Social Contexts for Detecting Deception:

Factors that Moderate the Effectiveness of Cognitive Load Approach Interviews

Elise Fenn

A Dissertation submitted to the Faculty of Claremont Graduate University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology

Claremont, California

2015

Approved by:

_____

Kathy Pezdek, Ph.D.

Claremont Graduate University

APPROVAL OF THE REVIEW COMMITTEE

Text will be inserted once dissertation is approved.

Kathy Pezdek
Claremont Graduate University
Professor of Psychology

Dale E. Berger
Claremont Graduate University
Professor of Psychology

Mark Costanzo
Claremont McKenna College
Professor of Psychology

Iris Blandón-Gitlin
California State University Fullerton
Associate Professor of Psychology

Abstract

Social Contexts for Detecting Deception:
Factors that Moderate the Effectiveness of Cognitive Load Approach Interviews

By

Elise Fenn

Claremont Graduate University: 2015

Cognitive Load Approach interviews are designed to improve deception detection accuracy by making an interview more difficult for liars than truth-tellers. However, factors such as the social context of an interview may reduce the feasibility of this interview approach. Individuals who perceive they are stereotyped as criminals may experience stereotype threat when interviewed by someone of a different ethnicity, resulting in high cognitive demand for both liars and truth-tellers. The similarity in demand may reduce lie-truth discrimination accuracy. These predictions were tested across two experiments. In Experiment 1, Hispanics and Whites were randomly assigned as liars or truth-tellers. They completed a mock-crime and were interviewed by a White interviewer. In Experiment 1a, Hispanic interviewees experienced more stereotype threat, arousal and cognitive load than Whites regardless of actual veracity. In Experiment 1b observers viewed videos of Experiment 1a interviewees and rated the degree that interviewees displayed behavioral cues related to lying or truth-telling. In Experiment 1c, another set of observers watched videos from Experiment 1a and made deception judgments. In Experiment 1b and 1c, the magnitude of behavioral differences between liars and truth-tellers was more similar for Hispanics than Whites, reducing lie-truth discrimination accuracy to about chance levels for Hispanics (44% accuracy) but not Whites (61% accuracy). In Experiment 1d, observers' lie-truth accuracy depended on an interviewee's working memory capacity; observers were less

accurate detecting liars among White interviewees with higher working memory capacity, whereas observer discrimination accuracy was similar for Hispanics with high and low working memory capacity. Identical procedures as Experiment 1a-d were used in Experiment 2, but with a Hispanic interviewer. Hispanic interviewees reported experiencing more threat, but their arousal and cognitive load ratings did not significantly differ from Whites. Observers' ratings and lie-truth discrimination accuracy of interviewees did not differ significantly as a function of ethnicity or veracity. However, observers' lie-truth accuracy depended on the working memory capacity of interviewees; observers were more accurate discriminating between White interviewees with low than high working memory capacity whereas observer discrimination accuracy was similar for Hispanics with high and low working memory capacity. Implications of these results are discussed.

Acknowledgements

The amount of gratitude I am feeling is overwhelming. First, I am forever grateful to Kathy. Kathy, you are my research Mom. You've coached me into becoming a professional, a scholar, and an adult. I am so lucky and blessed to have been paired with you throughout my graduate career. Thank you for teaching me how to think and for always supporting me – even when I was idiot (e.g., leaving my statistics book in the bathroom way back on my first-ever statistics exam with you).

Iris, thank you for mentoring me into my graduate school years so selflessly. I am so thankful that we are now collaborators. I have learned so much from you. Thank you also for your belief in me as a researcher. I still remember that the first time I considered pursuing this career was when you told me that my work was "PhD-student quality". That moment changed me forever, and now I am here today!

Dale, I will forever cherish the meetings that we had together throughout graduate school, where I often found that some of your advice and wisdom about research extended much farther. You always inspired all of us as students to "go out and do good things," as you would say at the end of your classes. I make it a point to tell my students this same thing now. Thank you for the wisdom and guidance that you've provided me throughout this journey.

Mark, thank you for your support, advice, and council. I have learned so much from you in your classes. I truly appreciate that you have taken time to provide more advice and council throughout my dissertation process.

Mom and Dads, your support, understanding, and unconditional love are unreal. You've taught me how to be a real human being and reminded me everyday that you believe in me. Your "dadisms," nice supportive texts, emails and calls, days at Glenn Ivy, and wonderful family

holidays have gotten me through some tough times. You two are the best. Aaron, thank you for comforting and loving me when it must have been very challenging to do so. Thank you for respecting and appreciating my work; for celebrating with me when things went well and hugging me when things went poorly. Most importantly, thank you for understanding that my main way of coping with all of this is to engage in random outbursts of singing and/or dancing. Johnny, thank you for your advice and for listening when I had questions. Thank you for taking time to help me with many issues (statistical, personal, etc.) throughout these past years, even when you were slammed with your own work. You're the best big brother! Bruce and Esther, thank you for encouraging me and understanding when I had to miss Thursday night dinners. Family, you've understood when I've been absent (absent-minded or just plain not around!), and you've pushed me when I didn't believe I could keep going. I love you more than anything.

Stacia, my first-ever roommate and now life long best boo; thank you for your constant support and being my rock the first few years of graduate school especially. It's time to throw some glitter! Now we are entering our next step of this academic journey together, and I couldn't be more thankful to share these experiences with you. Sara and Mike, we are forever the cognitive warriors together. I am so thankful to have gone through all of these moments together with each of you. One day our theory of life will change the world. Mollie, I am so grateful for our fun times we've shared working together. You push me to work harder and have taught me so much about life. Susan and Emily, thank you for reminding me to have fun and enjoy life beyond academia. Thank you for reminding me what it means to be a good friend. Susan, thank you for being my Prince forever and for understanding my soul more than most. Aspen, thank you for all of your hard work on this project and beyond. Thank you for making this a fun and bearable project. You are an inspiration! To all my lab mates, I am so lucky to have formed

relationships with you all! Thank you for all the fun happy hours, and exciting research endeavors that we have tackled together.

To all of the staff and community and Claremont Graduate University and California State University Fullerton, thank you. Thank you especially to Linda Pillow for all of your helpful support. Thank you also to all of the research assistants involved in this project: Andrew, Steven, Stephen, Kassidy, Courtney, Amelia and Gabrielle. This work could not have been possible without your hard work.

TABLE OF CONTENTS:

Chapter 1: Introduction

Since 2007, the Transportation Security Administration has spent $900 million training airport security officers to spot liars (Tierney, 2014). Unfortunately, a 2013 review by the Government Office of Accountability reported no evidence supporting the effectiveness of this training. In fact, current interview and interrogation techniques used by professionals are largely unsuccessful in obtaining information diagnostic of deception (Meissner, Russano, & Narchet, 2010). The fundamental problem is that human judges are typically not accurate discriminating between truths and lies (Bond & DePaulo, 2006). Even professionals "lie catchers," such as law enforcement officials, typically do not exceed chance levels (54%) in detecting deception (Bond & DePaulo, 2006; 2008; but see Whelan, Wagstaff, & Wheatcroft, 2015).

In recent years researchers have increased efforts to investigate ways to improve interviewing and the detection of deception. One proposal that has been advocated is to avoid arousal-based approaches and apply an information-gathering approach that induces greater mental demand on deceptive interviewees (Vrij & Granhag, 2012a). Cognitive Load Approaches are a new set of interview techniques based on the information-gathering approach (Vrij, Mann, Granhag, & Leal, 2011). This approach is based on the premise that lying is cognitively more demanding than truth telling; therefore inducing greater load with interview techniques will be more detrimental to the liar than the truth-teller. This increased load is hypothesized to result in greater behavioral differences between truth-tellers and liars, differences that are more diagnostic of deception. Several studies have supported the feasibility of this approach (for a review see Vrij & Granhag 2012a or Walczyk, Igour, Todorov, & Dixon, 2013), its use in

intelligence interviewing settings has been advocated (Evans, Houston & Meissner, 2012; Lane & Vieira, 2012; Tedeschini, 2012), and it is being applied in some forensic settings (Vrij & Granhag, 2012b).

While research shows promise for using Cognitive Load Approaches, social-contextual factors that may play a role in its effectiveness have been largely ignored. The current research investigates a Cognitive Load Approach interview in contexts defined by the factors of ethnicity and cognitive capacity. In various forensic settings, including intelligence gathering contexts, individuals from diverse cultural backgrounds may be interrogated or interviewed by individuals from different cultural backgrounds (Kleinman, 2006). It is in these information-gathering situations that cultural interactions and individual difference factors may affect behaviors, moderating the effectiveness of new interview approaches including Cognitive Load Approaches. In light of the pressing need for theory-based approaches to improve interviewing and the detection of deception (Blandón-Gitlin, Fenn, Masip, & Yoo, 2014), the goal of this dissertation is to identify possible conditions that impact the effectiveness of these interviewing techniques.

Deception Detection Approaches

When detecting deception, both professionals and laypeople tend to focus on irrelevant behavioral cues (Bond & DePaulo, 2008; Vrij, 2008; see Hartwig & Bond, 2011 for a different conceptualization of this view). Most practitioners rely on the folk-wisdom that liars, but not truth-tellers, leak emotional cues that exude "deception." Professionals tend to believe that liars appear more nervous, more anxious, and more stressed than truth-tellers (Vrij, 2008; Vrij, Fisher, Mann, & Leal, 2006). This is consistent with the robust finding that there is a faulty "liar's stereotype" (Costanzo &

Krauss, 2014, p. 49). For example, Bond and DePaulo (2008) found that 70% of a sample of more than 2,500 people believed that liars averted their gaze and stuttered more than truth-tellers. In reality, the behavioral differences between liars and truth-tellers are minimal under most conditions (DePaulo et al., 2003). Liars and truth-tellers tend to appear equally anxious, nervous, and stressed while liars frequently exhibit behaviors contrary to lay beliefs (e.g., less overall movement).

Professionals' incorrect beliefs regarding deceptive behaviors have led to the use of ineffective interview techniques. For example, one popular technique used by many American police departments -- the Reid technique -- uses the Behavioral Analysis Interview (BAI) to identify guilty suspects who are lying (Inbau, Reid, Buckley, & Jayne, 2001, as cited in Kassin & Gudjonsson, 2004). Interviewers using the BAI ask 15 behavior-provoking questions designed to elicit different reactions from guilty (deceptive) and innocent (truthful) suspects in an attempt to highlight behavioral cues diagnostic of guilt or deception. Examples of questions include, "Did you commit the crime?" and "Why do you think someone committed that crime?" (see Masip, Barba, & Herrero, 2012) However, several research assessments suggest that the BAI does not produce significant behavioral differences between liars and truth-tellers (e.g., Vrij, Mann, & Fisher, 2006). Further, the results of Vrij, Mann, and Fisher suggest that reactions elicited from liars during a BAI tend to be the opposite of the predicted outcome from the BAI manual (e.g., Inbau, Reid, Buckley & Jane, 2005); for example truth-tellers were rated as displaying more behavioral cues of nervousness than liars, such as shifting posture in their chair, (Cohen's d = .81). Masip and his collaborators (Masip, et al., 2012; Masip & Herrero, 2013; 2014; Masip, Herrero, Garrido, & Barba, 2011)

reported that the BAI's proposed cues to guilt and innocence are commonsense behavioral stereotypes. In addition, liars' strategies are based on these stereotypic notions. Techniques emphasizing emotion-leakage theories, such as the BAI, encourage interviewers to rely on stereotypical and inaccurate deception detection strategies. A narrow focus such as this has no doubt contributed to the low accuracy rate of 54% for detecting deception that has been reported in comprehensive meta-analyses (Bond & DePaulo, 2006; 2008).

One recent interview technique designed to amplify behavioral differences between truth-tellers and liars involves an approach that taxes the executive functioning resources of interviewees. These approaches are often referred to as Cognitive Load Approaches (Vrij et al., 2006; Vrij, Granhag, Mann & Leal, 2011). Several studies have empirically validated these approaches (for reviews see Walczyk et al., 2013 and Vrij & Granhag 2012a). For example, Vrij et al. (2008) had liars and truth-tellers describe an event in sequential or reverse order. Describing an event in reverse chronological order was predicted to place higher cognitive demands on the interviewees than describing an event in sequential order, especially those who were lying. In fact, as predicted, liars were more accurately detected in the cognitively demanding, reverse order condition (60% lie detection accuracy) than the sequential order condition (42% lie detection accuracy). Police observers rated liars in the cognitively demanding interview as "thinking harder," (d = 3.06) and "looking (more) nervous" (d = 2.56) in their responses than truth tellers, whereas there was no significant difference between liars and truth-tellers on these behaviors in the less demanding interview ("thinking harder": d = .46; "nervous" d = .00). Other studies have used other cognitively demanding techniques such as asking

unanticipated questions during interviews (Vrij et al., 2009), having interviewees keep eye gaze fixed on the interviewers (Vrij, Mann, Leal, & Fisher, 2010), or having participants recall an event in their second, non-native language (Evans, Michael, Meissner, & Brandon, 2013). Across all these studies, when compared to a non-demanding interview, liars showed more discernable signs of cognitive load and deception.

Underlying Mechanisms in Deception

Although recent research suggests that Cognitive Load Approaches elicit greater differences between truth-tellers and liars, the underlying mechanisms and the conditions that moderate the effectiveness of these approaches need to be investigated. Understanding these mechanisms will help researchers improve the technique, increasing the likelihood of successful field application.

Several lines of research suggest that lying involves a greater reliance on executive functioning resources than truth telling. Executive functioning resources include inhibition, attention, working memory, and control processes such as suppressing irrelevant thoughts or information, attending to salient or task-relevant information, and manipulating and monitoring incoming information. In a review of the literature on cognitive correlates of deception, Gombos (2006) concluded that the executive control processes of inhibition and working memory are central mechanisms in deception, and as they become limited, the lying task becomes more difficult. Similarly, in a meta-analysis of brain imaging studies of deception, Christ et al. (2009) reported that deception is associated with increased brain activity in areas particularly involved in task-switching, working memory, and inhibitory processes such as the dorsolateral prefrontal cortex

(DLPFC) and the anterior cingulate cortex (ACC). This meta-analysis did not examine effect sizes, but focused on the regions of interest (ROIs) that were consistently activated more in liars than truth-tellers across studies. In addition, Gamer, Bauermann, Stoeter, and Vosse (2007) found covariation between brain activity, autonomic response (skin conductance), and overt behaviors (reaction times) in response to deception. The main finding was that brain regions associated with executive processes were more active during deception than truth-telling conditions and this activity was correlated with slower reaction times and increased skin conductance. The Gamer et al. study was one of the first to show a link between the behavioral, physiological, and imaging data, providing a more complete picture of the physiological and psychological processes associated with deception. Results of meta-analyses examining cues to deception (verbal and nonverbal) also support the hypothesis that lying typically involves more cognitive resources than truth-telling (Bond & DePaulo, 2006, Hauch, Blandón-Gitlin, Masp, & Sporer, 2014). For example, Hauch et al. investigated verbal content of statements and found that across 44 studies, liars' statements were less complex (as measured by "content word complexity": $d = .44$), shorter (as measured by "word quantity": $d = .24$), and less elaborated than truth-tellers' statements (as measured by "verb quantity": $d = .48$). These cues are believed to be associated with a simpler story, an outcome partly due to the experience of cognitive load (Hauch et al., 2014)

Further, tasks that interfere directly with executive control processes tend to reduce a liar's ability to appear truthful (Debey, Verschuevere, & Crombez, 2012; Hu, Evans, Wu, Lee, & Foo, 2013; Visu-Petru, Varga, Miglea, & Visu-Petra, 2013). For example, performing an interfering secondary task (e.g., a dot-probe task) while

answering questions truthfully or deceptively produced greater discrimination accuracy, as measured by the signal detection measure area under the curve (AUC = .94) between liars and truth-tellers than a control condition (AUC = .88) (Hu et al., 2013). In another study, liars showed reduced control over their behaviors after completing a task that required executive functioning resources (i.e., the Attention Network Task, ANT) (Blandón-Gitlin, Echon, & Pineda, 2013). That is, liars who performed a task that depleted executive functioning resources prior to lying were more easily detected (lie detection accuracy: 64%) than liars who did not complete the task (lie detection accuracy: 27%). Debey, Verschuere, and Crombez (2012) manipulated the length of the response stimulus interval (RSI) between interview questions to affect liars' and truth-tellers' access to attentional resources. For liars more than truth-tellers, longer RSIs interfered with attentional resources more than shorter RSIs. That is there was a larger increase in the percentage of errors (d = .42) and a larger increase in reaction times (d = .26) from shorter to longer RSI trials for liars more than for truth-tellers. The results of these behavioral studies suggest that executive processes are central cognitive components involved in deception, and disrupting access to these processes increases the cognitive costs associated with lying.

The findings from behavioral and neuroimaging research suggest that lying is often more cognitively demanding than truth telling. One interpretation of these findings is that liars' more than truth-tellers must perform multiple tasks that rely on executive functioning resources such as working memory (Walczyk, Mahoney, Doverspike, & Griffith-Ross, 2009). Truth telling involves the retrieval of an experienced event. Although this typically involves reconstruction, recall of episodic memories is generally

automatic. Access to and use of existing memory traces can enhance the fluency of information and make it easier for the truth-teller than liar to provide an account, even if cognitive demand is increased during interviews. In contrast, when telling a successful lie an individual must keep the truth from being revealed by conveying a plausible narrative, controlling body movements, monitoring for errors, and tracking the observers' demeanor. These sorts of tasks are typically controlled by working memory, which allows an individual to monitor multiple tasks simultaneously such as the integration and encoding of incoming information, planning for future actions, and executing decision-making processes (Baddeley, 1992; 2003; 2012). Because working memory is a system of "limited capacity," taxing or overloading this system hinders an individual's ability to monitor other aspects of executive functioning such as inhibitory control and attention.

Conditions that Affect the Cognitive Costs of Telling a Truth or Lie

Cognitive Load Approaches make the already demanding process of lying even more taxing by placing demands on liars cognitive processing systems, such as working memory (Vrij et al., 2011a). The main assumption of this technique is that lying is more cognitively taxing than truth telling, and indeed, research on the underlying mechanisms of deception support this assumption (e.g., Christ et al., 2009). However, an interesting question is whether there are conditions that moderate this hypothesized relationship between cognitive load and ability to deceive. An emerging line of research suggests that the answer is "yes;" under certain conditions the difficulty of a lie is modified.

Vercheuere and colleagues hypothesized that the cognitive cost associated with lying and truth telling is malleable (Verschuere, Spruyt, Meijer, & Otgaar, 2011). There are four possible conditions that may affect the cognitive costs of lying or truth telling.

