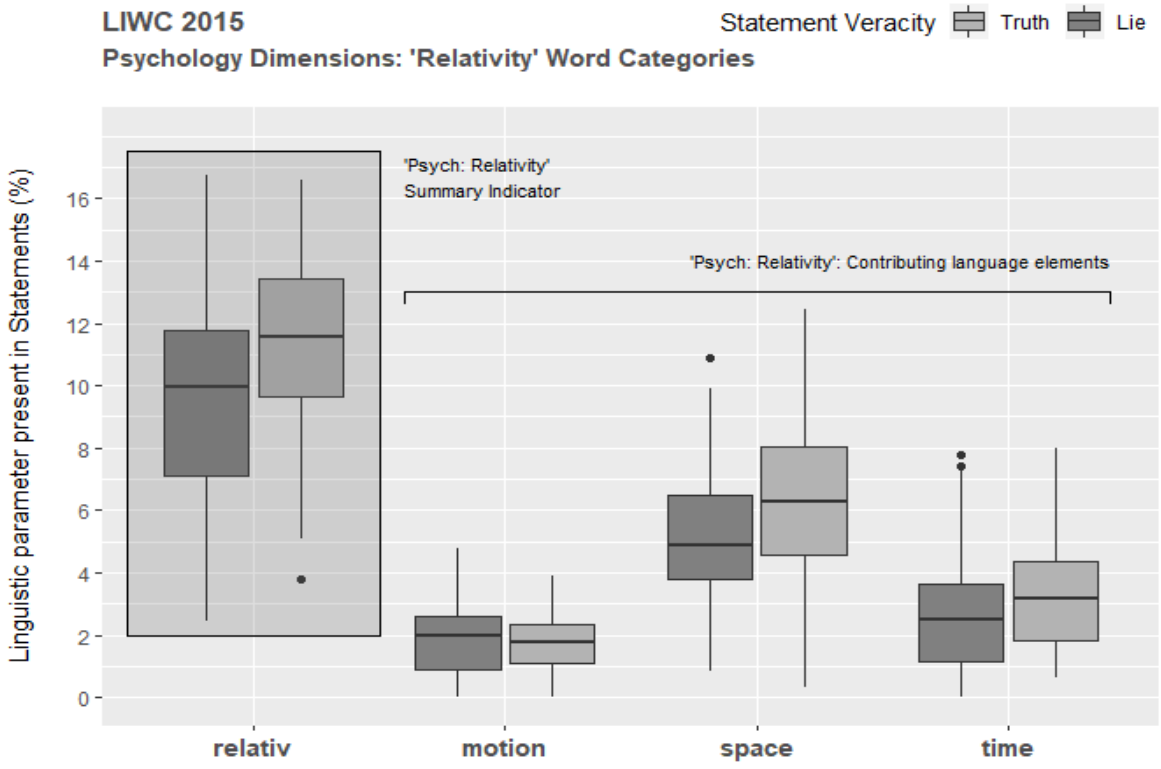

Figure 33 LIWC2015 Summary: 'Relative' Word Categories

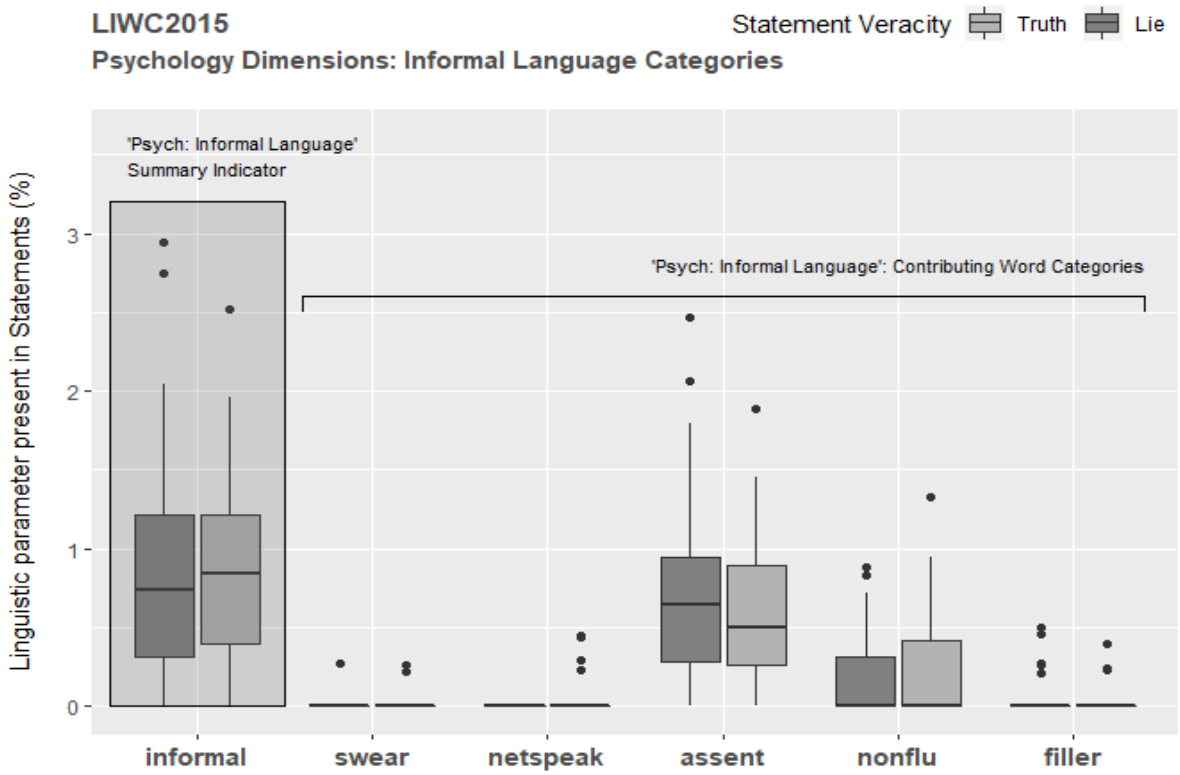

Figure 35 LIWC2015 Summary: Informal Language Categories

Appendix V: LIWC (language grouped) word category summaries

LIWC word categories presented as grouped by statements provided in a first or second language viewed against the presence (or not) of cognitive load.