First, results of several studies suggest that practicing lies can reduce the cognitive cost associated with lying (Blandón-Gitlin et al., 2013; Hu, Chen & Fu, 2012; Van Bockstaele et al., 2012; Verschuere et al., 2011). For example, Van Bockstaele et al. manipulated the frequency of lying – participants lied on 75% of trials (practiced liars), on 50% of trials (control group), or on 25% of trials (practiced truth-tellers). The more practiced liars made fewer errors and were faster than those who lied less frequently (on 25% of trials). Neuroimaging results of at least one study further suggest that practicing a lie (a rehearsed lie) alters the cognitive processes involved in lying, reducing activation in brain areas correlated with online processing systems such as working memory, while simultaneously increasing activation in brain areas related to autobiographical memory (Ganis, Kosslyn, Stose, Thompson, & Yurgelun-Todd, 2003). Rehearsed lies thus reduced the cognitive differences between truth-tellers and liars (Ganis et al., 2003)

Second, the content and motivations behind telling a lie or the truth, and the strategies used to do so may impact the cognitive costs. Fenn, McGuire, Langben, and Blandón-Gitlin (2014) found that explaining an intention (e.g., tell me about what you intend to do tomorrow) was a cognitively demanding task for truth-tellers. Truth-tellers who explained their future intentions during a cognitively demanding, reverse order interview were more likely to be judged as liars (as indicated by observers' bias ratings; d = .40), and displayed more deceptive behaviors (overall cognitive demand: d = .34; anxiousness: d = .24) than those in a less cognitively demanding, sequential order interview (but see Suchotzki, Crombez, Smulders, Meijer, & Verschuere, 2015 for a different conceptualization). Abe et al. (2014) found that the motivations behind a lie (whether lying for helpful or harmful reasons) impacted the neural substrates that

differentiated truth-telling and lying. Results of planned contrast tests suggest that lying about an event to avoid harmful consequences produced significantly greater brain activation in areas such as the right medial superior frontal gyrus (*Z* value = 3.56) and right temporoparietal junction (*Z* value = 3.92) compared to truth-tellers whereas there were no significant differences in activation of these areas between liars and truth-tellers in the helpful consequences condition (no statistical values reported). The type of lie appears to impact the cognitive costs and neural substrates involved in deception and truth-telling.

The strategy used by liars and truth-tellers may also impact the effectiveness of a cognitive load interview. Studies that have examined strategy use of liars have found that both liars and truth-tellers report using similar impression management strategies (e.g., avoid fidgeting, avoid looking nervous, etc.), whereas liars use more information management strategies than truth-tellers (e.g., staying consistent in their statement, avoiding incriminating details, "keeping (their statement) simple," etc.) (Hartwig, Granhag, Strömwall, & Doering, 2010; Hines et al., 2010; Strömwall, Hartwig, & Granhag, 2006). Liars also tend to report preparing for interviews more than truth-tellers, and tend to report using a larger number of strategies than truth-tellers (e.g., Masip & Herrero, 2013). Cognitive Load Approaches that aim to reduce liars' access to these strategies may improve lie detection accuracy.

Next, it may be easier than anticipated for liars to use countermeasures that reduce their executive functioning demands during Cognitive Load Approach interviews. Two possible types of countermeasures include rehearsing responses to anticipated interview questions, and choosing to perform a task during the interview that enhances executive

control. For example, the unanticipated questioning technique (e.g., asking participants to describe the spatial layout or temporal order of a target event, or to compose a drawing of the target event) is an important approach in interrogation settings because liars are less likely to anticipate and rehearse these questions than more general questions (e.g., "please tell me about what happened at dinner"). As a consequence, liars will have difficulty formulating a response to these questions during an interview, reducing their available executive functioning resources. However, a liar who is aware of these questions could prepare in advance (e.g., practice drawing a picture of their lie story that includes many details). This would make 'unanticipated' questions more anticipated, reducing a liars difficulty of formulating a response.

Additionally, countermeasures that facilitate liars' executive functioning abilities may also affect the cognitive cost of deception. For example, results of one study suggest that performing two simultaneous tasks that required access to inhibitory systems improved liars' ability to deceive (Fenn, Blandón-Gitlin, Coons, Echon, & Pineda, under review). Liars who performed a simultaneous task that required a high degree of inhibition (i.e., inhibiting urination urgency to a higher degree), displayed less deceptive behaviors, more truthful behaviors, and were detected at lower rates than liars performing a task that required a low degree of inhibition (i.e., inhibiting urination urgency to a lower degree).[1] These results align with observed phenomena in basic cognitive research

[1] It is expected that manipulating urination urgency may reach a ceiling threshold for participants to have control over their bladder, reducing the appearance of an ISE. That is, someone who has to urinate with extreme urgency (e.g., their bladder contains a physically unmanageable amount of fluid) may become consumed with thoughts about urinating, losing focus on the primary task (e.g., convincing the interviewer of innocence). The results of Fenn et al. (under review) and Tuk et al. (2011) suggest that when urination urgency was manipulated up to a level of higher urination urgency (i.e., the high control group that drank 750 mL of water and waited 45 minutes), inhibitory control ability was improved compared to a lower urination urgency (i.e. , the low control group that drank 500 mL of water and waited 45 minutes). All participants were initially told a cover story about the purpose of their participation, to hopefully avoid

suggesting that inhibition on a task from one domain (e.g., physiological bladder control) can "spill over" into other domains (e.g., behavioral control of deceptive appearance) (Berkman, Burklund, & Lieberman, 2009; Tuk, Trampe, & Warlop, 2011). More critically, these results provide intriguing evidence that it is possible for liars to reduce executive functioning demands by easy-to-manipulate countermeasures. Although it is likely that the observed increase in inhibitory control occurred without a person's conscious awareness, it is within the liars control to choose a countermeasure that would reduce the imposed cognitive load of a situation.

Finally, individual differences in executive functioning ability may affect the cognitive costs associated with lying. That is, liars with higher working memory capacity should find deception easier than liars with lower working memory capacity. Results of at least two studies support this assertion (Blandón-Gitlin, Arrieta, Gombos, Mayberry, & Gerkens, under review; Lane, Martin, Elliott, & Mennie, 2014). In both of these studies, observers were less accurate detecting liars with higher than lower WMC when under high cognitive demand. This new line of research suggests that the cognitive costs associated with lying and truth telling are malleable under certain conditions: when lies are practiced, when explaining a future intention, when countermeasures are used, and when individual differences in WMC are considered. As a consequence, these conditions may undermine the effectiveness of Cognitive Load Approach techniques.

Social context during interviews. One important and understudied variable that may also affect the cognitive difficulty associated with lying and truth telling is the social context of an interview. Evans, Meissner, Brandon, Russano and Kleinman (2010)

---

excessive focus on their urination urgency. Future studies should (cautiously) identify the threshold for using a bladder manipulation; examining whether the ISE disappears when participants are too consumed by thoughts of urination (although the ethics of such study are questionable!).

identified several interview and interrogation characteristics typical in a human intelligence-gathering context that need further investigation. Among these was a need to investigate the interviewer and interviewee relationship; namely, the influence of "a correspondence between interrogator and interviewee characteristics (including ethnicity, age, or gender)" (p. 238).

Based on a recent model by Najdowski (2011), it is predicted that the cross-ethnic relationship between the interviewer and interviewee will affect the diagnostic value of deception detection techniques such as a Cognitive Load Approach. According to Najdowski, minorities tend to appear more deceptive than non-minorities during interviews because of "stereotype threat." Stereotype threat, a term coined by Steele and Aronson (1995), refers to "being at risk of confirming, as self-characteristic, a negative stereotype about one's group" (p. 797). Stereotype threat applies to any stigmatized group. When the characteristics of social identity are more salient to a group (i.e., threatened condition), performance decrements are likely to be observed that ironically confirm the group stereotype. Examples of groups that experience stereotype threat are Whites with regard to appearing racist (Frantz, Cuddy, Burnestt, Ray & Hart, 2004), gay childcare workers with regard to appearing like child molesters (Bosson, Haymovitz & Pinnel, 2004), Blacks compared to Whites on intelligence tests (Steele & Aronson, 1995), and women compared to men on math assessments (Murphy, Steele, & Gross, 2007). Although Najdowski's focus was on Black minorities, she suggested that any group with stigmatized (negative) perceptions regarding criminality should experience stereotype threat when suspected of a crime. For example, (a) people from minority groups (e.g., Hispanics) with past negative police encounters, and (b) people from low

socioeconomic settings who tend to harbor negative attitudes toward law enforcement (Brown & Benedict, 2002). Thus, members of *any* stigmatized group that perceive a negative stereotype surrounding their relationship with criminality are likely to confirm this stereotype during an interview or interrogation, and this may be further exacerbated during interviews with someone of a different race (see also Davis & Leo, 2012).

The cognitive capacity of a stereotype-threatened individual, regardless of lie or truth status, is hypothesized to be reduced during an interview due to (a) excessive monitoring of the self and the interviewer and (b) increased physiological arousal (Schmader, Johns & Forbes, 2008). These predictions align with the predictions of Interpersonal Deception Theory, suggesting that during dynamic face-to-face interviews, behavior monitoring is an excessive burden for liars but not for truth-tellers (Buller & Burgoon, 1996). Other research on stereotype threat and evaluation anxiety supports these predictions. Experiencing stereotype threat reduces the availability of working memory resources, causing threatened persons to perform worse on tasks that rely on these resources (Beilock, Rydell, & McConnel 2007; Coy, O'Brien, Tabaczynski, Northern, & Carels, 2011; Murphy et al., 2007; Schamder et al.). For example, Beilock et al. manipulated stereotype threat by (a) informing women that their math performance on an exam would be evaluated against men's performance (threat condition), or (b) not providing these instructions (no threat condition). The math exam included easy or difficult problems that required verbal or visuo-spatial strategies to solve. These two types of math problems were chosen to examine the hypothesis that stereotype threat specifically interferes with the phonological loop of working memory, the system critical for storing and manipulating verbal information. Results confirmed their hypotheses.

When asked about their thoughts during the exam, women in the threat condition reported worrying about confirming their stereotype and monitoring their performance more than women in the non-threat condition, evidence that threat reduced the capacity of the phonological loop of working memory. As a result of this reduction, when compared to the no-threat condition, women in the threat condition were less accurate at solving difficult than easy math problems that required verbal strategies, but scored similarly on easy and difficult math problems that relied on visuo-spatial strategies. These results suggest that stereotype threat reduced task performance (e.g., accuracy on a math exam) by interfering with working memory resources (e.g., the phonological loop) necessary to complete a task (e.g., the verbal but not visuo-spatial math problems).

In addition, Trawalter, Richeson and Shelton (2009) have argued that for Whites and racial minorities, a dyadic, cross-cultural interaction is more threatening than a same-race interaction when negative racial stereotypes are primed. For example, Goff, Steele and Davies (2008) reported that when the stereotype of appearing racist was made salient (e.g., Whites appearing racist when discussing racial profiling), then Whites chose to distance themselves, or sit further away from a Black than a White conversational partner. More importantly, this difference in distance was significantly greater when White's explained their personal opinion compared to explaining another's opinion on racism. These results suggest that the context of the situation (explaining their own opinion versus another's opinion), interacted with interracial contact (same- or cross-race interview) to predict the level of threat experienced. In the proposed research, when the stereotype of criminality is made salient (e.g., after being accused of a crime), Hispanics

are predicted to experience more threat when interviewed by a White than Hispanic interviewer.

Further, Trawalter and Shapiro (2010) suggested that interracial contact can be more cognitively depleting than same-race contact for Whites and racial minorities due to excessive self-monitoring. For example, Richeson and Shelton (2003) reported that Whites who were prejudiced towards Blacks showed greater depletion of executive resources (lower scores on a Stroop task) after interacting with a Black than a White person. Threatened ethnic minorities may experience similar cognitive depletion during an interview with a member of a non-stigmatized ethnic group. According to Richeson and Shelton (2007), this is because minorities are often concerned about being the target of prejudice during interracial interactions, and as a consequence will engage in more self-regulatory behaviors. Based on these results, ethnic minorities who believe they are perceived as a criminal by the interviewer may have greater difficulty regulating their behaviors during a cross- than same-race interview designed to impose cognitive demands.

Several other studies have focused on differences in the nonverbal behaviors of interviewees from a stigmatized group (Vrij, 2008). Vrij and Winkel (1994) investigated differences in police officers' impressions of Dutch Caucasian and Black Surinamese actors. The actors lied or told the truth using nonverbal communication styles that mimicked either culture. Dutch Caucasian police officers rated interviewees as more suspicious when exhibiting Black than Caucasian nonverbal behaviors. Najdowski (2011) pointed out that some of the behavioral "symptoms" of stereotype threat such as gaze aversion are cues that law enforcement agents typically rely upon to determine deception

(e.g., the Reid technique). Thus, a suspect who is experiencing stereotype threat may be more likely to display cues to deception that officers are looking for regardless of actual guilt or innocence, leading to a possible increase in misidentifications of innocents. Whether Cognitive Load Approaches elicit more of these nonverbal cues in threatened truth-tellers and liars than those non-threatened, and how these conditions relate to detection accuracy are important questions that warrant investigation.

Based on the findings related to social contexts during interviewing (e.g., Najdowski, 2011; Richeson & Shelton, 2007; Vrij & Winkel, 1994), it is predicted that both liars and truth-tellers will experience more cognitive overload under stereotype threat than non-stereotype threat conditions. For non-stereotype threat conditions (in this study, White interviewees), a cognitive load approach interview is hypothesized to tax the executive functioning resources of liars more than truth-tellers. The Cognitive Load Approach interview questions, such as answering unanticipated questions or recalling an event in reverse order, are hypothesized to use more of liars' cognitive resources, but should not tax truth-tellers to the same degree. However, certain conditions such as stereotype threat may cause even a truth-teller to experience a reduction in cognitive resources. The application of a Cognitive Load Approach interview is predicted to add to the cognitive demands of interviewees, and consequently elicit weak diagnostic cues to deception at best, and at worst eliminate any differences between liars and truth-tellers. Investigating this potential constraint on the Cognitive Load Approach is one goal of the current research.

Rationale for the Proposed Research

The proposed research investigates factors that moderate the effectiveness of the Cognitive Load Approach. The underlying assumption of this approach is that liars will be more cognitively challenged than truth-tellers as information processing becomes more difficult during an information-gathering interview. This should amplify behavioral differences between liars and truth-tellers. This notion is supported by models of executive functioning suggesting that working memory is a finite resource that can be overloaded under certain conditions (Baddeley, 1992; 2003; 2012). Therefore, if the working memory capacity of both liars and truth-tellers is "tapped out," behavioral differences between the two are not likely to be revealed. Najdowski's (2011) model suggests that during interviewing, stigmatized ethnic minorities[2] would experience more cognitive load than non-stigmatized ethnic groups, especially when stereotypes are primed. Consistent with these models we predict that threatened persons (e.g., ethnic minorities under stereotype threat) will experience greater cognitive load during a cognitively demanding interview and therefore appear more deceptive than non-threatened persons regardless of the veracity of their accounts. Thus, the mechanisms that underlie stereotype threat, specifically increases in cognitive load and arousal, are

[2]Najdowski (2011) suggested that her model extends to "other groups who are stereotyped as criminals (e.g., Hispanics)," besides African Americans persons who identify as belonging to the Hispanic, Latino or Chicano culture tend to be stereotyped as criminals (Neimann, 2001; Mirande, 2008). Further, other results indicate that Hispanics are more likely to be convicted of a crime (perceived as guilty) than Whites (Espinoza & Willis-Esqueda, 2008; Mustard, 2001). Hispanics/Latinos/Chicanos, especially immigrants, tend to perceive law enforcement more negatively than do Whites, another potential factor leading to stereotype threat in criminal investigation contexts (Carter, 1983; Carter, 1985; Menjívar & Bejarano, 2004). This then provides the rationale for studying Hispanics, the population of participants selected for this research.

predicted to reduce the effectiveness of Cognitive Load Approaches; stereotype-threatened truth-tellers are likely to appear as deceptive as stereotype-threatened liars. The proposed research tests these predictions.

Overview of the Current Study and Hypotheses

A full list of independent and dependent variables, and the general design of each of four parts of the study is displayed in Table 1. The overall design includes the following independent variables: (a) Interviewee Ethnicity (White or Hispanic), (b) Interviewee Veracity Condition (truth or lie), and (c) Interviewee Working Memory Capacity (high or low). Experiment 1 (with White interviewer) and 2 (with Hispanic interviewer) (parts a through d) followed identical experimental procedures. In part a, participant-interviewees were assigned to lie or tell the truth during a Cognitive Load Approach interview about their involvement in a mock-crime scenario. These interviews were video-recorded and shown to two different samples of third-party observers (parts b and c). In part b, observers rated the interviewees on several behavioral cues related to truth (confident and convincing appearance) and deception (cognitive load and anxiety). In part c, a separate group of third-party observers made explicit veracity (deception) judgments by rating interviewees on how truthful or deceptive they appeared. In part d, observers' veracity ratings from parts c were examined as a function of interviewee's working memory capacity.

The principle predictions are based on comparing experiences for Hispanic and White interviewees in the White interviewer condition (Experiment 1). This is for two reasons. First, an applied forensic or intelligence interview is most likely to be conducted by a White interviewer.

Table 1. General Experimental Design for Experiments 1 and 2, separated by each sub-experiment (a, b, c, and d) and the corresponding participant pool, independent variables, and constructs measured (including tasks and dependent variables used to assess each construct).

Experiment and Participant Pool	Independent Variables	Construct	Task or Measure	Dependent Variable
A: Interviewee	Veracity: Lie or Truth Ethnicity: Hispanic or White	Stereotype Threat	15-item Memory Scale	Correct Response Rate[a]
			Interviewee Questionnaire	Self-Report experience of stereotype threat
		Cognitive Load	Auditory Probe Detection Task[b]	Reaction Time and Error Rate
			Interviewee Questionnaire	Self-Report of Mental Challenge
		Arousal	Physiological Measures	Heart Rate/Blood Pressure
			Interviewee Questionnaire	State-Trait Anxiety Index (STAI) Score
B: Third-Party Observers Sample 1[c]	Interviewee Veracity: Lie or Truth Interviewee Ethnicity: Hispanic or White	Detection Accuracy	1. Observer Ratings of Interviewee Veracity	1. Accuracy Discriminating Truths and Lies
C: Third-Party Observers Sample 2[c]	Interviewee Veracity: Lie or Truth Interviewee Ethnicity: Hispanic or White	Behavior Cues	1. General Behavior Assessment	1. Frequencies of target general behaviors
D: Third-Party Observers Sample 1 and 2	Interviewee Veracity: Lie or Truth Interviewee Ethnicity: Hispanic or White Interviewee Working Memory Capacity: High or Low[d]	Detection Accuracy	1. Observer Ratings from Experiment 1b and 1c	

[a]Participants scores on the 15-item memory scale are calculated as a percentage of items correctly recalled, based on the methods of Murphy et al. (2007).
[b]Auditory Probe Detection Task is the secondary task performed during the interview phase. The primary task for the interviewee is "convincing the interviewer you are telling the truth."
[c]Two independent samples of third-party observers participated in Experiments b (Experiment 1: N=184; Experiment 2: N=132), and c (Experiment 1: N=220; Experiment 2: N=134).
[d]Ospan score was calculated based on established method used in Unsworth et al., (2005).