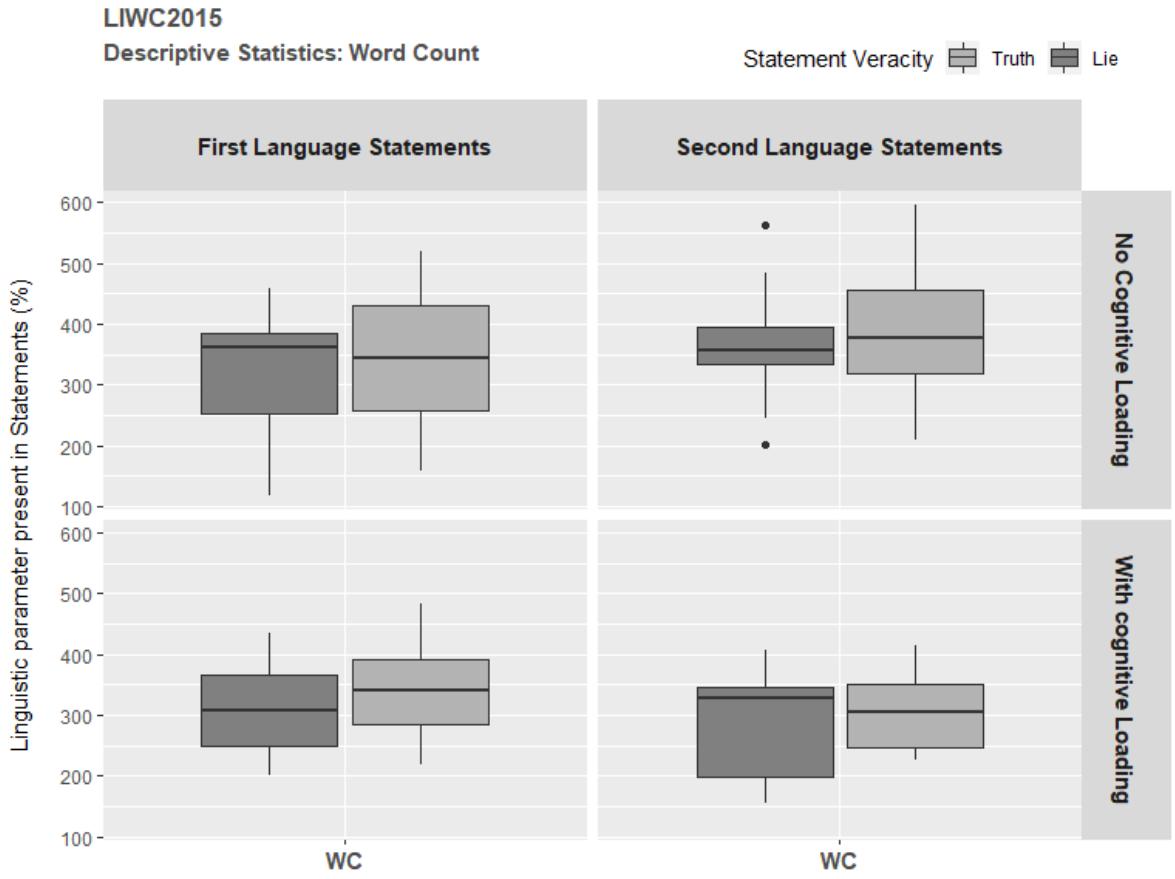

Figure 36 LIWC2015 (2x2 languages): Word Count

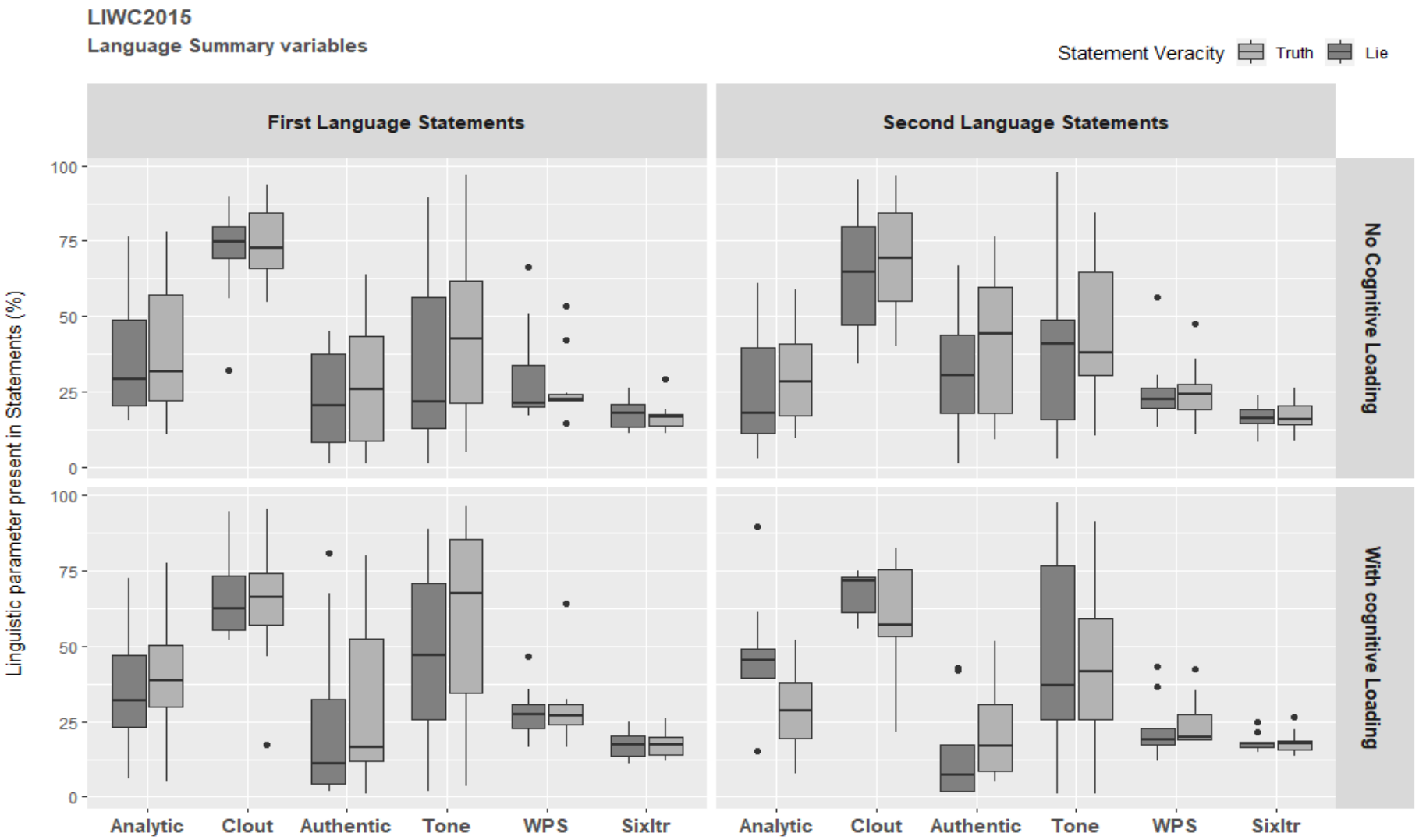

Figure 37 LIWC2015 (2x2 languages): Language Attribute Summaries