Whites are the majority (75%) among national and local (42%)[3] law enforcement (Newton, 2006; Reaves, 2010; see also Cleary, 2014). Second, because Whites comprise the plurality nationally (63%) and locally (44%[4]), they may be perceived as more dominant to minority groups such as those who identify as having Hispanic origins (17% of U.S. population) (U.S. Census Bureau, 2014a; 2014b). Thus, it is predicted that this condition will elicit the highest levels of stereotype threat, making this the critical comparison condition for determining how stereotype threat and veracity interact in predicting the effectiveness of Cognitive Load Approach interviews. *Experiment 1a through d* focused on this prediction by using a White interviewer only.

Experiment 2a, b, c, and d followed identical procedures as Experiment 1, except that the interviewer was Hispanic. It was predicted that the Hispanic interviewer would reduce the experiences of stereotype threat and arousal for Hispanic interviewees, and increase the magnitude of difference between Hispanic liars and truth-tellers as a consequence.

There are two novel aspects of the methods proposed in this study. First, stereotype threat is investigated for the first time in a context relevant to interviewing and interrogations. Second, direct measures of cognitive load are used to elucidate the mechanisms underlying Cognitive Load Approaches. These methods were designed to yield results that inform theory and advance application to investigative settings.

[3] This statistic is based on reports from the Los Angeles Police Department. Whites comprised the largest portion of officers (42%), although Hispanics were a close second (37%).
[4] Individuals who report they are White are a majority in both Los Angeles and Orange County (72 and 74% respectively), however those who report that they are White and *not of Hispanic origin* are the largest group in Orange County(43.5%) but not in Los Angeles County (28%). For detailed information on reporting of race please reference U.S. Census Bureau (2014a; b; c).

Chapter 2: Experiment 1a

The purpose of Experiment 1a was to examine the extent to which social context, defined by ethnicity of interviewee, primes stereotype threat, increases arousal, and imposes cognitive load more for Hispanic than White interviewees. To test this prediction, Hispanic and White participant-interviewees engaged in a mock-crime scenario and lied or told the truth about their experiences during a video-recorded interview. These recordings were used as stimuli in Experiments 1b and 1c.

Methods

Participants and design. A total of 47 community members and university students were recruited as *participant-interviewees* (i.e., interviewee; *M* age=23, *SD*=7.39; 41% male; 22 Hispanic). Participants were recruited using flyers, online recruitment systems through the university, and through referrals from friends and family members. Flyers and recruitment systems told participants that they would be participating in an "engaging, 1.5 hour study" where they would be "asked to plan and carry out a series of tasks possibly including a mock crime. .To provide an incentive to participate in the study, interviewees were entered in a raffle to receive $200. Interviewees were also told that they had the opportunity to receive $10 if they "passed the credibility assessment" to increase motivation to perform the tasks as instructed. For ethical reasons, all participants received the $10 whether or not they were included in the study. Exclusion criteria were: (a) video did not record properly, (b) participant did not comply with multiple parts of the experimental protocol intentionally, or (c) participant did not identify as White or Hispanic. A total of 15 interviews were excluded. After removing these participants, a total of 32 interviews were used as stimuli.

An interviewer (White male, 26 years) was trained using a Cognitive Load interview approach (see Vrij & Granhag, 2012a). Training for the interviewer consisted of an hour-long session that explained the interview protocol. The interviewer was instructed that the protocol was an information-gathering interview, and that the main goal of the interview was to elicit information about the mock-crime at hand. He was trained to adhere to the scripted interview when possible, and to deviate only if he felt more information was needed on a certain question. In this regard, the instructor explained that using open-ended questions that allow the interviewee to be in control of the interview are critical, so any deviations from the script should only include open-ended questions such as "please tell me more." After the initial training, the interviewer practiced trials of mock interviews and received feedback. The interviewer was blind to the hypotheses of the study and to the condition of each interviewee.

This study was a 2 (Veracity: Lie or Truth) x 2 (Ethnicity: Hispanic or White) between-subjects design. Several dependent variables were used to measure the constructs of Stereotype Threat (two measures of self-reported threat, a measure of cognitive vigilance) and Cognitive Load (a self-report scale, and error and RT on a secondary task). Measures of arousal were also computed pre- and post-interview [mean arterial blood pressure and pulse, and a self-report scale named the State-Trait Anxiety index (STAI)].

Materials. A summary of the measures used in this research is presented in Table 1. Measures of stereotype threat that are based on previous research (e.g., Blascovich, Spencer, Quinn & Steele, 2001; Murphy, Steele, & Gross, 2007; Steele & Aronson, 1995) included the following three self-report scales: (a) cognitive vigilance of situational

cues measured using a 15-item memory scale of details recalled during the interview (see Appendix 3); (b) self-report answers to the "interview experience" scale which included questions designed by Najdowski (2013) that more specifically address the extent that stereotype threat is experienced by participants in the context of an interview (e.g., "I was worried that the interviewer would suspect me of having committed a crime just because of my race"; and (c) a "general experience" scale adapted from Steel and Aronson (1995) with questions designed to address more general experiences of threat in contexts outside of this interview experience that are related to a criminal investigation (e.g., "In interviews, people of my race often face biased evaluations"). Appendix 4 presents the interview and general experience scales.

Arousal is a measure that is often correlated with stereotype threat and experiences of cognitive load. Measures of arousal in this study include: (a) anxiety and physiological arousal measured by mean arterial blood pressure (MABP) and pulse, two responses known to vary as a function of stress, arousal or emotional excitement, (b) a measure of self-report anxiety (named the State-Trait Anxiety index). These indicators of stereotype threat follow from the model by Najdowski (2011).

The cognitive load experienced by interviewees was objectively operationalized by error rate and reaction time (RT) in response to stimuli on a secondary task performed during the interview. Lying or telling the truth was the primary task. The secondary task was an auditory probe detection task requiring interviewees to press a mouse button as quickly as possible each time they heard a target beep in one ear. Beeps of high and low tones were periodically heard on a headset. Tones were counterbalanced across conditions so that approximately half the trials were presented with a high or low tone.

This task was chosen to minimize impact on performing the primary task (i.e., convincing the interviewer), while still being sensitive to assessing cognitive load experienced during the interview conditions (Brünken, Plass, & Leutner, 2004; Paas, Tuovinen, Tabbers & Van Gerven, 2003). At least one other study has used a similar auditory task to measure or manipulate cognitive load (Lane et al., 2014). Measures of subjective perception of cognitive load were included in post-interview ratings provided by interviewees Aligning with efforts of researchers interested in studying Cognitive Load Approaches, the study applied a direct measure of cognitive load (from the secondary task) to assess the interviewees' mental load experience during the interview (see Blandón-Gitlin et al., 2014 and Walczyk et al., 2013 for reviews).

Procedure. The experimental session consisted of four phases: (a) an *introductory* phase, (b) the *crime* phase, (c) the *interview* phase, and (d) the *post-interview-questionnaire* phase. Within ethnic and gender groups, half of the interviewees were randomly assigned to lie and half were assigned to tell the truth. Upon arriving to the experiment, participants read and signed the informed consent and completed a baseline measure of self-report anxiety using the STAI scale during the *introductory phase*. In the subsequent *crime phase*, interviewees participated in a complex scenario designed to mimic a situation relevant to intelligence-gathering forensic contexts. In these contexts eliciting accurate and complete information is as critical as detecting deception. Interviewees may conceal information and may not admit wrongdoing but the information they provide might be useful in determining the occurrence of events (Kleinman, 2006).

The crime scenario deviates from the typical scenario in laboratory-based research whereby one group performs a specific task and the comparison group does not. To enhance the applied value of this work, all of our interviewees participated in the scenario outside of a lab setting (retrieving journal articles from a professor's office to be delivered to another location), except that the liars had an additional "crime" task to carry out (steal a folder with confidential information from the professor's office). The research assistant provided instructions on the scenario. Participants were provided with the mission's instructions and a key to enter the professor's office. Liars were given the additional instruction to search for and steal a folder with sensitive information from the office. An interviewer was waiting for each interviewee at the location where the interview took place. The RA informed all participants that the interviewer did not know the specifics of the trip to the professor's office but she/he knows that sensitive information was stolen. The interviewees in the lie condition were instructed not to mention the search and the 'stealing' of the file. All groups were told to appear as truthful as possible so as not to raise suspicion of them. See Appendix 1 for the mock-crime "script" administered by the RA.

In the *interview phase*, the interviewer was blind as to interviewees' veracity condition. There were four parts of the interview phase. First, interviewees' blood pressure (systolic and diastolic) and pulse were measured pre-interview. Second, another measurement of anxiety using the STAI scale was administered. Third, interviewees received instructions on how to complete the auditory task and completed practice trials to ensure interviewees were comfortable with the sensors and procedures, and ensure that baseline measures of the secondary task were obtained. Finally, the interview occurred.

There were initial greetings and questions in the interview protocol. The questions included a series of open-ended prompts (e.g., "Tell me about your trip to the professor's office") and focused questions (e.g., "What did you see in the office?") that were part of an information-gathering approach that imposes cognitive load, especially in liars. This version of the protocol was an information-gathering interview adapted from the Cognitive Load Approach described by Vrij and colleagues (e.g., Vrij et al., 2011), and the Modified Cognitive Interview used in a number of laboratory and field studies (e.g., Morgan, Mishara, Christian, & Hazlett, 2008). Appendix 2 includes the interview protocol.

The interview used information-gathering questions intended to elicit as much information as possible about the target, mock-crime event. There were three aspects of the interview to induce more cognitive load for liars than for truth-tellers: (a) unanticipated questions (asking about the spatial layout of a target location and asking about the temporal order of events), (b) a reverse order question (asking the participant to recall their account in reverse chronological order), and (c) repeated questions (asking the participant to repeat everything that happened to them during the target event towards the end of the interview). All these question types have been shown to increase the cognitive difficulty of lying and increase cues to deception (Blandón-Gitlin et al., 2012; Vrij et al., 2008; 2009). While questions of this type should be difficult for liars, they were designed to assist truth-tellers with accessing their episodic memory of the event. For example, interviewees were asked questions like, "If I was next to you and I could see but not hear, tell me what I would see around the office and describe the layout of the office." Similar to the cognitive interview, these types of questions should help truth-tellers mentally

reinstate the context of their memories while still imposing cognitive demands on the liar. Interview sessions were videotaped for use in Experiments 1b and c.

During the *post-interview-questioning phase*, interviewees completed the 15-item memory scale assessing cognitive vigilance (Murphy et al., 2007), a self-report measure of cognitive load (e.g., "How mentally challenging was the interview session?"), and a third measurement of the STAI. Finally, individual differences in interviewee's working memory capacity was assessed via an automated version of the Operation Span Task (OSPAN) and used for analysis in Experiment 1d. Scores on the OSPAN task are related to other complex span tasks, such as reading span tasks and counting span tasks, in assessing working memory capacity (see Conway et al., 2005). They also tend to have good internal consistency and test-retest reliability (Conway et al., 2005; Unsworth, Heitz, Schrock, & Engle, 2005). In this OSPAN task participants solve a series of simple mathematical problems (e.g., 9/9 +7 = 8; TRUE or FALSE) while holding target letters in memory. A score is computed based on the number of letters recalled in the correct order across all trials. For example, if a participant recalled 6 letters correctly out of a set of 6, then "6" would be added to their total score from that 6-set of letters. If on the next set presented that same participant recalled 4 letters correctly out of a set of 5, then "0" would be added to their total score. Thus, their cumulative OSPAN score would be 6 across these two sets. Interviewees were removed from these analyses if their accuracy on the math problems fell below 85%; one participant was removed from this sample based on these criteria. This method of scoring is utilized in the original article for the Automated OSPAN task (Unsworth et al. 2005), and is similar to the all-or-nothing-load scoring described in Conway et al. (2005). Another participant's data did not record due

to a computer malfunction. Higher scores indicate higher working memory capacity. Interviewees completed this task at the end of their participation so as not to confound the experimental conditions with performance on this task.

Results

A total of 32 interviewees participated, eight in each condition. Approximately 41% were male. The mean age of interviewee was 23.0 years (*SD*=7.39). The mean age was similar across conditions. Hispanic liars (3 males) had a mean age of 24.8 (*SD* = 9.11); White liars (3 males) had a mean age of 22.1 (*SD* = 2.54); Hispanic truth-tellers (3 males) had a mean age of 24.4 (*SD* = 11.03); White truth-tellers (4 males) had a mean age of 20.6 (*SD* = 3.29).

Motivation. Participants were asked to rate on a 1 (not at all motivated) to 7 (very motivated) scale "How motivated were you to convince the interviewer that you were truthful." On average, participants were motivated to appear truthful during the interview (*M* = 6.00, *SD* = .95). Critically, a 2 (Interviewee Ethnicity) x 2 (Veracity) between-subjects Analysis of Variance (ANOVA) suggested that there were no significant differences in motivation ratings across conditions (all *F* values < 1).

Stereotype threat. Three self-report measures of stereotype threat were analyzed. Two scales of stereotype threat were analyzed to examine the extent to which participants experienced threat during the interview (interview experience scale) and during their life generally (general scale). A memory questionnaire was analyzed separately to examine cognitive vigilance, the extent to which participants attended to stereotype-relevant cues during their interview.

Stereotype threat experience scales. Two items on the general threat scale were reverse coded prior to analyses so that higher averages on each scale indicated higher

perception of stereotype threat. Means were computed on two scales for each participant: (a) general threat and (b) threat of interview experience (see Figure 1). Univariate and multivariate assumptions were investigated prior to analyses by examining standardized values, skew and kurtosis values, and a Mahalanobis distance test. There were no significant violations to assumptions. Results of a 2 (Interviewee Ethnicity) x 2 (Veracity) between-subjects Multivariate Analysis of Variance (MANOVA) indicated that Hispanic interviewees reported experiencing more stereotype threat than Whites on both the interview experience scale (Hispanics: $M = 4.05$, $SD = .72$; Whites: $M = 2.53$, $SD = .68$) and the general threat scale (Hispanics: $M = 3.02$, $SD = 1.69$; Whites: $M = 1.51$, $SD = .51$), $F(2, 27) = 19.85$, $d = 1.63$. Univariate tests showed that Hispanics reported significantly more threat than Whites on both the general, $F(1, 31) = 41.04$, $p < .001$. $d = 1.21$) and interview experience scales, $F(1,31) = 11.89$, $p = .002$, $d = 2.17$. No other effects were significant (F's < 1.59).

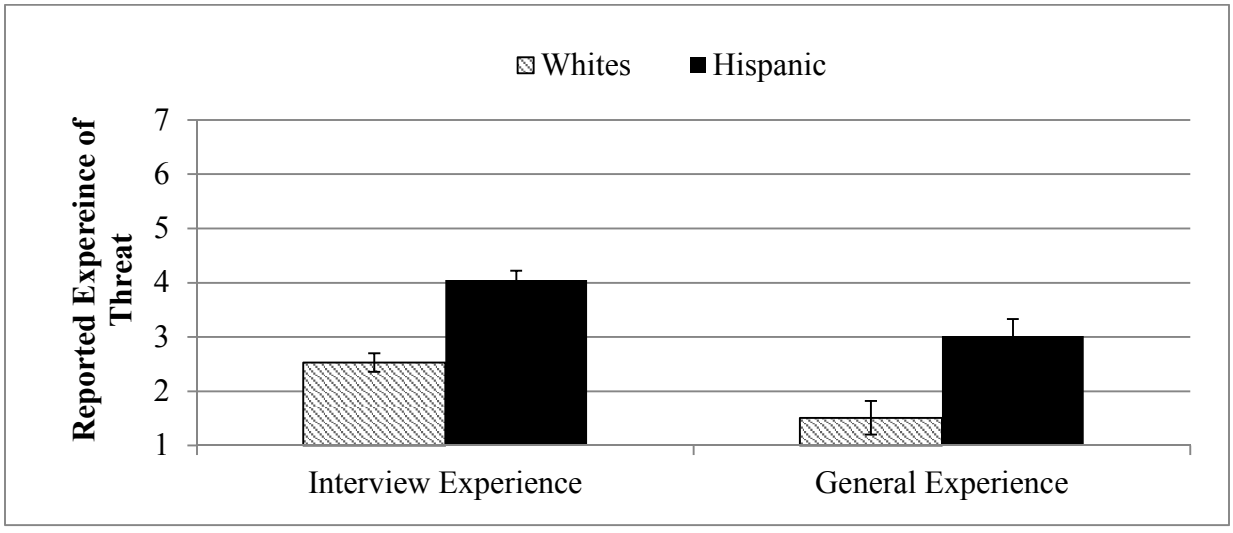

Figure 1. Stereotype Threat Scales Experiment 1a. Average scores on stereotype threat scales, as a function of interviewee ethnicity and scale-type (general or interview experience). Error bars represent standard error of the mean.

Cognitive vigilance scale. Cognitive vigilance was assessed on a 27-point scale, with higher values indicating higher vigilance to possible stereotype cues. Univariate and multivariate assumptions were investigated prior to analyses by examining standardized values, and skew and kurtosis values. There were no significant violations to assumptions. A 2 (ethnicity) x 2 (veracity) between-subjects ANOVA was conducted on cognitive vigilance scores. Counter to our hypotheses, White interviewees (M = 22.56, SD = 3.09) had significantly higher scores than Hispanic interviewees (M = 19.33, SD = 3.56), $F(1,31) = 8.07$, p = .008, d = .97. No other effects were significant (F's<4.18). All means and standard deviations presented in Appendix 5, Table A. One explanation for these results is that the questions used to measure vigilance were not sufficiently sensitive to capture the types of vigilant cues that people experiencing stereotype threat during an interview attend to. To my knowledge there is no standardized "vigilance" measure in studies investigating stereotype threat. Some researchers use open-ended questionnaires (such as the Verbal Thoughts Questionnaire, Beilock et al., 2007) whereas other researchers use direct measurements such as response time to questions (Seibt & Foerster, 2004).

Cognitive load. Cognitive load was assessed using: (a) error rates and reaction time on a secondary task that participants completed during their interview, and (b) six self-report questions regarding experiences during the interview.

Secondary task: error and RT. One participant was detected as an outlier with an error rate of 93%. This person was removed from all analyses. Two separate 2 (Interviewee Ethnicity) x 2 (Veracity) x 2 (Time: baseline or interview) mixed design ANOVAs with Time as the within-subjects variable were conducted on error and RT

data. Error rates were calculated as the proportion of times a participant incorrectly responded (missed detecting the tone, or incorrectly indicated the tone was present) out of the total number of trials at baseline (for baseline error rate) or during the interview (for the interview error rate). There was a significant interaction of time and veracity as predicted, $F(1, 27) = 4.86$, $p = .036$, $d = .81$. As can be seen in Figure 2, right panel, whereas liars made significantly more errors during the interview ($M = .24$, $SD = .13$) than during baseline ($M = .07$, $SD = .12$), $t(14) = 3.93$, $p = .002$, $d = 1.36$; truth-tellers had similar error rates at baseline ($M = .06$, $SD = .09$) and during the interview ($M = .11$, $SD = .08$), $t(15) = 1.74$, $p = .102$, $d = .59$. No other statistical significance was found. All means and standard deviations presented in Appendix 5, Table B.