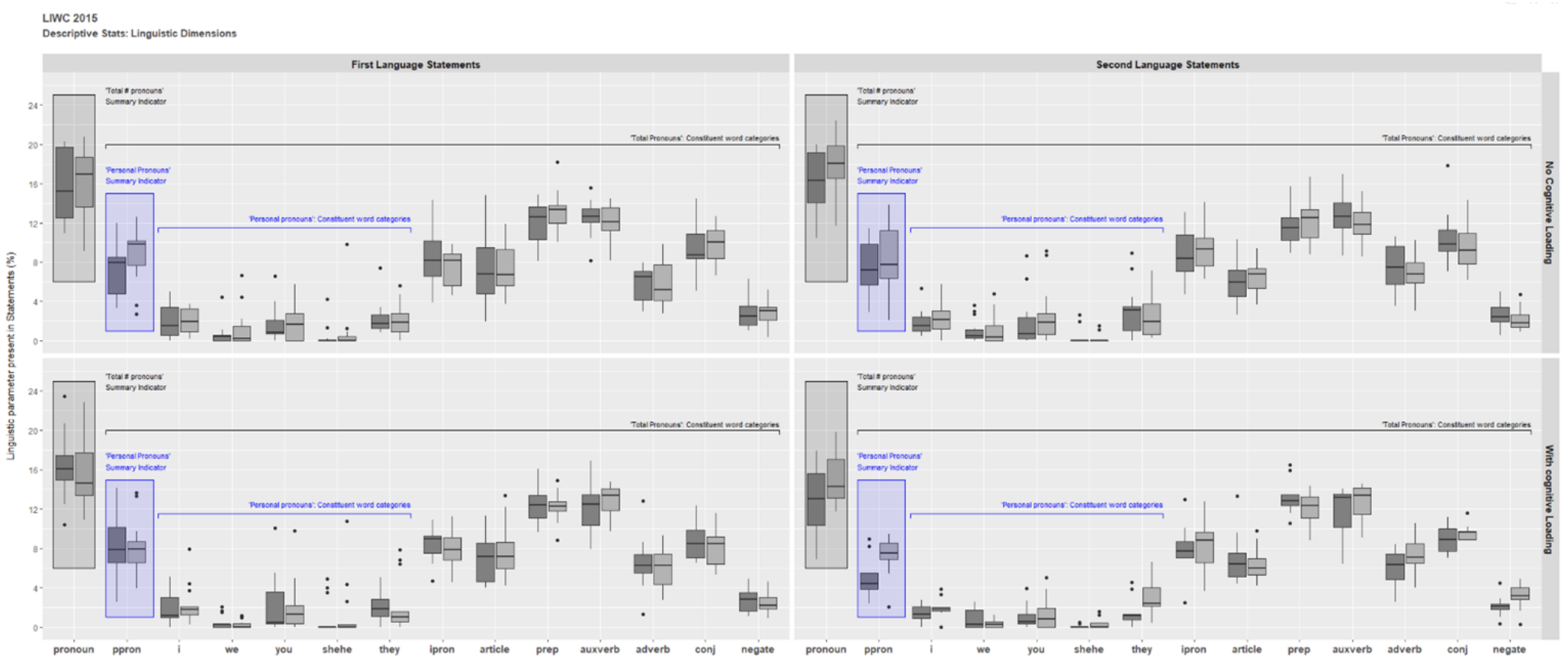

Figure 38 LIWC2015 (2x2 languages): Linguistic Dimensions

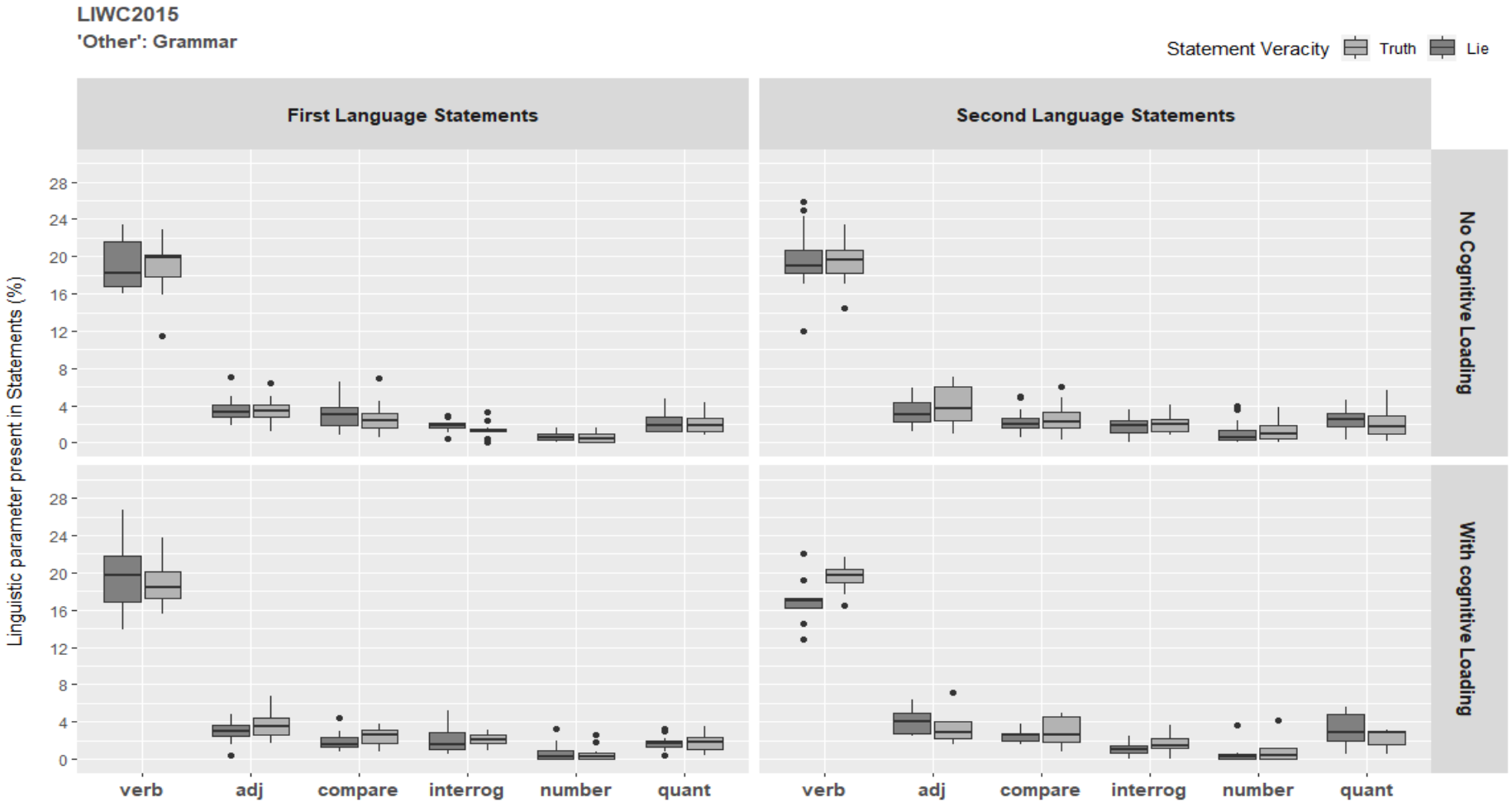

Figure 39 LIWC2015 (2x2 languages): Grammar Word Categories

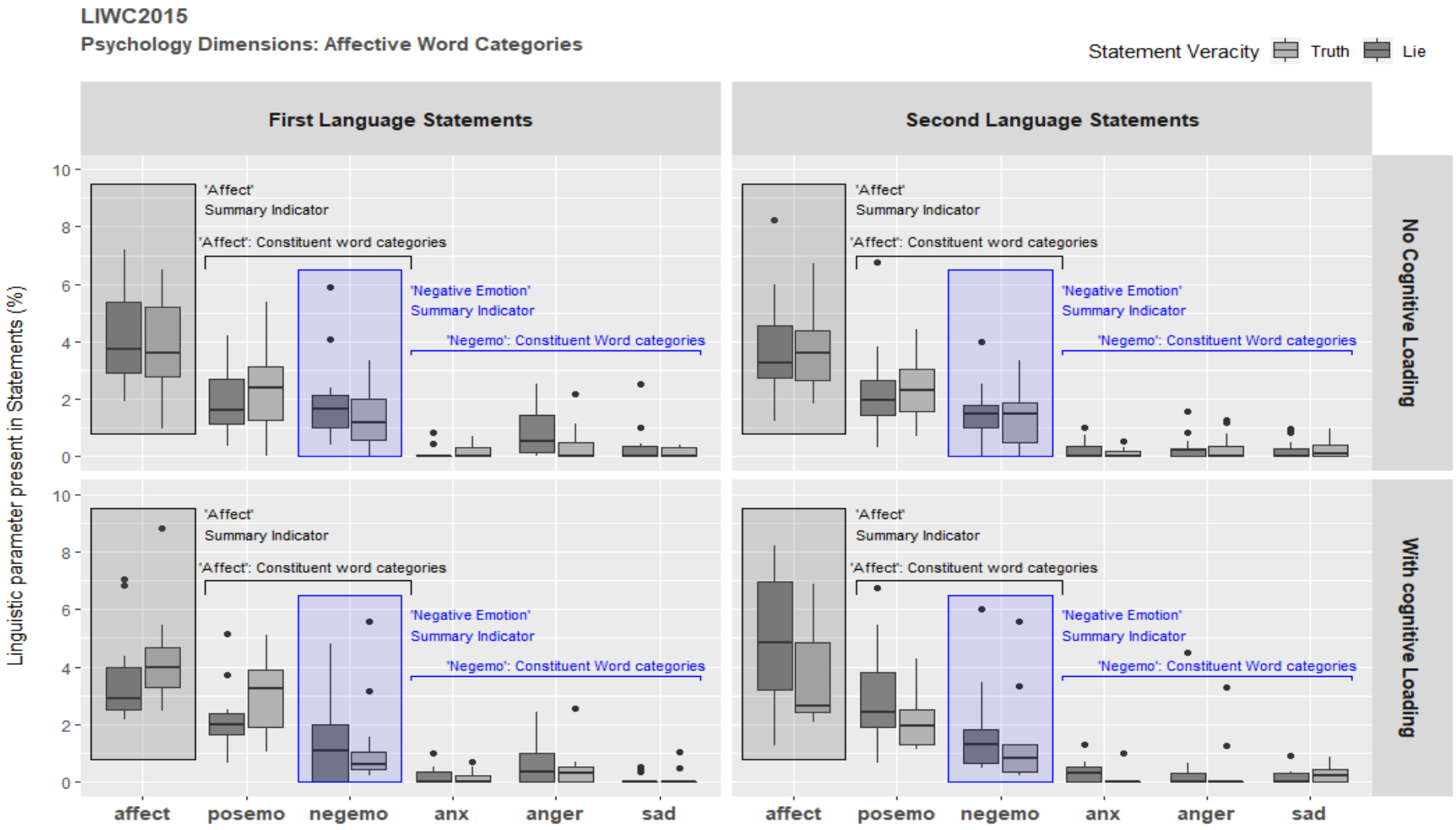