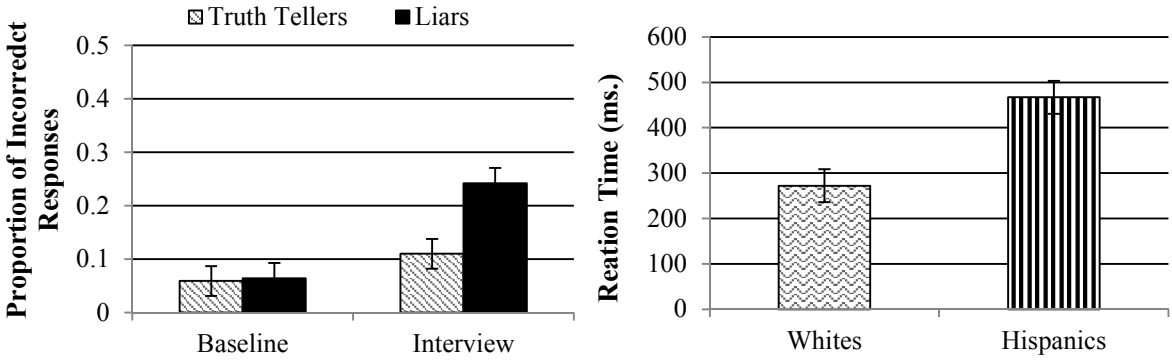

Figure 2. Experiment 1a Secondary Task Results. Mean RT (left) and error rate (right) on the secondary task, as a function of interviewee ethnicity and veracity. Error bars represent standard error of the mean.

The second measure of cognitive load was reaction time (RT). As predicted there was a significant main effect of interviewee ethnicity; Hispanics ($M = 467.08$, $SD = 180.56$) had longer RTs than Whites ($M = 261.56$, $SD = 83.33$), $F(1, 27) = 13.95$, $p = .001$, $d = 1.46$. No other results were significant (F's < 2.02). Hispanics took longer to respond than Whites, indicating greater cognitive overload. Hispanics in this study

experienced a speed-accuracy tradeoff: Hispanics took longer to respond and did not show an increase in making errors during the interview compared to baseline. This suggests that Hispanics monitored their responses during their interview more closely, even at the expense of increased reaction time. This aligns with predictions of stereotype threat; threatened individuals tend to monitor their behaviors more than non-threatened individuals (e.g., Coy et al., 2011).

Self-Report. Participants' answers to six questions regarding the difficulty of the interview on 1(not at all difficult) to 7(very difficult) scales were analyzed. Self-report measures of "cognitive load" have been used in other studies examining the cognitive load approach (e.g., Vrij et al., 2008; 2009). Univariate and multivariate assumptions were investigated prior to analyses by examining standardized values, skew and kurtosis values, and a Mahalanobis distance test. There were no significant violations to assumptions. In a 2 (Interviewee Ethnicity) x 2 (Veracity) between-subjects Multivariate Analysis of Variance (MANOVA) there were no significant multivariate (F's < 2.32) or univariate (F's < 3.33) effects. Table C in Appendix 5 includes the means and standard deviations in each condition.

Arousal and anxiety. Measures of arousal and anxiety were analyzed using (a) Mean Arterial Blood Pressure and Pulse measure before and after the interview, and (b) STAI responses measured at three points across participation (before the mock-crime, before the interview, and at the conclusion of their participation). Anxiety was included because it is hypothesized to be one mechanism that accounts for the debilitating effects of stereotype threat on performance: as threat increases anxiety increases also, reducing

access to executive functioning resources as a consequence (Najdowski, 2011; Schmader, et al., 2008).

Mean arterial blood pressure (MABP) and pulse. One participant was removed because their blood pressure and pulse did not record due to an equipment failure following the interview. Univariate and multivariate assumptions were investigated prior to analyses by examining standardized values, skew and kurtosis values, and a Mahalanobis distance test. One multivariate outlier was removed from MABP and pulse analyses.

MABP was calculated using the formula, (Systolic BP + 2 (Diastolic BP))/ 3 (Ward & Langton, 2007; for a different formula see Zheng, Amoore, Mieke, & Murray, 2011). Two measurements of MABP were recorded pre- and post-interview, and an average of these measurements was used[5] to create a MABP score per participant. A 2 (Ethnicity) x 2 (Veracity) x 2(Time: pre- or post-interview) mixed-design ANOVA, with time as the within-subjects variable was computed using MABP values. As predicted there was a significant main effect of ethnicity; Hispanics (M = 101.14, SD = 17.94) had significantly higher MABP than Whites (M = 82.03, SD = 22.63), $F(1, 26) = 6.45$, p = .017, d = .94. As can be seen in Figure 3, there was a significant interaction of time and veracity, $F(1, 26) = 4.31$, p = .048, d = .77. Liars MABP decreased significantly from pre- (M = 103.17, SD = 23.80) to post-interview (M = 96.59 SD = 21.73), $t(14) = 2.88$, p = .012, d = .29; whereas Truth-tellers had a similar MABP pre- (M = 82.60, SD = 19.92) and post-interview (M = 83.99, SD = 21.54), $t(14) = .39$, p = .704, d = .07. Liars' showed a greater reduction in arousal from pre- to post-interview than truth-tellers. These results

[5] Some participants (n=7) blood pressure and pulse recorded only once at pre- or post- interview. These participants MABP were computed using only one measurement of systolic and diastolic blood pressure.

suggest that liars were more aroused than truth-tellers, and showed a greater reduction in arousal after being interviewed.

A 2 (Ethnicity) x 2 (Veracity) x 2 (Time: pre- or post-interview) mixed-design ANOVA, with Time as the within-subjects variable was computed using pulse values. There were no significant effects or interaction (*F*'s < 2.65). Table D in Appendix 5 includes the means and standard deviations in each condition.

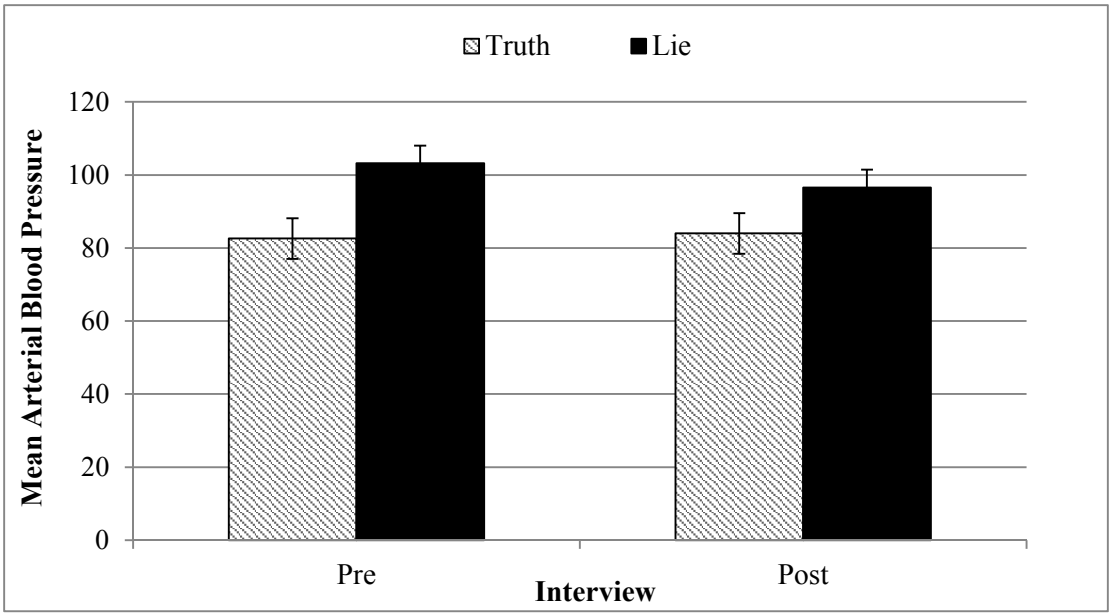

Figure 3. MABP Values. Experiment 1a results of average MABP values, as a function of interview (pre- or post-) and veracity. Error bars represent standard error of the mean.

State-Trait Anxiety Index (STAI). STAI values were conducted by summing across all items on the scale. Nine items were reverse coded prior to analyses so that higher values on the scale indicate higher levels of anxiety. Values on this scale range from 20 to 80. Univariate and multivariate assumptions were investigated prior to analyses by examining standardized values, skew and kurtosis values, and a Mahalanobis distance test. There were no significant violations to assumptions. A 2 (Ethnicity) x 2 (Veracity) x 3 (Time: pre-mock crime, pre-interview, post-interview) mixed-design

ANOVA, with time as the within-subjects variable was conducted on STAI values. As predicted there was a significant main effect of time, $F(2, 27) = 8.82$, $p = .001$, $d = 1.07$. Pairwise comparisons suggest that participants were significantly less anxious pre-mock crime ($M = 34.78$, $SD = 6.53$) than immediately pre-interview ($M = 39.05$, $SD = 11.76$), $p = .009$, $d = .45$. Participants were significantly less anxious post-interview (M=30.59, SD=7.86) than at pre-interview, $p < .001$, $d = .85$, or pre-mock crime, $p = .002$, $d = .58$. Put another way, all participants reported the highest levels of anxiety immediately before the interview. No other effects were significant (F's < 1). Table E in Appendix 5 presents the means and standard deviations in each condition.

The hypotheses for Experiment 1a were partially supported. First, the significant interaction of veracity and time on error rate on the secondary task supports the hypothesis that the interview was more cognitively demanding for liars than truth-tellers. Second, for self-report measures of threat across two different scales, Hispanics reported experiencing more stereotype threat than Whites during their interview experience, and reported experiencing more stereotype threat generally during their lives. However, contrary to hypotheses, Hispanics scored lower than Whites on the cognitive vigilance scale. One possible explanation is that the cognitive vigilance scale was not measuring the intended construct of vigilance towards stereotype threat cues. Finally, as hypothesized Hispanics experienced higher objective signs of arousal and cognitive load than Whites; Hispanics had higher MABP values and longer RTs during the interview.

Chapter 3: Experiment 1b

The results of Experiment 1a support the hypothesis that during a cognitively demanding interview, Hispanic interviewees accused of a crime experienced stereotype threat and show signs of increased cognitive load compared to White interviewees. Najdowski (2011) suggested that regardless of actual veracity, stereotype threat, mediated by cognitive load, would exacerbate in interviewees the display of deceptive behaviors. Given that Hispanic interviewees reported experiencing higher levels of stereotype threat, it is predicted that Hispanic liars and truth-tellers displayed similar high rates of behaviors that are diagnostic of deceit (e.g., thinking hard or appearing anxious), whereas White liars would show significantly more behavioral cues that indicate deceit than truth-tellers. Put another way, because Hispanic liars and truth-tellers experienced more similar levels of load and arousal than White liars and truth-tellers, the difference in behaviors between Hispanic interviewees should be smaller than the difference in behaviors between White interviewees. Experiment 1b tests these predictions. Third-party observers viewed the video-recorded interviews from Experiment 1a and rated interviewees on several behavioral indices designed to test the presence or absence of six diagnostic cues to deceit or truth. The purpose of Experiment 1b was to examine the extent to which interviewee behaviors differ as a function of interviewee race and veracity.

Methods

Participants and design. A total of 184 observers (55 % female; *M* age[6] = 35.84, *SD* = 12.78; 69% White, 12% African American, 8% Hispanic, 10% Asian, 2% other) watched videos of interviews collected during Experiment 1a and rated behavioral cues related to deception or truth. Observers were recruited using Amazon's Mechanical Turk (M-Turk)[7], a crowd-sourcing website used increasingly by social science researchers (Buhrmester, Kwang, & Gosling, 2011). This experiment was a 2 (Veracity: truth or lie) x 2 (Ethnicity: Hispanic or White) within-subjects design using ratings of three composite variables (appearance of cognitive load, anxiousness, and confident and convincing appearance) as the dependent variable.

Materials and procedure.

Videos of interviewees. Videos were created for each account given (true or false, Hispanic or White participant). Each video clip included a participant's entire response to the temporal-order three-part question: "(1) Tell me about the first… (2) next…. (3) last couple of things that you did during your time in the office." This resulted in 32 independent video clips, one clip for each participant. The average length of video clips was 77 seconds (*SD* = 28), ranging from 26 to 145 seconds. The average length of time was similar across all conditions (Hispanic Liars: *M* = 81, *SD* = 29; Hispanic Truth-Teller: *M* = 80, *SD* = 30; White Liars: *M* = 77, *SD* = 29; White Truth-Tellers: *M* = 75, *SD* = 16). Each clip captured a suspect from the waist up.[8] To ensure that observers in

[6] One person did not report their age.
[7] A total of 16 observers were removed prior to analyses because they failed the attention check, their location was outside the United States, or their IP address was identified twice.
[8] Some clips may have appeared to show certain interviewees closer to the camera than others. This effect was caused primarily by the varying body sizes of the interviewees. The effect of these slight differences is

Experiment 1b had a manageable set of videos to evaluate, the final set of 32 video clips was separated into two sets that each included a total of 16 video clips, counterbalanced so that the same number of each condition was included in each set. Observers were given only one set of 16 video clips.

Behavior ratings of videos by third-party observers. The observers rated the video-recordings of interviewees from Experiment 1a on six behavioral cues: two cues related to truth and four cues related to deception (see Table 2 for all questions and average responses per condition). Results of several studies, including comprehensive meta-analyses of cues to deception, suggest that certain patterns of behavior tend to be related to lying or telling the truth (DePaulo et al., 2003; Hartwig & Bond, 2011; Hauch, et al., 2014). Hartwig and Bond (2011) meta-analyzed the relationship between the appearance of lie and truth behavioral cues and actual veracity across 134 cues. In addition to other cues, their results suggested that liars appeared to be thinking harder, more indifferent, and less spontaneous than truth-tellers. Truth-tellers appeared to be more cooperative and relaxed, and produced stories that were more realistic than liars.

The six cues used in the current experiment were chosen based on results of Hartwig and Bond (2011), along with other comprehensive meta-analyses (e.g., DePaulo et al., 2003; Hauch et al., 2014). Third-party observers in Experiment 1b rated the presence or absence of each of six cues on a 1 (not at all present) to 7 (very present) scale. An attention check was also used; participants answered, "The answer to this question is 15," and had to choose the option "15."

---

minimal as all clips ensured that observers had a clear view of participants from the same angle and approximately the same distance.

Table 2.

Experiment 1 White Interviewer: Third-Party Observers' Ratings on the Six Behavioral Cues, as a Function of Cues, Interviewee Race and Veracity Condition

	Hispanic Interviewee						White Interviewee					
Behavioral Cues	Truth		Lie				Truth		Lie			
	M	*SD*	*M*	*SD*	*t*	*d*	*M*	*SD*	*M*	*SD*	*t*	*d*
Cognitive Load	**3.94**	**1.09**	**3.51**	**1.02**	**6.35***	**.41**	**3.21***	**1.04**	**3.92**	**.88**	**10.91***	**.74**
1. How hard does the interviewee appear to be thinking?	4.16	1.19	3.74	1.10	5.90*	.37	3.56	1.18	4.01	.90	6.27*	.43
2. How much difficulty does the interviewee appear to be having in answering the questions?	3.72	1.20	3.29	1.14	5.24*	.37	2.87	1.08	3.82	.97	12.54*	.92
Anxiousness	**3.57**	**1.16**	**3.37**	**1.10**	**2.27***	**.18**	**2.83**	**1.08**	**3.46**	**1.02**	**9.78***	**.60**
3. How anxious does the interviewee appear?	3.75	1.19	3.47	1.12	3.40*	.24	2.96	1.10	3.60	1.02	9.31*	.60
4. How stressed does the interviewee appear?	3.39	1.21	3.27	1.15	1.77*	.10	2.70	1.11	3.32	1.10	9.24*	.56
Confident and Convincing Appearance	**4.27**	**1.01**	**4.51**	**.90**	**3.55***	**.25**	**4.84**	**.88**	**4.04**	**.87**	**12.77***	**.91**
5. How confident does the interviewee appear?	4.14	1.06	4.41	.92	3.52	.27	4.74	.90	3.94	.85	11.34	.91
6. How convincing is the interviewee's story?	4.40	1.08	4.61	1.01	2.98	.20	4.94	.97	4.14	1.02	11.79	.80

Note. All ratings were on a 1 (not at all) -7 (very) scale. Cognitive demand and anxiousness involve cues to deception. Confidence and convincing appearance indicate cues to truthfulness. Bolded rows represent the mean of each cluster (e.g., Cognitive Load).

*p<.001

The associated behavioral cue ratings from observers were averaged to create three composite variables: cognitive load, anxiety, and confident and convincing appearance. Cronbach alpha values were examined between to-be-combined items. All alphas were greater than or equal to .78.

Results

Figure 4 presents the relationship between ethnicity and veracity for each cluster of behaviors (cognitive load, anxiety, and confident and convincing appearance). Univariate and multivariate assumptions were investigated prior to analyses by examining standardized values, skew and kurtosis values, and a Mahalanobis distance test. Three multivariate outliers were detected using a Mahalanobis distance test. Sensitivity analyses suggest that significance tests for the critical composite variables did not differ when excluding or including these cases; outliers were not removed.

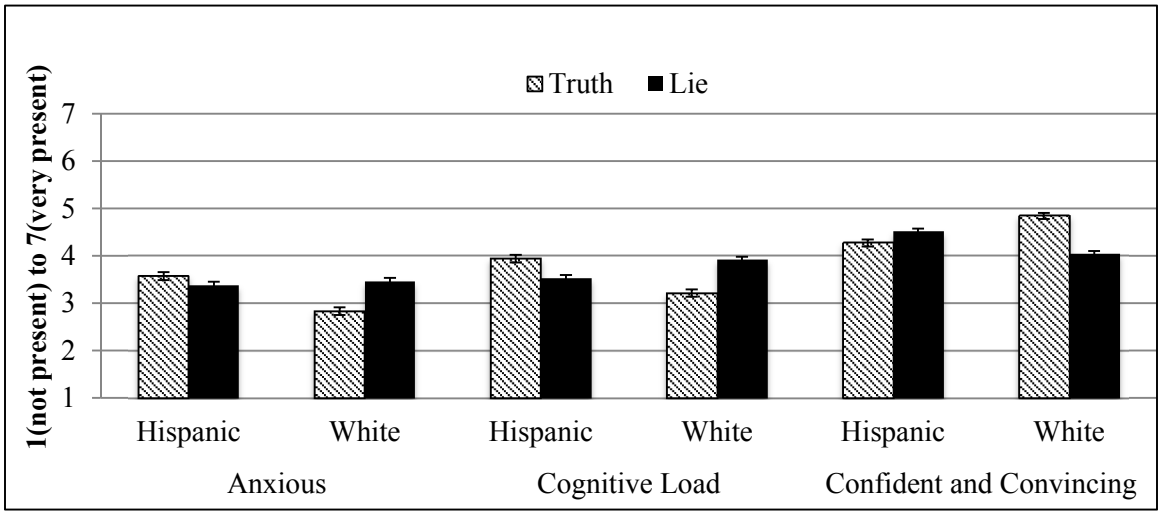

Figure 4. Behavioral Cue Ratings. Average observer ratings of interviewee behavior for each cluster of cues as a function of interviewee ethnicity and actual veracity. Error bars represent standard error of the mean.