Figure 40 LIWC2015 (2x2 languages): 'Affective' Word Categories

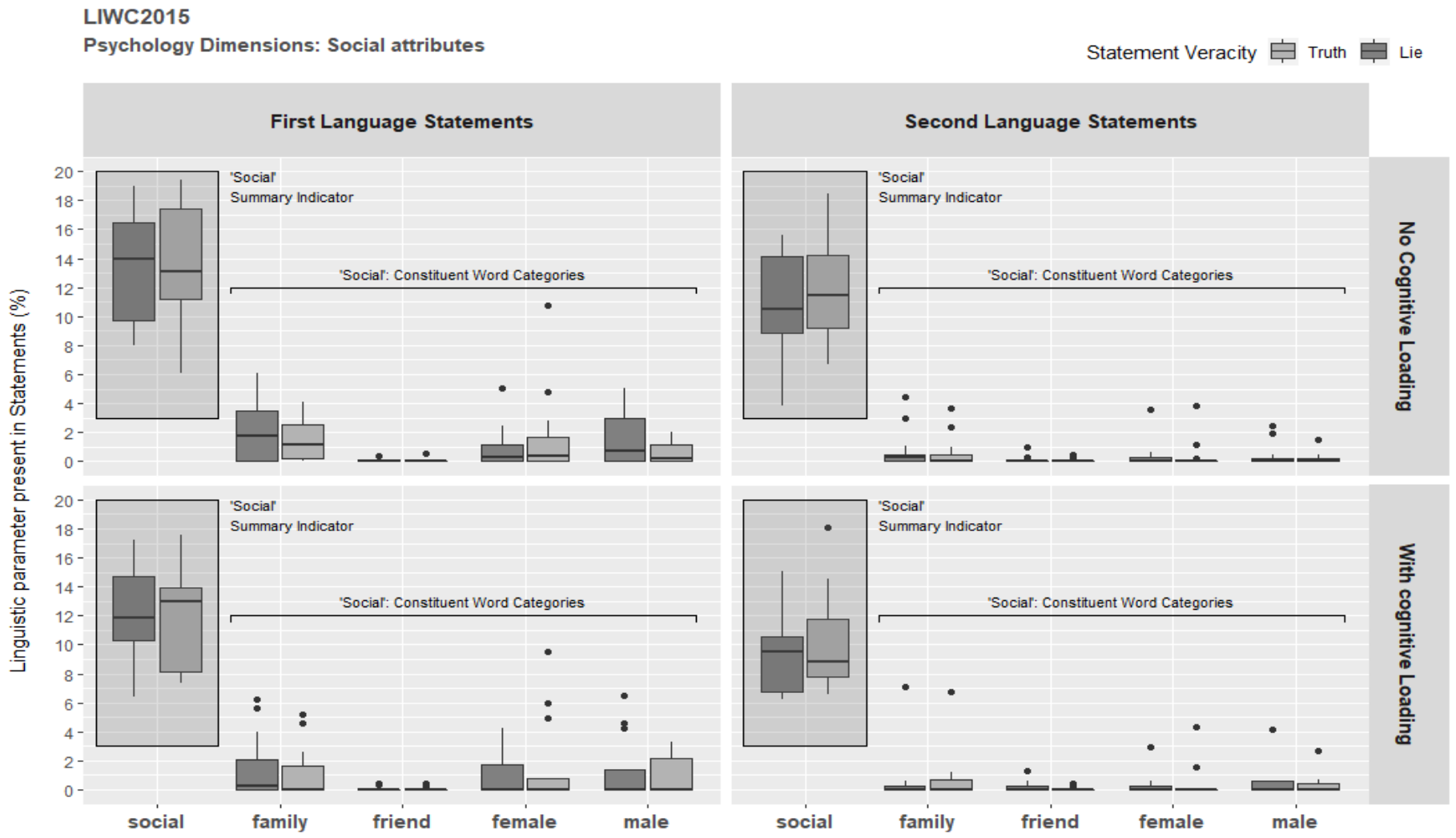

Figure 41 LIWC2015 (2x2 languages): 'Social Attributes' Word Categories