Observers' average responses to rating scales for each composite variable were analyzed using three 2 (Interviewee Veracity: truth or lie) x 2 (Ethnicity of Interviewee:

Hispanic or White) within-subjects ANOVAs: one for cognitive load, one for anxiousness, and one for confident and convincing appearance. There was a significant interaction of veracity and ethnicity on the rating of cognitive load, $F(1, 183) = 157.03$, $p < .001$, $d = .92$ (see Figure 4). Whereas White liars ($M = 3.92$, $SD = .84$) were rated as significantly more cognitively loaded than truth-tellers ($M = 3.21$, $SD = 1.04$), $t(183) = 10.91$, $p < .001$, $d = .75$, for Hispanic interviewees the effect was reversed, Hispanic liars ($M = 3.51$, $SD = 1.02$) appeared less cognitively loaded than truth-tellers ($M = 3.94$, $SD = 1.09$), $t(183) = 6.35$, $p < .001$, $d = .41$. Most critically the magnitude of difference between liars and truth-tellers was smaller for Hispanics ($d = .41$) than Whites ($d = .75$) These results align with Experiment 1a, in which Hispanic liars and truth-tellers experienced more similar levels of cognitive load, and thus displayed more similar behaviors of cognitive load in Experiment 1b. A paired t-test was conducted comparing the absolute difference in observer ratings of cognitive load between liars and truth-tellers that were Hispanic or White. The average absolute difference in observer ratings between Hispanics liars and truth-tellers (M = .75, SD = .66) was non-significantly smaller than White liars and truth-tellers (M = .85, SD = .74), $t(183) = 1.36$, $p = .176$, $d = .14$.

For ratings of anxiety by the observers, a significant interaction of veracity and ethnicity was detected, $F(1,183) = 74.93$, $p < .001$, $d = .64$. Whereas White liars ($M = 3.46$, $SD = 1.02$) were rated as significantly more anxious than truth-tellers ($M = 2.83$, $SD = 1.08$), $t(183) = 9.78$, $p < .001$, $d = .60$, Hispanic liars ($M = 3.37$, $SD = 1.10$) appeared less anxious than truth-tellers ($M = 3.57$, $SD = 1.16$), $t(183) = 2.77$, $p = .006$, $d = .18$. Consistent with hypotheses, the magnitude of difference between liars and truth-tellers was smaller for Hispanics ($d = .18$) than Whites ($d = .60$). However, this difference was

non-significant. The average absolute difference in observer ratings between Hispanics liars and truth-tellers (M = .75, SD = .66) was non-significantly smaller than White liars and truth-tellers (M = .85, SD = .65), $t(183) = 1.56$, $p = .120$, $d = .15$.

As predicted, there was a significant interaction of veracity and ethnicity on confidence and convincing appearance, $F(1,183) = 122.85$, $p < .001$, $d = .82$. Whereas White truth-tellers ($M = 4.84$, $SD = .88$) were rated as significantly more confident and convincing than liars ($M = 4.04$, $SD = .87$), $t(183) = 12.77$, $p < .001$, $d = .91$, Hispanic liars ($M = 4.51$, $SD = .90$) were rated as more confident and convincing than truth-tellers ($M = 4.27$, $SD = 1.01$), $t(183) = 3.55$, $p < .001$, $d = .25$. Most critically, as hypothesized the magnitude of difference between liars and truth-tellers was smaller for Hispanics ($d = .25$) than Whites ($d = .91$). As predicted, this difference was significant such that the average absolute difference in observer ratings between Hispanics liars and truth-tellers (M = .73, SD = .61) was smaller than White liars and truth-tellers (M = .93, SD = .70), $t(183) = 3.21$, $p = .002$, $d = .30$.

Two main results are notable. First, when compared to truth-tellers, observers rated liars as displaying more behavior cues that are diagnostic of deception (e.g., appearing cognitively loaded or appearing anxious) and fewer cues related to truthfulness (e.g., appearing confident and convincing) only when interviewees were White. This supports the hypothesis that the interview techniques used in this study magnified behavioral differences between liars and truth-tellers, when interviewees experienced lower levels of stereotype threat. However for interviewees who experienced higher levels of stereotype threat (Hispanics), an opposite pattern resulted such that truth-tellers displayed more behaviors of deception and less behaviors of truth than liars. One

interpretation for these results is that the Hispanic liars may have experienced lower levels of threat than Hispanic truth-tellers, perhaps because Hispanic liars knew they were safe from any accusations, and because Hispanic liars knew what to expect in the upcoming interview. That is, Hispanic liars were told to steal and lie whereas the truth-tellers thought they were being accused of something (stealing the folder) that they could not defend against if confronted. The Hispanic truth-tellers' naiveté about what exactly was reported stolen from the professor's office may have created additional apprehension beyond just experiencing stereotype threat when entering the interview, causing truth-tellers to experience more anxiety and feel more threatened. As predicted, the White interviewees may not have felt as much threat from a same-ethnicity interviewer, allowing the Cognitive Load Approach interview to induce higher cognitive load on liars than truth-tellers and exaggerate the behavioral differences between liars and truth-tellers as a consequence. Another explanation for this finding is that Hispanic liars *are* displaying behavioral cues related to lying, but those cues were not measured in this study. Hispanics may experience certain emotions such as nervousness or guilt to a greater degree than Whites because they are experiencing more stereotype threat. The results of Experiment 1a partially support this, such that Hispanics had higher average MABP values than Whites. As a consequence of this heightened arousal, Hispanics may display more stereotypical cues to deceit, such as fidgeting or gaze aversion.

Second, the predictions from Experiment 1a were partially supported. Across all cues, the magnitude of difference in observer's behavior ratings between liars and truth-tellers was smaller for Hispanic than White liars and truth-tellers. However this difference was statistically significant for confident and convincing appearance only.

These results suggest that behavioral differences between liars and truth-tellers were smaller for Hispanics than Whites, but only when observers judged confident and convincing appearance. Similar to the explanation for why Hispanic liars displayed less lie behaviors than truth-tellers, it is possible that Hispanic liars and truth-tellers displayed more stereotypical cues to deceit that were not measured in this study.

Chapter 4: Experiment 1c

The results of Experiment 1b suggest that Hispanic liars and truth-tellers tended to look more similar during their interview than White liars and truth-tellers, especially for the confident and convincing appearance cues. Further, observers rated White liars as displaying more behavioral cues of deception and fewer cues of truth than White truth-teller. Given these results, it is predicted that observers making explicit lie-truth judgments (e.g., "how truthful or deceptive is this person?") will be less accurate for Hispanic than for White interviewees. The purpose of Experiment 1c was to investigate this hypothesis by examining observer discrimination accuracy detecting truths and lies. The discrimination accuracy of observers' from a large sample of the US population was assessed.

Methods

Participants and design. A total of 220 observers (55 % female; *M* age[9] = 35.23, *SD*= 17.33; 78% White, 9% African American, 9% Hispanic, 5% Asian, .5% Middle Eastern, 1.5% Native American, and 2%[10] other) [11] watched videos of interviews collected during Experiment 1a. Observers were recruited using M-Turk. This was a 2 (Veracity: truth or lie) x 2 (Ethnicity: Hispanic or White) within-subjects design using three dependent variables: observer's continuous ratings of veracity, proportion correct lie-truth judgments, and a signal detection measure named AUC, which estimates area under the ROC curve as a function of ethnicity.

[9] One person did not report their age.
[10] 4 people chose not to report their ethnicity.
[11] A total of 16 observers were removed prior to analyses because they failed the attention check, their location was outside the United States, or their IP address was identified twice. The same attention check was used as in Experiment 1b.

Materials and procedure.

Videos. Videos and survey sets were identical to those used in Experiment 1b.

Veracity ratings. Observers rated the interviewees on deceptive appearance on a 1 (definitely truthful) to 6 (definitely lying) scale.

Results

Veracity ratings. Figure 5 displays observer's veracity ratings as a function of interviewee veracity and ethnicity. Univariate and multivariate assumptions were investigated prior to analyses by examining standardized values, skew and kurtosis values, and a Mahalanobis distance test. There were no significant violations to assumptions. A 2 (Veracity) x 2 (Ethnicity) within-subjects ANOVA on average veracity ratings was conducted. There was a significant interaction of veracity and ethnicity, $F(1, 219) = 68.24$, $p < .001$, $d = 1.12$.

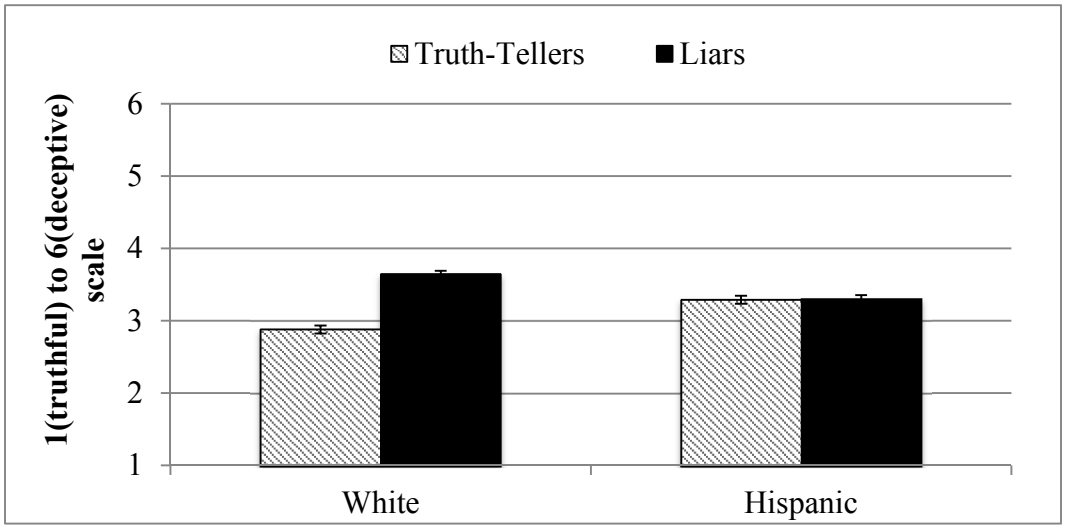

Figure 5. Experiment 1c Veracity Ratings. Third-party observers' average ratings of interviewee veracity as a function of interviewee ethnicity and actual veracity. Error bars represent standard error of the mean.

Whereas White liars ($M = 3.64$, $SD = .76$) were rated as more deceptive than truth-tellers ($M = 2.88$, $SD = .83$), $t(219) = 11.41$, $p < .001$, $d = .96$, Hispanic liars ($M =$

3.30, *SD* = .82) and truth-tellers (*M* = 3.29, *SD* = .81) appeared similarly deceptive, *t(*219) = .24, *p* = .811, *d* = .01. As predicted, Hispanic liars and truth-tellers appeared similar, whereas White liars appeared more deceptive than truth-tellers. These results support the predictions from Experiment 1a, and the results of the pattern of data in Experiment 1b, such that the behavioral differences between Hispanic liars and truth-tellers are more similar than Whites.

Proportion of correct responses. The observers' proportion of correct lie-truth judgments for Hispanic and White interviewees was examined. Proportion of correct judgments was computed by taking the sum of the correctly identified lies and truths out of the total number of lie and truth observations. In Experiment 1c the proportion of correctly identified liars and truth-tellers was significantly greater when identifying Whites (about 62% accurate) than Hispanics (about 50% accurate), $Z = 6.82$, $p < .001$. Critically, Whites were identified at a significantly greater rate than 54%, the average detection accuracy rate indicated by Bond and DePaulo (2006), $Z = 6.16$, $p < .001$. As predicted, the cognitive load interview improved accuracy rates beyond average levels of detection accuracy (54%), as reported by an extensive meta-analysis averaging across 24,482 observer judgments (Bond & DePaulo, 2006). Further, observers were more likely to misidentify Hispanic truth-tellers as liars (44%) than White truth-tellers (31%), $Z = 6.01$, $p < .001$.

ROC analyses. As an additional measure of truth-lie discrimination accuracy, Receiver Operating Characteristic (ROC) curves were calculated using observer's veracity ratings to compare differences in area under the ROC curve (AUC) when viewing Hispanic and White interviewees. ROC curves are based on Signal Detection

Theory (see Banks, 1970 and Stanislaw & Todorov, 1999 for reviews). The ROC curve plots the proportion of correctly identified liars (hit rate) against the proportion of truth-tellers falsely identified as liars (false alarm rate). This allows for comparison of sensitivity (hit rates, the proportion of correctly identified liars) and specificity (the correct rejection rate, or proportion of correctly identified truth-tellers) within a given distribution. After plotting the ROC curve for a given sample, the AUC is computed to investigate the overall performance of an observer at classifying signal trials (liars) and noise trials (truth-tellers), with higher values indicating better overall classification accuracy. Using an empirical ROC curve for investigating discrimination accuracy is recommended over other signal detection measures such as *d'* or *A'* because it is difficult to test whether assumptions of these measures are met within a data set (Gronlund, Wixted, & Mickes, 2014; Zhang & Mueller, 2005). Given the within-subjects design, hit rate and false alarm rates were computed using the total number of observations of liars (n = 880) and truth-tellers (n = 880). The pROC package using R statistical software (Robin et al., 2011) was used to compare the AUC for observers judging Hispanic compared to White interviewees. The pROC software package uses DeLong's method for comparing the difference between two paired ROC curves when examining AUC (DeLong, DeLong, & Clarke-Pearson, 1988). The AUC was significantly greater when observers were judging White (AUC = .64, 95% CI [.62, .67]) than Hispanic (AUC = .47, 95% CI [.44, .50]) interviewees, Z = -32.65, p < .001. This indicates that observers were more accurate discriminating between White, than Hispanic, liars and truth-tellers. Figure 6 displays observer's accuracy discriminating between truths and lies as a function of Ethnicity.

Consistent with predictions, a cognitively demanding interview produced above-chance levels of discrimination accuracy for White but not Hispanic interviewees. A cognitively demanding interview exaggerated observable behavioral differences between liars and truth-tellers (Experiment 1b) and increased observers' lie-truth discrimination accuracy (Experiment 1c) *only* when an interviewee experienced low levels of stereotype threat in Experiment 1a (the White interviewee condition). Hispanic interviewees who experienced higher levels of threat in Experiment 1a were discriminated at about chance levels.

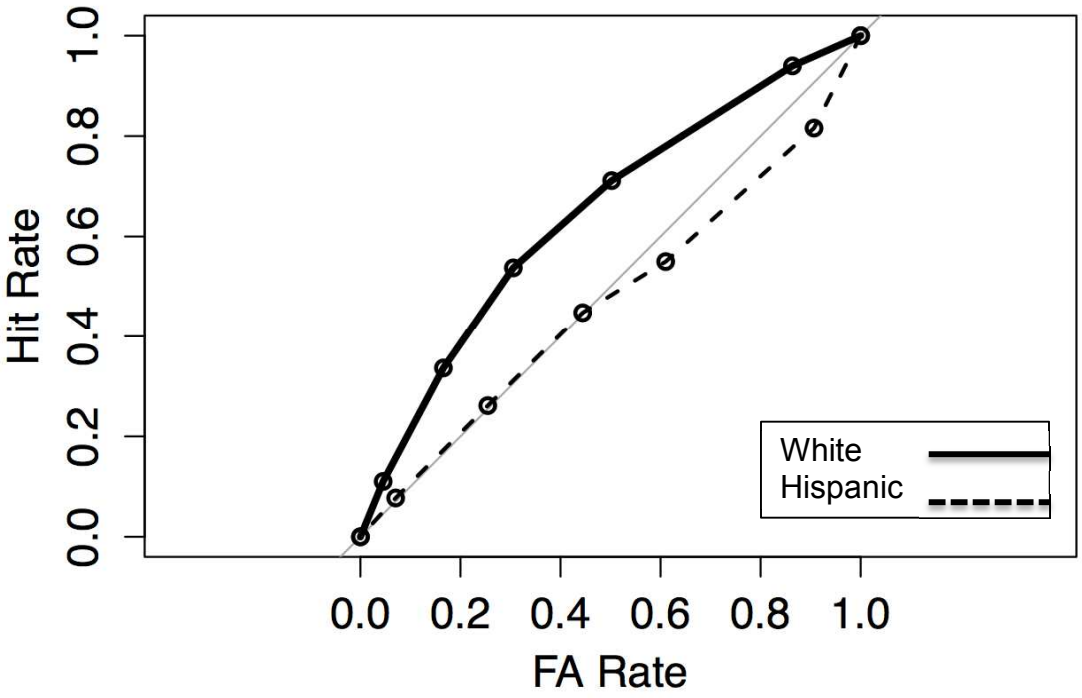

Figure 6. ROC Curves Experiment 1c. Observers' ratings for Hispanic and White interviewees, plotted as a function of hit and false alarm rates. The grey line indicates chance-level responding.

Why did the degree that an interviewee experienced stereotype threat impact his or her ability to appear truthful? As predicted, the results of Experiment 1a suggest that at high levels of threat (Hispanic interviewees), liars and truth tellers experienced similarly

high levels of cognitive load and arousal during their interview. As a consequence, the cognitively demanding interview interfered with Hispanic liars *and* truth-tellers ability to convince an observer that they were truthful. At lower levels of threat (White interviewees), liars experienced higher levels of load and arousal than truth tellers. As a consequence, White liars and truth-tellers were identified at above-chance levels.

These results extend research on Cognitive Load Approach techniques in two important ways. First, success at increasing behavioral differences and increasing lie-truth discrimination accuracy to above chance levels was replicated in the White interviewee condition. Critically, the approach was successful even though liars and truth-tellers participated in a similar mock-crime scenario. Liars and truth-tellers completed identical tasks during the mock-crime scenario with only one critical difference: truth-tellers did not search for or steal a confidential folder. This mock-crime scenario deviates from the typical scenario in laboratory-based lie-detection research whereby one group performs a specific task and the comparison group does not (e.g., Vrij et al., 2008). Compared to other studies, liars and truth-tellers relied on more similar, rich memories when retrieving answers during the interview. Second, the effectiveness of the approach was reduced when interviewees experienced stereotype threat; Hispanic liars and truth-tellers reported experiencing more stereotype threat than Whites, and were discriminated at lower rates of accuracy than Whites as a consequence. Put another way, at higher levels of stereotype threat, interviewees experienced more cognitive load, and arousal, liars and truth-tellers appeared similar to one another and observers had low levels of discrimination accuracy as a consequence.

Chapter 5: Experiment 1d

The results of Experiments 1a, b, and c suggest that a cognitively demanding interview increases observer lie-truth discrimination accuracy even when interviewees are relying on similar memories to answer interview questions. However, the effectiveness of a Cognitive Load Approach was reduced when interviewee's experienced greater stereotype threat. Hispanic liars and truth-tellers experienced similar levels of cognitive load and arousal, and as a consequence were rated similarly by third-party observers.