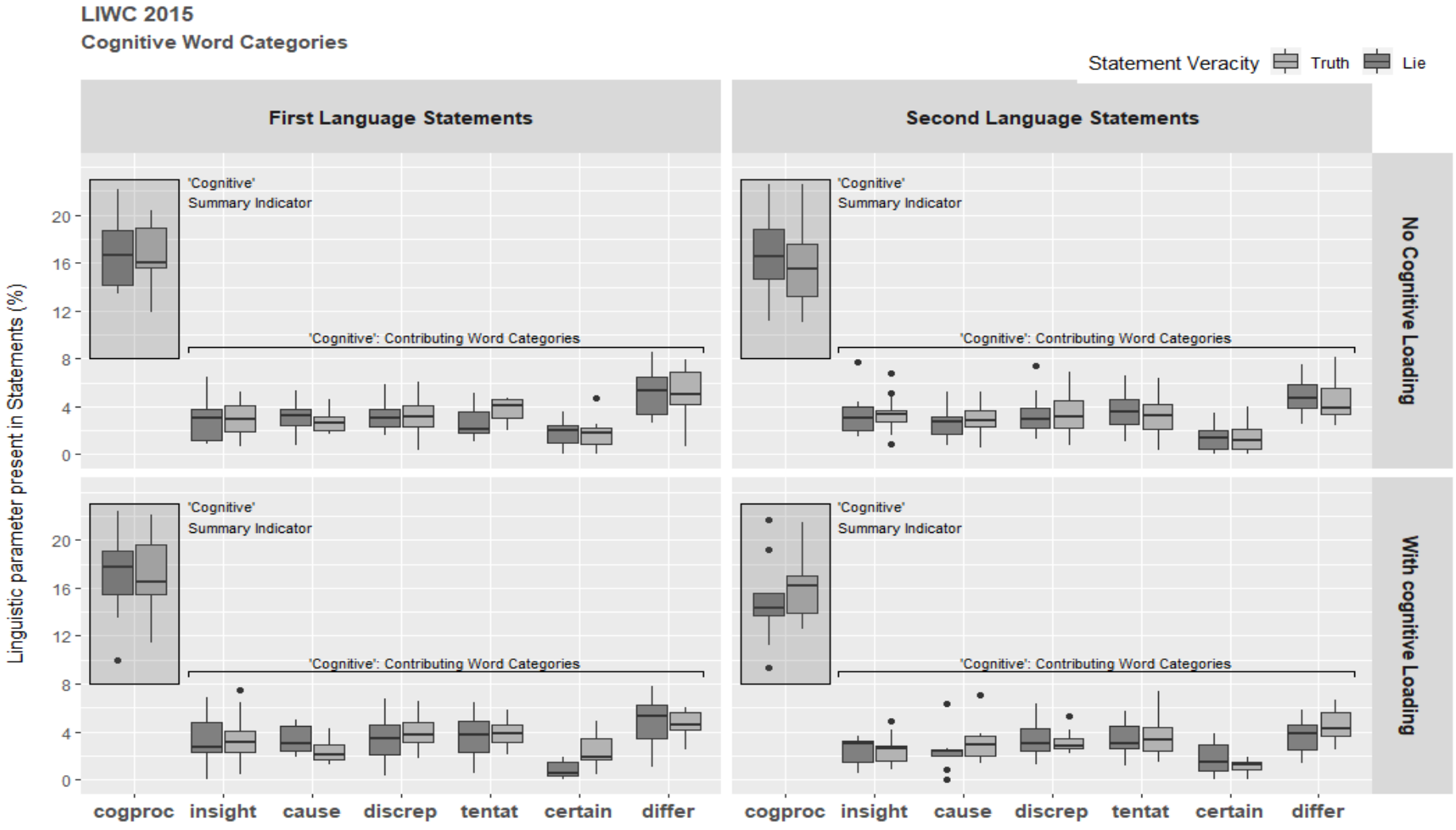

Figure 42 LIWC2015 (2x2 languages): Cognitive Attributes' Word Categories

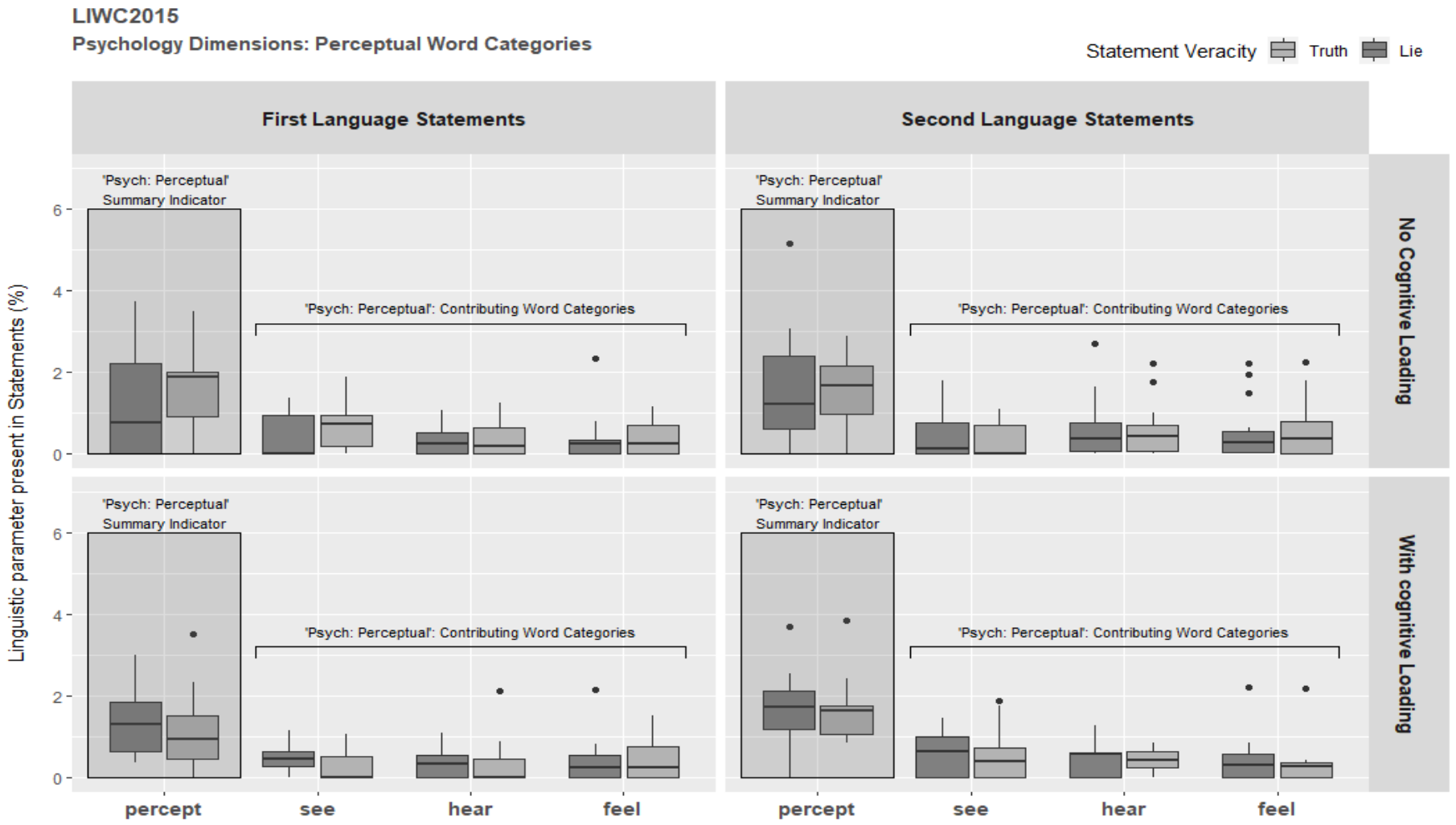

Figure 43 LIWC2015 (2x2 languages): 'Perception' Word Categories