The purpose of Experiment 1d was to examine one possible underlying mechanism of the relationship between veracity and ethnicity in predicting deceptive appearance: working memory. Results of behavioral and neuroimaging studies support the point that executive functioning processes, such as working memory, are involved to a greater extent in deceptive than truthful responses (Christ et al., 2009; Walczyk et al., 2009). Recent evidence also suggests that interviewees' working memory capacity is related to observers' deception detection accuracy. Blandón-Gitlin et al. (under review) reported that whereas observers' deception detection accuracy for high working memory capacity interviewees improved in a high cognitive load condition over a control condition, detection accuracy was similar across interview conditions for low working memory capacity interviewees. Thus, the working memory capacity of interviewees was hypothesized to moderate the relationship between actual veracity and deceptive appearance during a cognitively demanding interview.

There are two possible predictions. Based on the results of Blandón-Gitlin et al., it follows that the magnitude of difference in observer's accuracy and veracity ratings

(overall deceptive appearance) between liars and truth-tellers would be smaller for high than low working memory capacity interviewees. However, an opposing prediction exists. Neuroimaging research suggests that areas of the brain related to working memory are activated more when lying than when telling the truth (Christ et al., 2009). Working memory is comprised of several, limited capacity sub-systems (Baddeley et al., 2012). Therefore, tasks designed to overload the already-limited capacity systems of working memory, such as the interview task used in this study, may produce more cognitive overload for liars' who have access to fewer resources (low working memory capacity interviewees) than those who begin the lie task with more resources (higher working memory capacity). Following this, high working memory capacity liars may appear more believable because they have more access to working memory resources, even during a demanding interview. That is, the magnitude of difference in observers' accuracy and veracity ratings may be smaller for high than low working memory capacity interviewees.

The relationship between working memory capacity and veracity in predicting discrimination accuracy may also depend on an interviewee's ethnicity. As predicted from work on the relationship between stereotype threat and working memory, stereotype threat should reduce access to working memory resources such as the phonological loop and central executive (Shmader et al., 2008). For example, results of Beilock et al. (2007) suggest that when compared to no-stereotype threat controls, stereotype threat manipulations increase in negative, off-task thoughts related to anxiety about conforming to a particular stereotype. As a consequence, the capacity of the phonological loop and central executive components of working memory is reduced and performance on tasks

that rely on these resources is impaired (see also Coy et al., 2011 for how evaluation anxiety operates in a similar manner). Claude Steele (in Abumrad, 2013) describes this relationship as the consequence of excess "mental chatter" that occurs for individuals experiencing stereotype threat. When experiencing threat, individuals will internalize their worries regarding stereotype threat, mentally "discussing" these worries with themselves. For example, participants experiencing stereotype threat in Beilock's study reported a higher frequency of thoughts like "I thought about how boys are usually better than girls at math so I was trying harder not to make mistakes [even though I did]" when performing math problems. Steele suggests that this excess mental chatter competes with the limited availability of working memory resources, especially the phonological loop and the central executive (Abumrad, 2013).

Given the results of studies on stereotype threat and working memory, and lie detection research there are several competing predictions regarding the interaction of ethnicity, veracity and working memory capacity. Neuroimaging and behavioral research suggest that lying involves working memory resources more than truth-telling. Therefore observers' may be better able to detect Hispanic liars with high than low working memory capacity, whereas truth-tellers may not be affected by stereotype threat at all. That is, liars should be affected more than truth-tellers, especially when those liars are experiencing stereotype threat.

However, it is also possible that in the context of an investigative interview, liars and truth-tellers rely on certain components of working memory, such as the phonological loop, to a similar degree. When access to the phonological loop is reduced, such as is the case when experiencing stereotype threat, liars and truth-tellers may be

affected similarly. That is, Hispanic liars and truth-tellers may appear similar to one another at high and low working memory capacity, and observers may be less able to discriminate between Hispanic liars and truth-tellers regardless of working memory capacity.

Method

Participants and design.

Observer's accuracy ratings from Experiment 1b (N = 220) were re-analyzed as a function of interviewee working memory capacity. The design of this was a 2 (Interviewee Ethnicity: Hispanic, White) x 2 (Interviewee Veracity: Lie, Truth) x 2 (Interviewee Working Memory Capacity: High, Low).

Materials and procedure.

The same veracity ratings analyzed in Experiment 1c were used in this Experiment. The critical difference was that the ratings data were dichotomized based on *participant*-i*nterviewee*'s working memory capacity. Interviewee's were categorized as high or low working memory capacity using a median split analysis on their OSPAN scores (median OSPAN score: 35; M = 36.13, SD = 16.18). Higher scores on the OSPAN are related to higher working memory capacity (Conway et al., 2005). This method of analysis has been used in similar studies such as Blandón-Gitlin et al. (under review) and Lane et al. (2014) (see also research on "choking under pressure," e.g., Beilock et al., 2007). One interviewee was removed from analyses because his/her accuracy on the math problems portion of the OPSAN was below 85% (Unsworth et al. 2005). Another interviewee was removed because of an equipment malfunction in recording the working memory task. In total observers ratings were calculated for 15 high

(10 liars and 8 truth-tellers) and 15 low (8 liars and 10 truth-tellers) working memory capacity participants.

Results

Veracity ratings. Observer's average veracity ratings of interviewees were analyzed using a 2 (Veracity) x 2 (Ethnicity) x 2 (Working Memory Capacity) within-subjects ANOVA. Univariate and multivariate assumptions were investigated prior to analyses by examining standardized values, skew and kurtosis values, and Mahalanobis distance values. There were no significant violations to assumptions. Bonferonni adjustments to α (.05) were used to correct for family-wise error rate for all effects and pairwise comparisons tested, setting α equal to .003. There was a significant three-way interaction of veracity, ethnicity and working memory capacity, $F(1, 219) = 19.21$, $p < .001$, $d = .59$. A simple effects analysis was conducted to compare the interaction of veracity and working memory capacity when observing Hispanics or Whites. When viewing White interviewees with low working memory capacity, observers rated liars ($M = 3.53$ $SD = 1.32$) and truth-tellers ($M = 3.46$, $SD = 1.43$) similarly, $t(218) = .57$, $p = .567$, $d = .05$; whereas when viewing interviewees with high working memory capacity, liars ($M = 3.61$, $SD = .98$) were rated as significantly more deceptive than truth-tellers ($M = 2.76$, $SD = 1.02$), $t(218) = 8.77$, $p < .001$, $d = .85$. When viewing Hispanic interviewees, observers rated liars and truth-tellers with high working memory capacity (liars $M = 2.94$, $SD = 1.48$; truth-tellers $M = 3.13$, $SD = 1.12$), $t(218) = 1.45$, $p = .145$, $d = .14$, and low working memory capacity (liars $M = 3.24$, $SD = .86$; truth-tellers $M = 3.15$, $SD = 1.07$), $t(218) = 1.03$, $p = .303$, $d = .09$, at similar rates.

Figure 7 presents this three-way interaction. All main effects and interactions were significant besides the two-way interaction of veracity and ethnicity. These values are presented in Table 4 and are not interpreted in light of the significant three-way interaction.

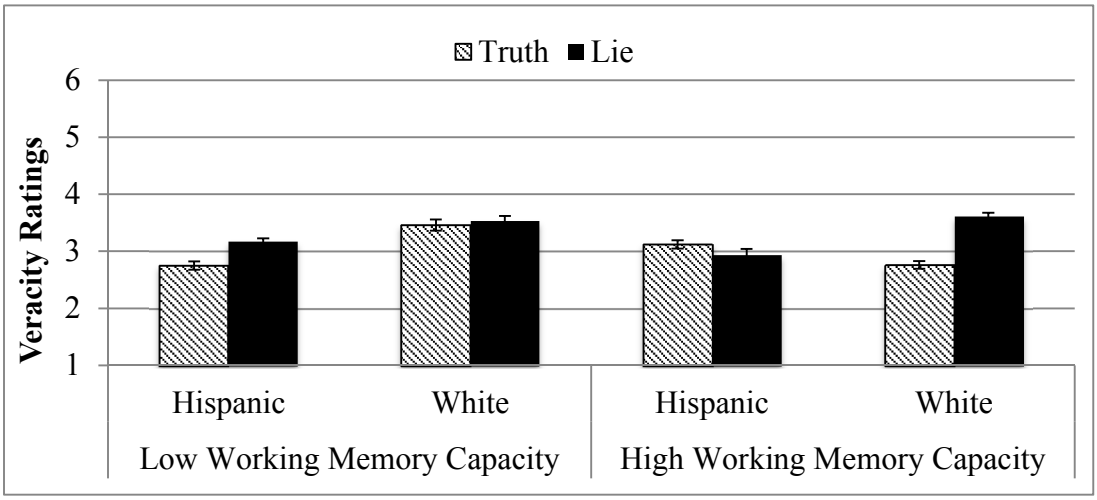

Figure 7. Observer Veracity Ratings of Experiment 1d. Average observer veracity ratings as a function of interviewee ethnicity, working memory capacity, and actual veracity. Error bars represent standard error of the mean.

Table 4.

Experiment 1d F and p values from the 2 (Ethnicity) X 2 (Veracity) x 2 (Working Memory Capacity) ANOVA on veracity ratings.

	F (1, 219)	*p*
Ethnicity	16.99	<.001
Veracity	15.57	<.001
Working Memory Capacity	17.98	<.001
Ethnicity * Veracity	23.05	<.001
Ethnicity * Working Memory Capacity	2.07	0.152
Veracity * Working Memory Capacity	5.00	0.026

These results support the hypotheses. Replicating research by Blandón-Gitlin (under review), a cognitively demanding interview reduced the believability of high more than low working memory capacity liars, but only when experiencing low levels of

stereotype threat (White interviewees). When experiencing stereotype threat, Hispanic interviewees with low compared to high working memory capacity were rated similarly.

ROC analyses. Behavioral ratings data reported above are useful for understanding observers' average ratings for liars and truth-tellers, but they do not provide an analysis of discrimination accuracy. In other words, these data do not assess whether observers accurately discriminated between liars and truth-tellers at above chance levels as a function of veracity, ethnicity and working memory capacity. To address this question, an ROC analysis was conducted to examine the interaction between these variables. A bootstrap analysis with 2,000 samples was used to compare AUC values for two critical comparisons: observer lie-truth discrimination accuracy when rating (a) Hispanic interviewees with high compared to low working memory capacity and (b) White interviewees with high compared to low working memory capacity. Bootstrapping was chosen to ensure the robustness of the test given the uneven sample size per condition. As predicted, AUC values of observer discrimination accuracy were significantly higher for White interviewees with high (AUC = .65, 95%CI [.612, .679]) than low working memory capacity (AUC = .50, 95% CI [95% CI: 0.455-0.536], D = 5.57, $p < .001$. On the other hand, the difference in observer AUC values for Hispanic interviewees with high working memory capacity (AUC = .55, 95% CI [.520, .587]) and those with low working memory capacity (AUC = .55, 95% CI [.465, .541]) did not attain statistical significance, D[12] = 1.94, p = .053. Figure 8 presents this interaction.

[12] D = (AUC1-AUC2)/s. The standard error of the difference (s) between the two AUCs is determined using bootstrapping (Robin et al., 2011).

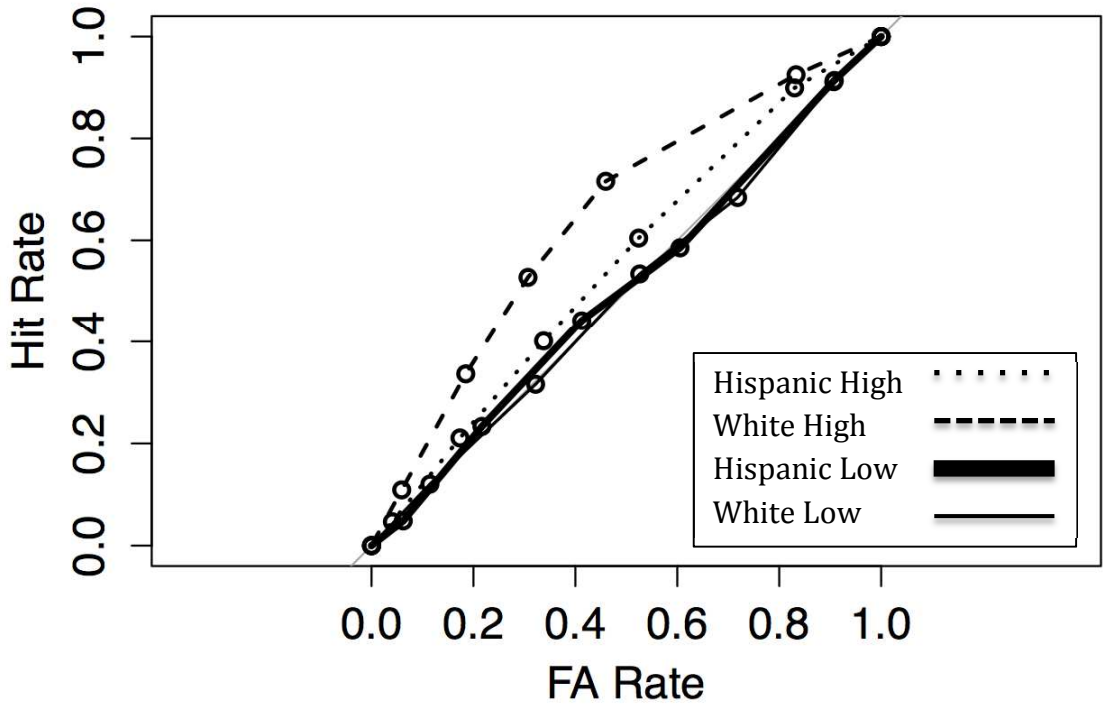

Figure 8. ROC Curves Experiment 1d. Observers' ratings for Hispanic and White interviewees at low or high levels of working memory capacity, plotted as a function of hit and false alarm rates. The grey line indicates chance-level responding.

Observer discrimination accuracy between liars and truth-tellers differed as a function of veracity, ethnicity and working memory capacity. That is, observer discrimination accuracy was improved for individuals with high working memory capacity but only for White but not Hispanic interviewees. These results align with the hypothesis that liars and truth-tellers experiencing stereotype threat rely on certain components of working memory, such as the phonological loop, to a similar degree. When access to the phonological loop is reduced, such as is the case when experiencing stereotype threat, liars and truth-tellers were affected similarly. That is, Hispanic liars and truth-tellers appeared similar to one another at both high and low working memory capacities.

Chapter 6: Experiment 2a

The results from Experiment 1 are consistent with hypotheses. When a group (e.g., Hispanics) feels stigma regarding police interaction, an inter-ethnic interview can prime these feelings and increase stereotype threat, cognitive load, and arousal as a consequence. However, the design of Experiment 1 was not completely crossed; Whites appeared with a same- but not cross-ethnic interviewer. Experiment 2 was designed to address this issue by replicating Experiment 1 but with a Hispanic interviewer. It was predicted that a Hispanic interviewer would not elicit a high-degree of stereotype threat in Hispanic or White interviewees. According to Najdowski (2011), simply being accused of a crime may elicit threat. However, the same-ethnicity interviewer should not add to this threat-level, and may possibly reduce the experience of threat. Whites are not predicted to have a stigma regarding police interaction, so an interethnic interview should not prime any feelings of threat. When interviewed by a Hispanic, Hispanics and Whites are predicted to experience similar levels of arousal and cognitive load. As a consequence, liars should display more behavioral cues related to deception than truth-tellers (e.g., liars should appear as though they are thinking harder than truth tellers), producing similar rates of discrimination accuracy for White and Hispanic participants. Experiments 2a, b, c, and d test these hypotheses.

Methods

Participants and design. Approximately 36 community members and university students were recruited as *participant-interviewees* (i.e., interviewee; *M* age = 23, *SD* = 7.39; 41% male; 22 Hispanic) using the same recruitment procedure as in Experiment 1a. Of these 36, a total of 32 interviews were recorded and used as stimuli in Experiment 1.

A total of four interviews were excluded because the eyes or face of the interviewee were obstructed by hair or small eyeglasses (n = 2), they indicated an eye disorder that caused their gaze to shift inadvertently (n = 1), or their interview did not record properly (n = 1). The same incentives were used as in Experiment 1a.

A Hispanic interviewer (22 years of age) was recruited and trained for Experiment 2. The same training protocol was used as in Experiment 1a. The Hispanic interviewer was selected to be similar to the White interviewer on traits besides ethnicity (age, gender, background knowledge regarding interview techniques such that both interviewers were criminal justice majors who had worked with law enforcement).

This was a 2 (Veracity: Lie or Truth) x 2 (Ethnicity: Hispanic or White) between-subjects design. The same dependent variables were used as in Experiment 1a.

Materials and Procedure. The materials and procedure were identical to Experiment 1a, except that a different interviewer was trained.

Results

A total of 32 interviewees participated, eight in each condition. Approximately 41% were male. The mean age of interviewee was 27 years (SD=13.91). The mean age was similar for liars and truth-tellers, but Hispanics liars and truth-tellers were younger on average than Whites liars and truth-tellers. Hispanic liars (3 males) had a mean age of 20.00 (SD = 2.31); White liars (4 males) had a mean age of 32.88 (SD = 16.47); Hispanic truth-tellers (3 males) had a mean age of 22.88 (SD = 3.72); White truth-tellers (3 males) had a mean age of 32.13 (SD = 20.04).

Motivation. Participants were asked to rate on a 1 (not at all motivated) to 7 (very motivated) scale "How motivated were you to convince the interviewer that you were

truthful." On average, participants were motivated to appear truthful during the interview (M = 6.06, SD = 1.34). Critically, a 2 (Interviewee Ethnicity) x 2 (Veracity) between-subjects Analysis of Variance (ANOVA) suggested that there were no significant differences in motivation ratings across conditions (F values < 2.54).

Stereotype threat. Three measures of stereotype threat were assessed and analyzed. Two self-report scales on stereotype threat were analyzed to examine the extent to which participants experienced threat during the interview (interview experience scale) and during their life (general scale). A memory questionnaire was analyzed separately to examine cognitive vigilance, the extent to which participants attended to stereotype-relevant cues during their interview.

Stereotype threat experience scales. Two items on the general-threat scale were reverse coded prior to analyses so that higher averages on each scale indicated higher perception of stereotype threat. As in Experiment 1, means were computed on two scales for each participant: (a) general threat and (b) threat of interview experience

Univariate and multivariate assumptions were investigated prior to analyses by examining standardized values, skew and kurtosis values, and a Mahalanobis distance test. There were no significant violations to assumptions. A 2 (Interviewee Ethnicity) x 2 (Veracity) between-subjects MANOVA suggests that Hispanics reported experiencing more stereotype threat than Whites on the interview experience scale (Hispanic: M = 3.15, SD = 1.77; White: M = 1.50, SD = .63) and general experience scale (Hispanic M = 3.96, SD = .98; White: M = 2.76, SD = .81), $F(2,27) = 10.22$, $p < .001$, $d = 1.17$. Univariate tests suggest that Hispanics reported significantly more threat than Whites on general, $F(1, 31) = 13.86$, $p = .001$, $d = 1.33$) and interview experience scales, $F(1, 31)$ =

11.57, $p = .002$, $d = 1.24$). No other significant effects were detected (F's < 1; See Figure 9).