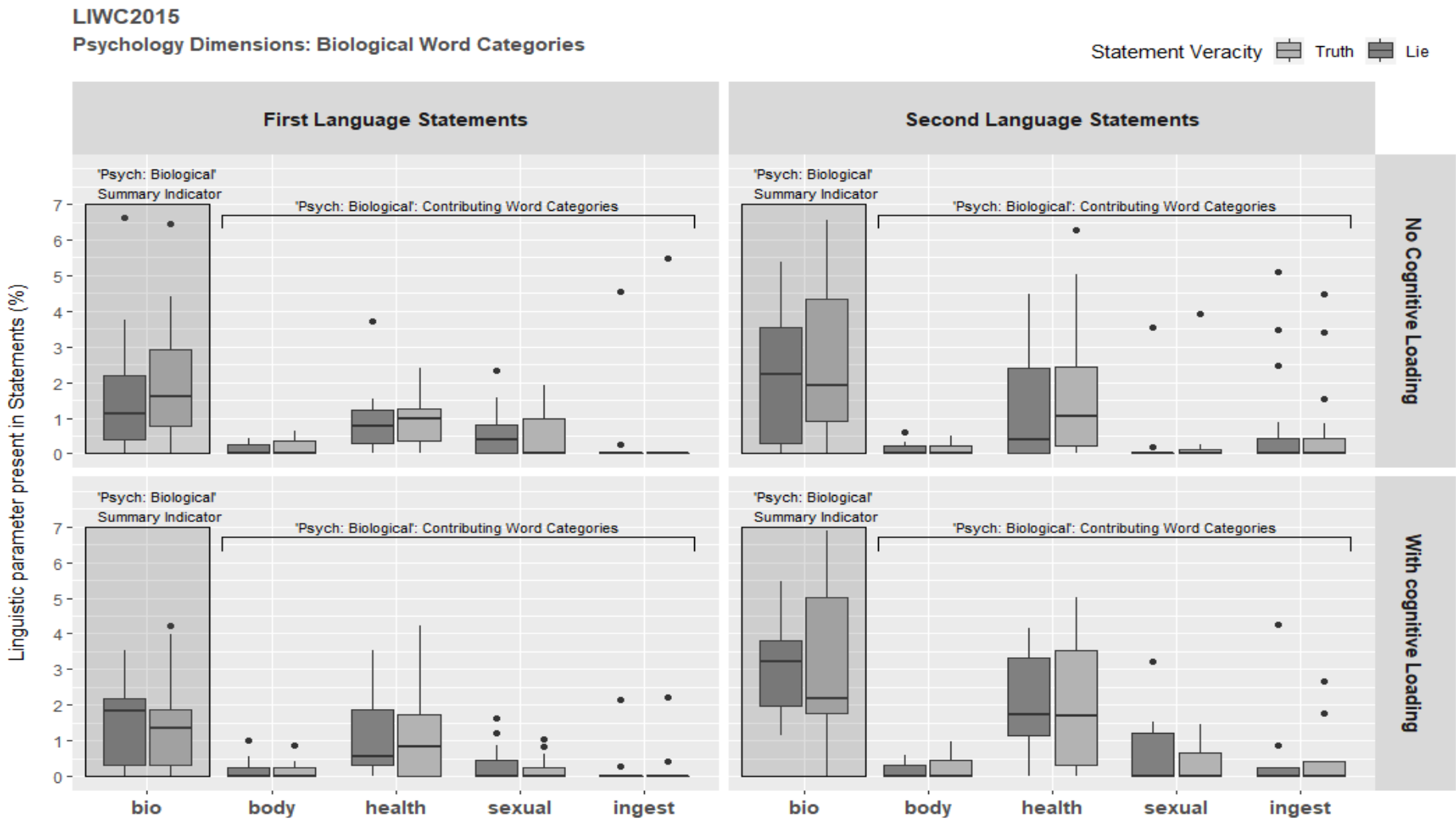

Figure 44 LIWC2015 (2x2 languages): Biological Word Categories

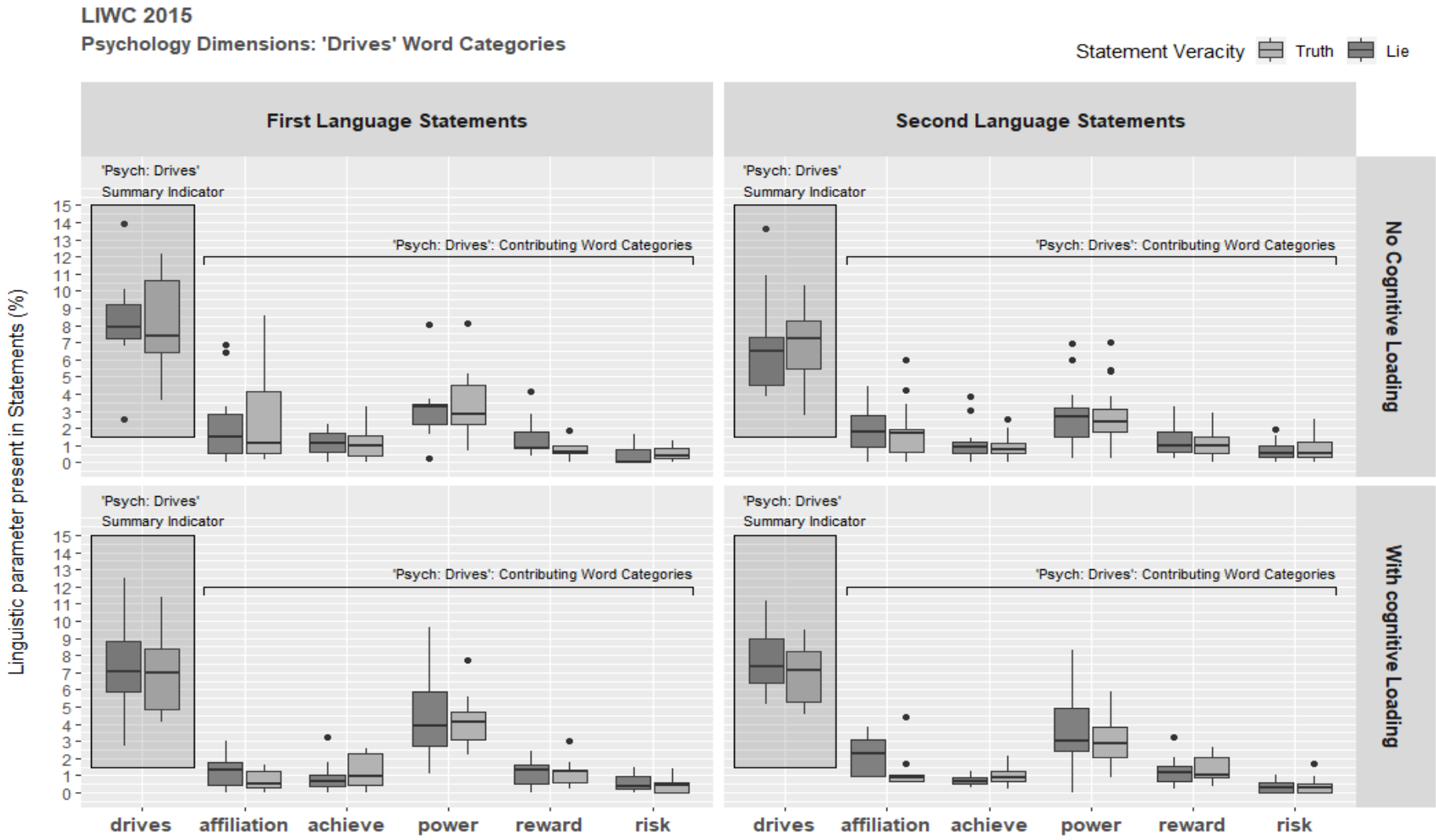

Figure 45 LIWC2015 (2x2 languages): 'Drives' Word Categories