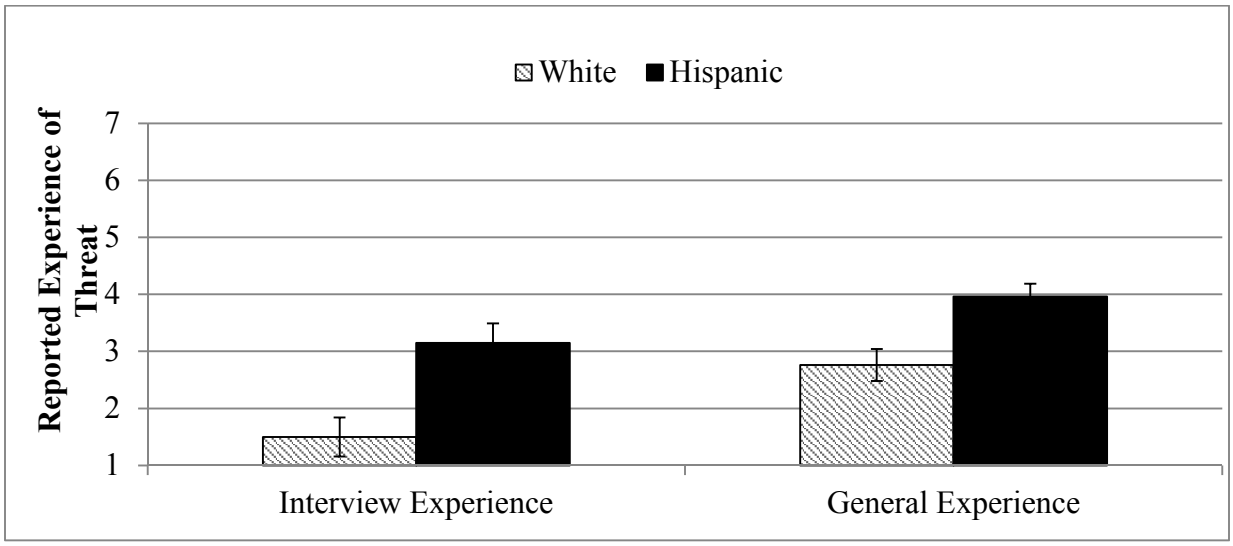

Figure 9. Stereotype Threat Scales Experiment 2a. Average scores on stereotype threat scales when interviewed by a Hispanic, as a function of interviewee ethnicity and scale-type (general or interview experience). Error bars represent standard error of the mean.

Cognitive vigilance. Cognitive vigilance was computed on a 27-point scale, with higher values indicating higher vigilance to possible stereotype cues. Univariate and multivariate assumptions were investigated prior to analyses by examining standardized values, and skew and kurtosis values. There were no significant violations to assumptions. A 2 (Interviewee Ethnicity) x 2 (Veracity) between-subjects ANOVA was conducted on vigilance scores. White interviewees ($M = 19.97$, $SD = 3.51$) had significantly lower vigilance than Hispanic interviewees ($M = 23.22$, $SD = 2.94$), $F(1, 31) = 7.53$, $p = .010$, $d = 1.00$. There were no other significant effects (F's < 1). These results contradict Experiment 1 results, where White interviewees had higher vigilance scores than Hispanics, further suggesting that the 15-item memory scale used in this study may not be a valid measurement of vigilance.

Cognitive Load. Cognitive load was assessed using: (a) error rates and reaction time on the same secondary task used in Experiment 1, which participants completed during their interview, and (b) the same six self-report questions regarding experiences during the interview answered during the post-interview-questionnaire phase.

Secondary task: error and RT. Univariate and multivariate assumptions were investigated prior to analyses by examining standardized values, and skew and kurtosis values. One participant was detected as an outlier with an error rate of 67%. This person was removed from all analyses. Two separate 2 (Interviewee Ethnicity) x 2 (Veracity) x 2 (Time: baseline or interview) mixed design ANOVAs with time as the within-subjects variable were conducted on error rate and RT data.

Error rate was calculated as the proportion of times a participant incorrectly responded (missed detecting the tone, or incorrectly indicated the tone was present) out of the total number of trials at baseline (for baseline error rate) or during the interview (for interview error rate)[13]. There was a significant main effect of time; as predicted, participants made more errors during the interview (M = .22, SD = .13) than baseline (M = .04, SD = .05), $F(1, 31) = 53.94$, $p < .001$, $d = 1.83$. There were no other significant effects (F's < 3.49). The second measure of cognitive load was RT. There were no significant effects for RT (F's < 2.08). Unlike in Experiment 1, no significant speed-accuracy tradeoff was detected for Hispanic interviewees in Experiment 2. This supports the prediction that a Hispanic interviewer reduced overall load experienced for Hispanic interviewees, such that Hispanic and White interviewees showed similar error and RT rates on the secondary task. However, there were also no significant differences between

[13] During their interview, two participant's secondary tasks stopped recording during the reverse order question.

truth-tellers and liars. Table B in Appendix 6 includes mean and standard deviations for RT and error rate in each condition.

Self-Report. Participants' answers to six questions regarding the difficulty of the interview on 1(not at all difficult) to 7(very difficult) scales were analyzed. Univariate and multivariate assumptions were investigated prior to analyses by examining standardized values, skew and kurtosis values, and a Mahalanobis distance test. There were no significant violations to assumptions. A 2 (Interviewee Ethnicity) x 2 (Veracity) between-subjects Multivariate Analysis of Variance (MANOVA) detected no significant multivariate (F's < 1.29) or univariate (F's < 2.81) main effects or interaction. Table C in Appendix 6 includes means and standard deviations for all conditions.

Arousal and anxiety. Measures of arousal and anxiety were analyzed using (a) Mean Arterial Blood Pressure and Pulse measured before and after the interview, and (b) STAI responses measured at three points across participation (before the mock-crime, before the interview, and at the conclusion of their participation). Anxiety was included because it is hypothesized to be one mechanism that accounts for the debilitating effects of stereotype threat on performance: as threat increases, anxiety increases also, reducing access to executive functioning resources as a consequence (Najdowski, 2011; Schmader, et al., 2008).

Mean Arterial Blood Pressure (MABP) and pulse. One participant was removed because their blood pressure and pulse did not record due to an equipment failure following the interview. Univariate and multivariate assumptions were investigated prior to analyses by examining standardized values, skew and kurtosis values, and a Mahalanobis distance test. There were no significant violations to assumptions. MABP

was calculated using the same method as in Experiment 1a. A separate 2 (Ethnicity) x 2 (Veracity) x 2 (Time: pre- or post-interview) mixed-design ANOVA, with time as the within-subjects variable was computed for MABP values and for pulse values. There were no significant effects of MABP (F's < 1.47) or pulse (F's < 1). Table D Appendix 6 includes the means and standard deviations in each condition.

State-Trait Anxiety Index (STAI). STAI scores were computed by summing across all items. Values on this scale range from 20 (indicating low levels of anxiety) – 80 (indicating high levels of anxiety). Nine items were reverse coded prior to analyses. Univariate and multivariate assumptions were investigated prior to analyses by examining standardized values, skew and kurtosis values, and a Mahalanobis distance test. There were no significant violations to assumptions. A 2 (Ethnicity) x 2 (Veracity) x 3 (Time: pre-mock crime, pre-interview, post-interview) mixed-design ANOVA, with time as the within-subjects variable was conducted on STAI values. As predicted there was a significant main effect of time, $F(2, 27) = 13.84$, $p < .001$, $d = 1.36$. Pairwise comparisons suggest that participants were significantly less anxious pre-mock crime ($M = 38.31$, $SD = 6.23$) than immediately pre-interview ($M = 41.28$, $SD = 10.17$), $p = .046$, $d = .35$. Participants were significantly less anxious post-interview ($M = 32.03$, $SD = 8.13$) than at pre-interview, $p < .001$, $d = 1.00$, or pre-mock crime, $p < .001$, $d = .87$. No other effects were significant (F's < 3.62). As predicted, all participants reported the highest levels of anxiety immediately before the interview. Table E in Appendix 6 includes the means and standard deviations in each condition.

The hypotheses for Experiment 2a were partially supported. First, self-report measures of threat across two different scales indicate that with a Hispanic interviewer,

Hispanic interviewees experienced more stereotype threat than did White interviewees and reported experiencing more stereotype threat generally. The inclusion of a Hispanic interviewer did not reduce the self-reported experiences of threat by the Hispanic interviewees. However, as predicted Hispanic and White interviewees experienced similar levels of arousal and cognitive load as indicated by objective measurements of these constructs. Contrary to hypotheses, no significant differences in cognitive load or arousal were detected between liars and truth-tellers. In this sample the interview did not elicit significant differences in cognitive load to a level detectable by the objective measures used in this study.

Chapter 7: Experiment 2b

The results of Experiment 2a partially support the hypothesis that in the presence of a Hispanic interviewer some of the perceived threat experienced during a cognitively demanding interview was less than with a White interviewer, especially for Hispanic interviewees. Hispanic interviewees reported experiencing more stereotype threat than White interviewees, but their cognitive load and arousal levels were not significantly different. Experiment 2b examines the extent to which deceptive appearance (as defined by behavioral cues in Experiment 2b and deception detection accuracy in Experiment 2c) is affected by the contextual conditions of the interviewees. Najdowski (2011) suggested that stereotype threat, mediated by cognitive load, would exacerbate deceptive behaviors in interviewees. Given that the objective measurements of stereotype threat (e.g., MABP) and cognitive load (e.g., error rate and RT) showed no significant differences between Hispanic and White participants, it is predicted that the behavioral differences between Hispanic liars and truth-tellers should be similar to the difference between White liars and truth-tellers. Third-party observers watched video-recorded interviews from Experiment 2a and rated interviewees on six behavioral cues.

Methods

Participants and design. A total of 132 observers (*M* age = 34.44, *SD* = 10.97; 52% female; 70% White, 13% African American, 13% Hispanic, 8% Asian, and 2% identified as Mixed Ethnicity or European[14]) watched videos of interviews collected in Experiment 2a. Data were collected using Amazon's Mechanical Turk (M-TurK). This study was a 2 (Veracity: truth or lie) x 2 (Ethnicity: Hispanic or White) within-subjects

[14] A total of three participants chose not to identify their ethnicity.

design using ratings of three composite variables (appearance of cognitive load, anxiousness, and confident and convincing appearance) as the dependent variable.

Materials and Procedure.

Videos of interviewees. The materials and procedure were created identically to Experiment 1, using the new set of videos collected during Experiment 2a. The average length of video clips was 101 seconds (*SD* = 41), ranging from 56 to 242 seconds. The average length of time was similar across all conditions (Hispanic Liars: *M* = 97, *SD* = 30; Hispanic Truth-Teller: *M* = 107, *SD* = 47; White Liars: *M* = 93, *SD* = 25; White Truth-Tellers: *M* = 117, *SD* = 55).

Behavior ratings of videos by third-party observers. The observers rated the interviewees on the same six behavioral cues to truth and deception as in Experiment 1 (see Table 5).[15]

Results

Ratings data were analyzed using 2 (Veracity) x 2 (Ethnicity of Interviewee) within-subjects ANOVAs on each of three separate dependent variables, (a) cognitive load, (b) anxiousness, and (c) confident and convincing appearance. As predicted, liars (*M* = 3.89, *SD* = .77) were rated as experiencing more cognitive load than truth-tellers (*M* = 3.60, *SD* = .90), $F(1, 131) = 30.07$, $p < .001$, $d = .35$. There were no other significant effects (*F*'s < 1.77).

[15] Cronhbach alpha values were examined between to-be-combined items. All alphas were greater than .78

Table 5.
Experiment 2 Hispanic Interviewer: *Third-Party Observers '* (*N* = 132) *Ratings on Six Behavioral Cues, as a Function of Cluster, Interviewee Race and Veracity Condition*

	Hispanic Interviewees						White Interviewees					
Behavioral Cues	Truth		Lie				Truth		Lie			
	M	*SD*	*M*	*SD*	*t*	*d*	*M*	*SD*	*M*	*SD*	*t*	*d*
Cognitive Load	**3.61**	**1.02**	**3.96**	**.84**	**5.02****	**.37**	**3.59**	**.97**	**3.83**	**.92**	**3.02***	**.25**
1. How hard does the interviewee appear to be thinking?	4.00	1.18	4.06	.99	.79	.06	3.90	1.07	3.98	1.07	.91	.07
2. How much difficulty does the interviewee appear to be having in answering the questions?	3.23	1.04	3.85	.90	7.67**	.64	3.29	1.06	3.68	1.02	4.60**	.37
Anxiousness	**3.08**	**1.09**	**3.55**	**.97**	**6.52****	**.46**	**3.02**	**1.07**	**3.40**	**1.03**	**4.98****	**.36**
3. How anxious does the interviewee appear?	3.12	1.09	3.66	1.04	6.82**	.51	3.08	1.07	3.47	1.05	4.81**	.37
4. How stressed does the interviewee appear?	3.03	1.14	3.44	1.00	5.26**	.38	2.95	1.12	3.33	1.08	4.64**	.35
Confident and Convincing Appearance	**4.79**	**.77**	**4.24**	**.74**	**7.76**	**.73**	**4.88**	**.82**	**4.54**	**.82**	**4.56**	**.41**
5. How confident does the interviewee appear?	4.58	.85	4.06	.79	6.36**	.63	4.71	.87	4.49	.88	2.69*	.25
6. How convincing is the interviewee's story?	5.01	.84	4.41	.86	7.53**	.71	5.06	.89	4.60	.98	5.50**	.49

Note. All ratings were on a 1 (not at all) -7 (very) scale. Cognitive demand and anxiousness involve cues to deception. Confidence and convincing appearance indicate cues to truthfulness. Bolded rows represent the mean of each cluster (e.g., Cognitive Load).

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

Also as predicted, liars (M = 3.56, SD = .93) were rated as more overall anxious than truth-tellers (M = 3.10, SD = .98), $F(1, 131) = 70.02$, $p < .001$, d = .48. Hispanics (M = 3.47, SD = 1.02) were rated as more anxious than Whites (M = 3.14, SD = .96), $F(1, 131) = 4.25$, $p = .041$, d = .33. The interaction was non-significant, $F < 1$.

Contrary to predictions, there was a significant interaction of veracity and ethnicity for confident and convincing appearance, $F(1, 131)= 4.38$, $p = .038$, d = .37. The magnitude of difference between Hispanic truth-tellers (M = 4.79, SD = .77) and liars (M = 4.24, SD = .74), $t(131) = 7.76$, $p < .001$, d = .73 was greater than between White truth-tellers (M = 4.88, SD = .82) and liars (M = 4.54, SD = .82), $t(131) = 4.56$, $p < .001$, d = .41 .

Across cues to deception, observers rated liars as displaying more cues related to cognitive load, anxiousness, and less cues related to confident and convincing appearance than truth-tellers, regardless of interviewee ethnicity. While liars and truth-tellers did not significantly differ on objective measurements of cognitive load or arousal in Experiment 2a, implicit judgments of load suggest that the interview was successful in eliciting several discernable differences between liars and truth-tellers in behavioral cues to deceit. The relationship between ethnicity and veracity in predicting observer behavior ratings was unclear. For cues to deception, the difference in observer behavioral ratings for liars and truth-tellers did not differ for Hispanic or White interviewees. These results are consistent with the non-significant differences between Hispanic and White interviewees on the objective measurement results (e.g., error and RT on the secondary task) in Experiment 2a. However for truthful behaviors, the magnitude of the difference between

liars and truth-tellers was greater for Hispanic than White interviewees for truthful cues (confident and convincing appearance).

Chapter 8: Experiment 2c

The results of Experiment 2b partially align with the results of Experiment 2a; the magnitude of difference between liars and truth-tellers was similar for Hispanic and White interviewees for cues to deception (cognitive load and anxiety), but not cues to truth (confident and convincing appearance). Given the results of Experiment 2b, it is predicted that Hispanic liars and truth-tellers should be discriminated at similar accuracy rates to White liars and truth-tellers. The purpose of Experiment 2c was to investigate this hypothesis by examining observer accuracy detecting truths and lies.

Methods

Participants and design. A total of 134 observers (49 % female; *M* age = 34.62, *SD* = 11.39; 76% White, 7% African American, 9% Hispanic, 8% Asian) watched videos of interviews collected in Experiment 2a. Observers were recruited using Amazon's Mechanical Turk (M-Turk)[16]. This study was a 2 (Veracity: truth or lie) x 2 (Ethnicity: Hispanic or White) within-subjects design using two dependent variables: observer's continuous ratings of veracity, and a signal detection measure AUC values of the ROC curves as a function of ethnicity.

Materials and procedure.

Videos. Videos and survey sets were identical to those used in Experiment 2b.

Veracity ratings. Observers rated the interviewees on deceptive appearance on a 1 (definitely truthful) to 6 (definitely lying) scale.

[16] A total of 14 observers were removed prior to analyses because their location was identified as outside the United States, or their IP address was identified twice. No participants failed the attention check. The same attention check was used as in Experiment 1.

Results

Veracity ratings. Observer ratings of interviewees from Experiment 2a were analyzed using a 2 (Veracity) x 2 (Ethnicity) within-subjects ANOVA. Higher average ratings indicate greater deceptive appearance. Univariate and multivariate assumptions were investigated prior to analyses by examining standardized values, skew and kurtosis values, and a Mahalanobis distance test. One possible univariate outlier was detected; their standardized z-score on veracity ratings ($z = 3.31$) exceeded the recommended cutoff value (greater than +/- 3.29, $p < .001$, two-tailed) by Tabacnick and Fidell (2007, p. 79). A sensitivity analysis suggested no significant differences were found when excluding or including this participant; all participants were left in for the remaining analyses. As predicted there was a significant main effect of veracity such that liars ($M = 3.21$, $SD = .71$) were rated significantly more deceptive than truth-tellers ($M = 2.93$, $SD = .70$), $F(1, 133) = 21.59$, $p < .001$, $d = .40$. There were no other significant effects (F's < 1.38). Table F in Appendix 6 includes the means and standard deviations in each condition.

Proportion of correct lie-truth responses. Observer's proportion of correct lie-truth judgments for Hispanic and White interviewees was examined. Proportion of correct judgments was computed by taking the sum of the correctly identified lies and truths out of the total number of lie and truth observations. In Experiment 2c the proportion of correctly identified liars and truth-tellers was not significantly different when observing White (about 54% accurate) and Hispanic interviewees (about 52% accurate), $Z = .476$, $p = .634$. Further, observers were equally likely to misidentify Hispanic (34%) and White (36%) truth-tellers as liars, $Z = .64$, $p = .521$. Contrary to

hypotheses, the cognitive load interview in Experiment 2 did not improve detection accuracy significantly above the 54% accuracy rate reported by an extensive meta-analysis averaging across many judgments (Bond & DePaulo, 2006).

ROC analyses. Receiver Operating Characteristic (ROC) curves were calculated to compare differences in discrimination accuracy between Hispanic interviewees and White interviewees by computing AUC values for each ROC curve. After plotting the ROC curve for a given distribution, the AUC was computed to investigate the overall classification accuracy for observers at classifying signal trials (liars) and noise trials (truth-tellers), with higher values indicating better overall classification accuracy. Given the within-subjects design, hit rate and false alarm rates were computed using the total number of observations of liars (n = 536) and truth-tellers (n = 536). There was no significant difference between AUC values when observers were judging White (AUC = .553, 95% CI [.520, .587]) or Hispanic (ACU = .553, 95% CI [.520, .587]) interviewees, Z = .097, p = .923.