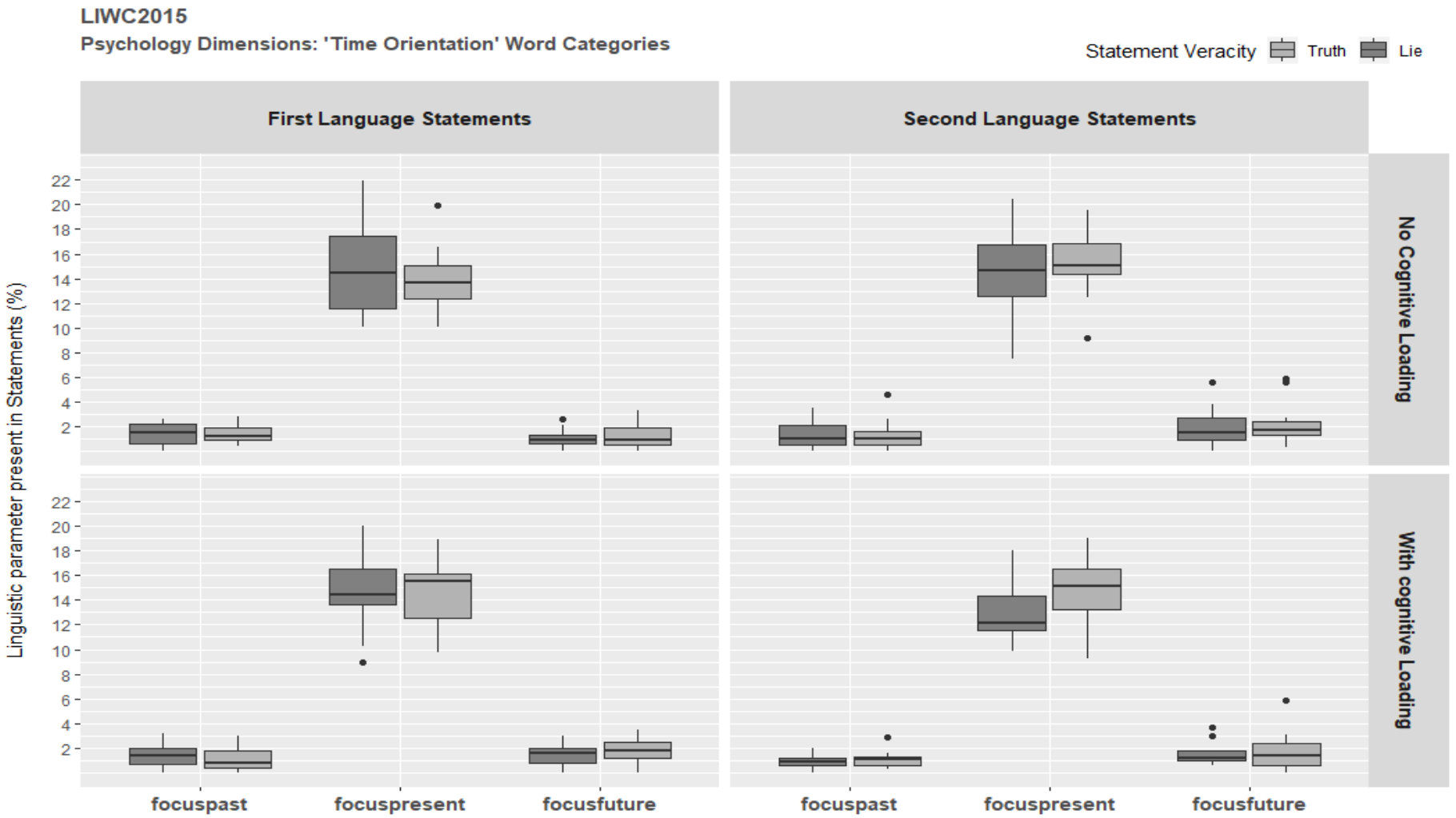

Figure 46 LIWC2015 (2x2 languages): 'Time Orientation' Word Categories

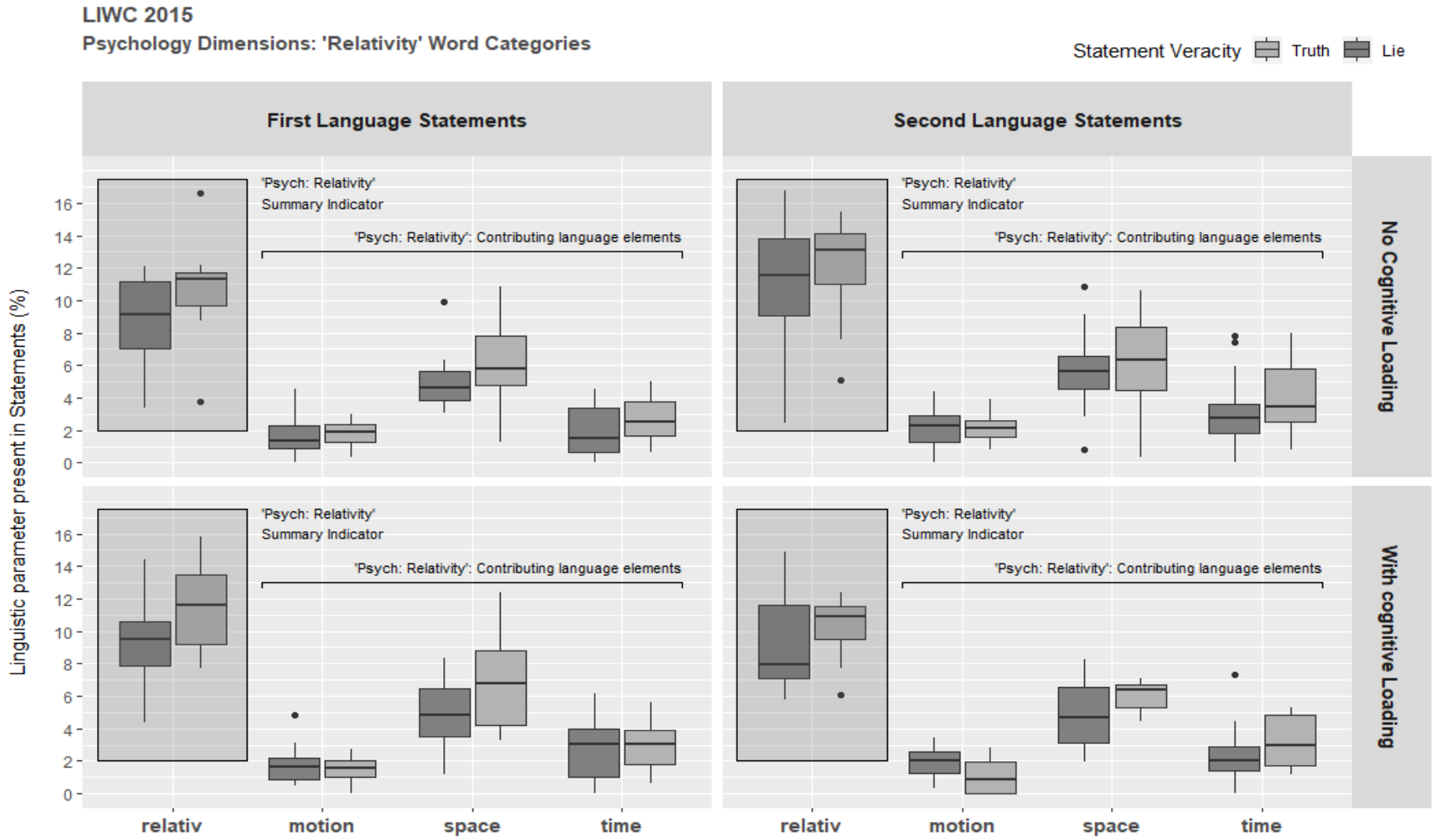

Figure 47 LIWC2015 (2x2 languages): 'Relative' Word Categories

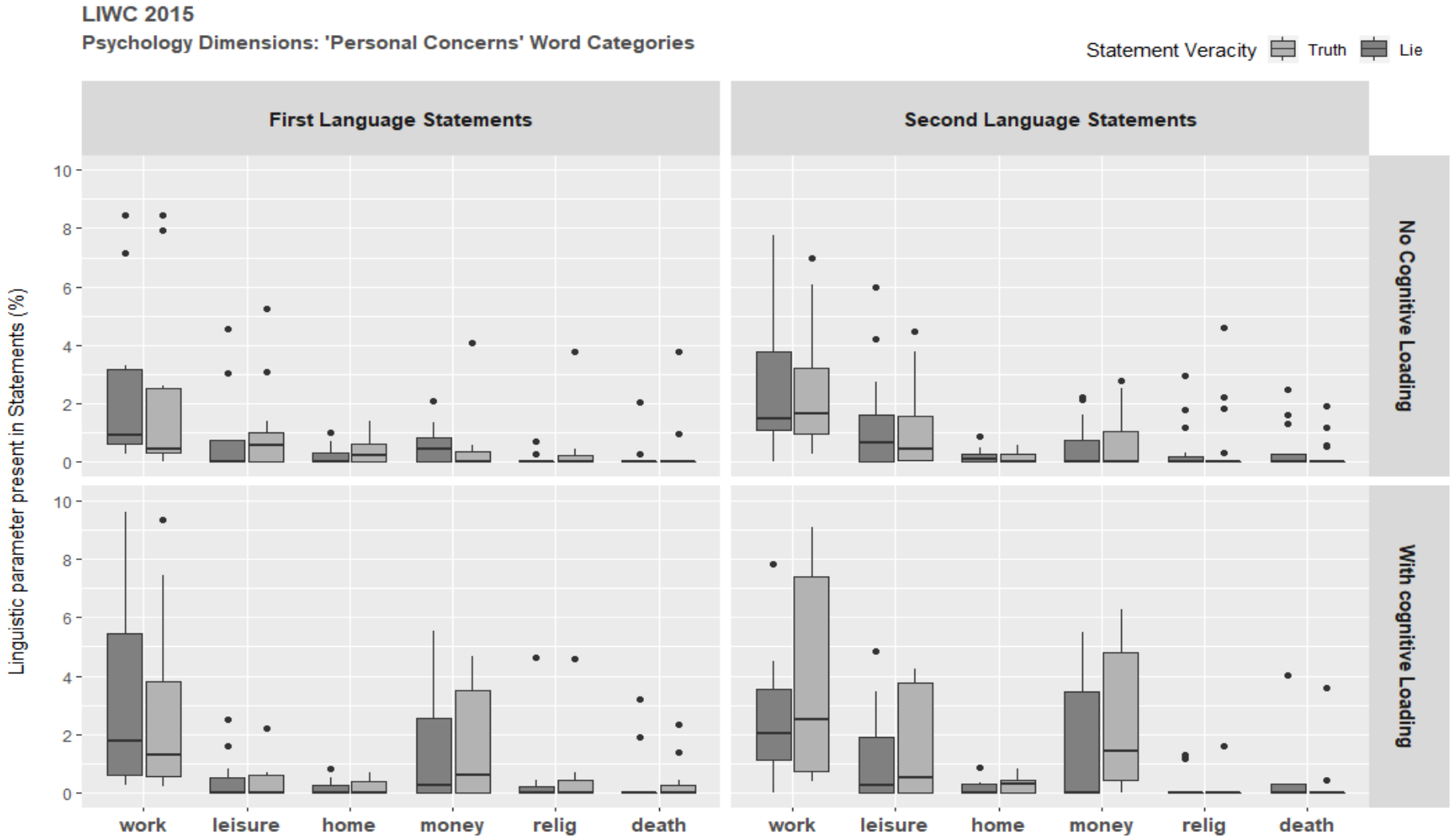

Figure 48 LIWC2015 (2x2 languages): 'Personal Concerns' Word Categories

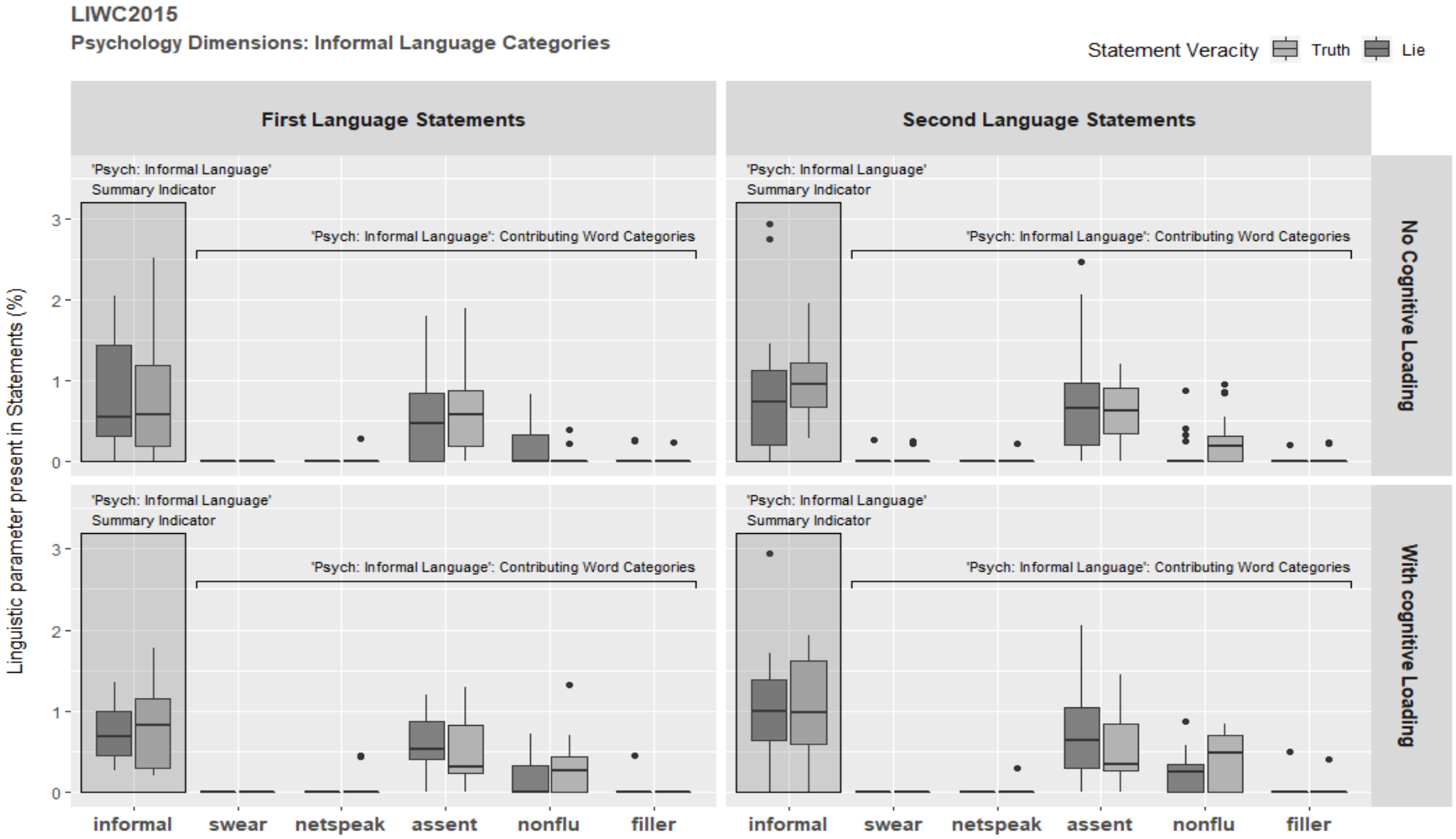

Figure 49 LIWC2015 (2x2 languages): Informal Language Word Categories