Consistent with predictions and the results of Experiments 1 and 2a and 2b, a cognitively demanding interview produced above-chance levels of discrimination accuracy only when liars were experiencing more cognitive load and arousal than truth-tellers. In Experiment 2, Hispanic liars and truth-tellers experienced similar levels of load and arousal to White liars and truth-tellers (Experiment 2a). Critically, Hispanic and White liars were rated as significantly more cognitively loaded and anxious than truth-tellers. Consistent with the prediction that Hispanics experience less stereotype threat in the context of an inter-ethnic interview, Hispanics reported that they felt more threatened

and stereotyped during the interview than Whites, but did not significantly differ from Whites on objective measures of arousal or cognitive load.

Chapter 9: Experiment 2d

The results of Experiments 2a, b, and c offer partial support for the hypothesis that Hispanic interviewees experienced more comparable levels of stereotype threat to White interviewees when interviewed by an interviewer of the same ethnicity. Contrary to predictions, all interviewees experienced similar levels of cognitive load and arousal. In Experiment 2a liars and truth-tellers had similar error rates and RT on a secondary task, and had similar average MABP and pulse levels. As a consequence, observer discrimination accuracy was only at about chance levels for Hispanic and White interviewees in Experiment 2c. However, observer ratings of cognitive load and anxiousness in Experiment 2b offer some support for the point that the interview was more cognitively demanding for liars than truth-tellers; liars were rated as significantly more anxious and more cognitively loaded than truth-tellers regardless of interviewee ethnicity.

The purpose of Experiment 2d was to unpack the relationship between interviewee ethnicity and veracity by dividing interviewee's based on their working memory capacity. If the interview was of similar cognitive demand for liars and truth-tellers, then observer ratings of deceptive appearance were predicted to be similar for liars and truth-tellers at high or low levels of working memory. Further, if Hispanic and White interviewees experienced similar levels of stereotype threat, then observer ratings of deceptive appearance were predicted to be similar for White and Hispanic interviewees at high and low levels of working memory capacity.

Methods

Participants and design.

Observer's accuracy ratings from Experiment 2c (N = 134) were analyzed as a function of interviewee working memory capacity. The design of this study was a 2 (Interviewee Ethnicity: Hispanic, White) x 2 (Interviewee Veracity: Lie, Truth) x 2 (Interviewee Working Memory Capacity: High, Low) within-subjects design.

Materials and procedure.

Observer veracity ratings from Experiment 2c were re-analyzed in this Experiment. The critical difference was that the ratings data were divided based on *participant*-i*nterviewee*'s working memory capacity. Interviewee's were categorized as high or low working memory capacity using a median split analysis on participant-interviewees OSPAN scores (median OSPAN score: 38.5; M = 37.06 SD = 16.76). Higher scores on the OSPAN are related to higher working memory capacity (Conway et al., 2005).

Results

Veracity ratings. Observer's average veracity ratings of interviewees were analyzed using a 2 (Veracity) x 2 (Ethnicity) x 2 (Working Memory Capacity) within-subjects ANOVA. Univariate assumptions were investigated prior to analyses by examining standardized values, and skew and kurtosis values. There were no significant violations to assumptions. Bonferroni adjustments to α ($p < .05$) were used to correct for family-wise error rate, setting α equal to $p < .003$. There was a significant interaction between veracity and working memory capacity, $F(1, 133) = 13.14$, $p < .001$, $d = .63$. As

predicted, truth-tellers were rated similar regardless of their working memory capacity (High: $M = 2.99$, $SD = .92$; Low: $M = 2.86$, $SD = .83$), $t(132) = 1.39$, $p = .166$, $d = .15$; whereas liars with low working memory capacity ($M = 3.46$, $SD = .99$) were rated as more deceptive than liars with high working memory capacity ($M = 3.15$, $SD = .79$), $t(132) = 3.41$, $p = .001$, $d = .35$. There were no other significant effects (F's < 7.26; all p > .003) These results partially support the explanation that a Hispanic interviewer compared to a White interviewer presented a lower level of threat for Hispanic interviewees: regardless of ethnicity, the pattern of deceptive behaviors between liars and truth-tellers was similar at high or low levels of working memory capacity.

The results partially support that the interview was more cognitively demanding for liars than truth-tellers. For interviewees with low working memory capacity, liars were rated as more deceptive than truth-tellers; however, for interviewees with high working memory capacity, there was no significant difference between liars and truth-tellers (see Figure 10).

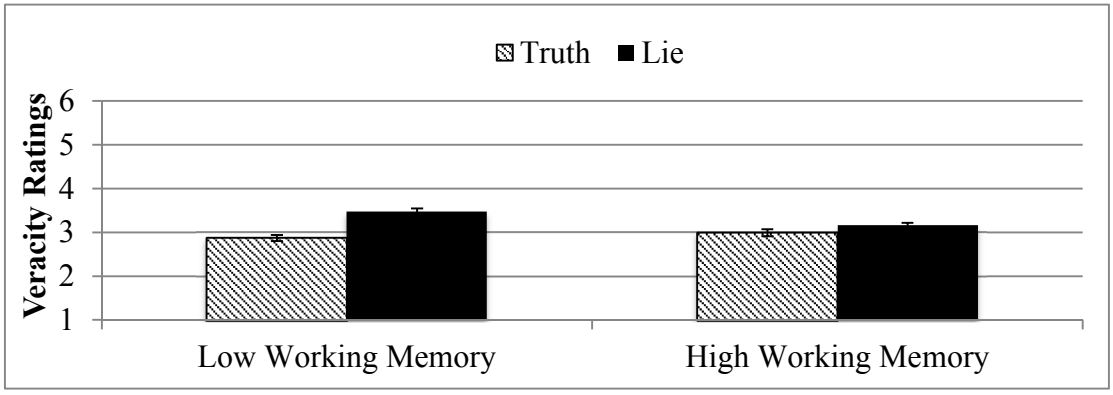

Figure 10. Observer Veracity Ratings Experiment 1d. Observers' average veracity ratings presented as a function of interviewee working memory capacity and veracity. Error bars represent standard error of the mean.

Although this pattern of data contrasts with Experiment 1, these results fit with the explanation that a cognitively demanding interview designed to overload the already-limited capacity systems of working memory, such as the interview task in this study, produced more cognitive overload for liars' with low working memory capacity. That is, the magnitude of difference in observer's veracity ratings was smaller for high than low working memory capacity interviewees

ROC analyses. The ratings data is useful to understand observers' averages for liars and truth-tellers, but it does not provide an analysis of discrimination accuracy – that is, are observers accurately discriminating between liars and truth-tellers at above chance levels as a function of veracity, ethnicity, and working memory capacity? To address this question, an ROC analysis was conducted to examine the interaction of these variables. A stratified bootstrap analysis with 2,000 resamples was used to compare AUC for the two critical comparisons: observer lie-truth discrimination accuracy for (a) Hispanic high compared to low working memory capacity interviewees and (b) White high compared to low working memory capacity interviewees. Observers had significantly higher AUC values when rating White interviewees with low (AUC = .621, 95%CI [.573, .668]) than high working memory capacity (AUC = .448, 95% CI [95% CI: .399, .497], $D = 5.10$, $p < .001$, whereas observer ratings of Hispanic interviewees were similar at high (AUC = .56, 95% CI [.513, .612]) and low levels of working memory capacity (AUC = .57, 95% CI [.522, 0.620]), D[17] = .24, $p = .814$. Figure 11 presents this interaction.

Contrary to hypotheses, observer ratings of deceptive appearance differed as a function of interviewee ethnicity and working memory capacity. Whereas when rating

[17] D = (AUC1-AUC2)/s. The standard error of the difference (s) between the two AUCs is determined using bootstrapping (Robin et al., 2011).

White interviewees, observers were less accurate detecting liars in the high than low working memory condition, when rating Hispanic interviewees observer ratings were similar when judging interviewees with high or low working memory capacity. As in Experiment 1, observers detected Hispanic liars and truth-tellers at similar, about-chance levels regardless of Working Memory Capacity. One explanation of these results is that Hispanic interviewees experienced higher levels of stereotype threat than Whites. However, the results of observers' behaviors ratings in Experiment 2b suggest that Hispanic and White interviewees did not necessarily experience differences in stereotype threat.

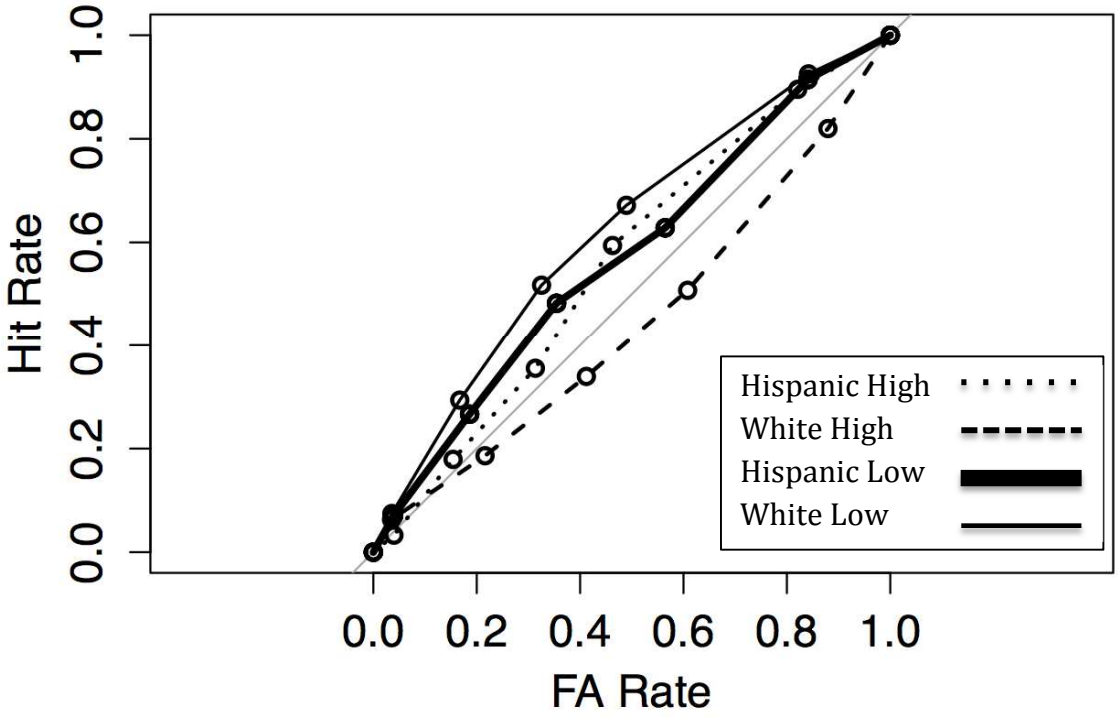

Figure 11. ROC Curves Experiment 2d. Observers' ratings for Hispanic and White interviewees at low or high levels of working memory capacity, plotted as a function of hit and false alarm rates. The grey line indicates chance-level responding.

Chapter 10: General Discussion

This study investigated differences in the cognitive processes and emotional experiences between liars and truth-tellers of different ethnicities during a cognitively demanding interview. Observer ratings of behavioral cues and judgments of veracity were also assessed to test whether cognitive processes and emotions impacted observers' accuracy discriminating between liars and truth-tellers of different ethnicities. Liars and truth-tellers experiencing similar levels of stereotype threat, cognitive load, and arousal were predicted to appear similarly deceptive and be detected with lower levels of accuracy when compared to liars experiencing more cognitive load than truth-tellers. The ethnicity of interviewee and interviewer were manipulated within the context of a realistic mock-crime scenario using objective measurements of underlying mechanisms (e.g., executive processes such as attention or working memory) that are typically more involved in deception than truth-telling. Across two experiments, the accuracy of observers' ability to discriminate between liars and truth-tellers was affected by the social context of the interview. As predicted, Hispanic interviewees experienced higher levels of arousal and cognitive load than White interviewees when interviewed by a White (Experiment 1) but not a Hispanic (Experiment 2) interviewer.

Cognitive Load Approaches to Detect Deception

The results of this study suggest that a cognitively demanding interview, such as the protocol used in this study, was effective to the extent that liars and truth-tellers' cognitive and emotional demands differed. When truth-tellers and liars experienced similar, and high levels of arousal and cognitive load, responding to a demanding interview elicited more cues to deception and fewer cues to truth regardless of veracity or

ethnicity. These similar cognitively demanding experiences reduced detection accuracy, increasing the rate of falsely identified truth-tellers when compared to liars and truth-tellers that differed on ratings of cognitive demand. Especially in Experiment 1 with a White interviewer, the magnitude of differences (cognitively and behaviorally) for liars and truth-tellers were smaller for Hispanic than White interviewees. Under these conditions, the ability to discriminate between liars and truth-tellers was also reduced. In Experiment 2, when Hispanic and White interviewees experienced more similar levels of arousal and cognitive load, third-party observers achieved similar rates of lie-truth discrimination accuracy. Further, as liars and truth-tellers in Experiment 2 experienced similar rates of cognitive load and arousal (as indicated by the non-significant differences in MABP and secondary task measures of error rate and RT), observers achieved lie-truth discrimination accuracy rates at about chance levels. Figures 11 and 12 compare effect sizes in Experiment 1 and 2 side-by-side as a function of interviewee ethnicity. The effect size (Cohen's *d*) represents the magnitude of difference between liars and truth-tellers average scores on composite values of cognitive load and arousal, and observer ratings of deception. In Experiment 1 (with a White interviewer), the difference between White liars' and truth-tellers' experiences of cognitive load, arousal, and observers' behavioral ratings of veracity were greater than Hispanics, whereas in Experiment 2 (with an Hispanic interviewer), the difference in cognitive load and arousal between liars and truth-tellers was smaller for both ethnicities when compared to Experiment 1.

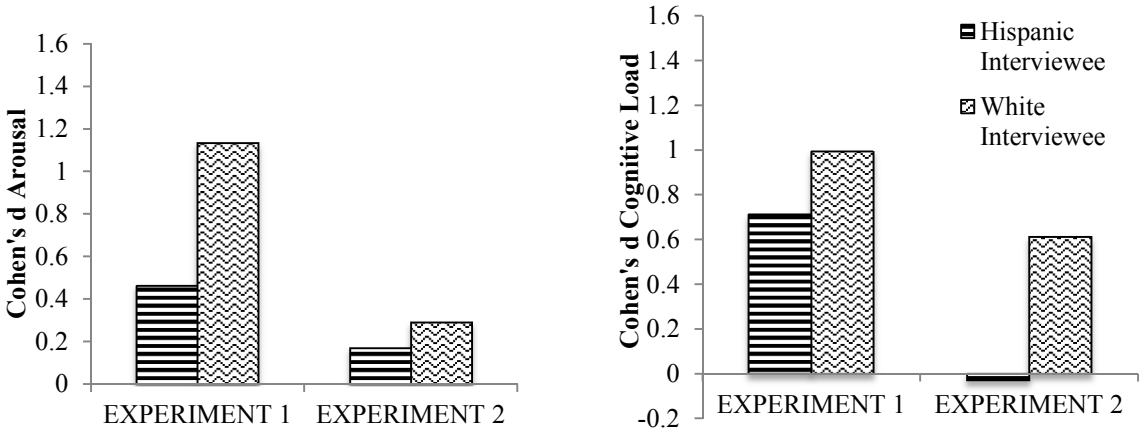

Figure 11. Effect sizes (Cohen's *d*), representing the difference between liars and truth-tellers average scores on composite values of arousal and cognitive load[18], are presented as a function of interviewee ethnicity and Experiment. Larger effect sizes indicate that liars had higher cognitive load or arousal composite scores.

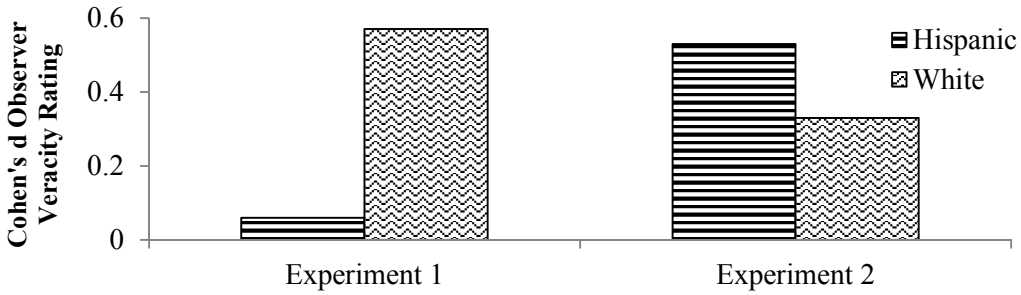

Figure 12. Effect sizes (Cohen's *d*), representing the difference in observers' veracity ratings (Experiments 1c and 2c) between liars and truth-tellers, are presented as a function of interviewee ethnicity and Experiment. Larger effect sizes indicate that observers rated liars as more deceptive than truth-tellers.

[18] Arousal and cognitive load composite scores were calculated by averaging the standardized scores of objective dependent variables measuring each construct. For arousal, this included mean arterial blood pressure (pre- and post-interview), pulse (pre- and post-interview), and STAI values (pre-mock crime, pre-, and post-interview). For cognitive load, this included reaction time values (interview), and error rates (interview) Reaction time values were reverse coded so that all positive values indicated "more" cognitive load.

In Experiments 1 and 2, observers' accuracy discriminating between liars and truth-tellers depended on an interviewees' working memory capacity. In Experiment 1, observers were worse at discriminating between liars and truth-tellers with high than low cognitive demand (Figure 13, left panel). However, when liars experienced stereotype threat, observers rated Hispanic liars and truth-tellers similarly regardless of working memory capacity (Figure 13, right panel). These results suggest that liars and truth-tellers may rely on certain components of working memory, such as the phonological loop, to a similar degree when experiencing stereotype threat, or when interviewees have lower working memory capacity generally (even White truth-tellers with low working memory capacity were rated more similarly to liars). When access to the phonological loop is reduced, such as is the case when experiencing stereotype threat; liars and truth-tellers were affected similarly. As a consequence, observers in this experiment were less able to discriminate between Hispanic liars and truth-tellers at high and low working memory capacity.

Observers showed different patterns of lie-truth discrimination accuracy when viewing interviewees in Experiment 2 (see right panels of Figure 13 and 14). When interviewed by a Hispanic in Experiment 2, White liars with low working memory capacity appeared more deceptive than truth-tellers. One notable difference between the results of these two experiments is that White liars and truth-tellers in Experiment 2 had more similar cognitive and emotional states than in Experiment 1 (see Figure 11); there was no significant difference between liars and truth tellers on the secondary task or arousal data. However, White liars were significantly more cognitive loaded and had higher MABP values than truth-tellers in Experiment 1. It is possible then that the level

of cognitive load experienced by liars interacts with working memory capacity, veracity, and ethnicity in predicting overall deceptive appearance.

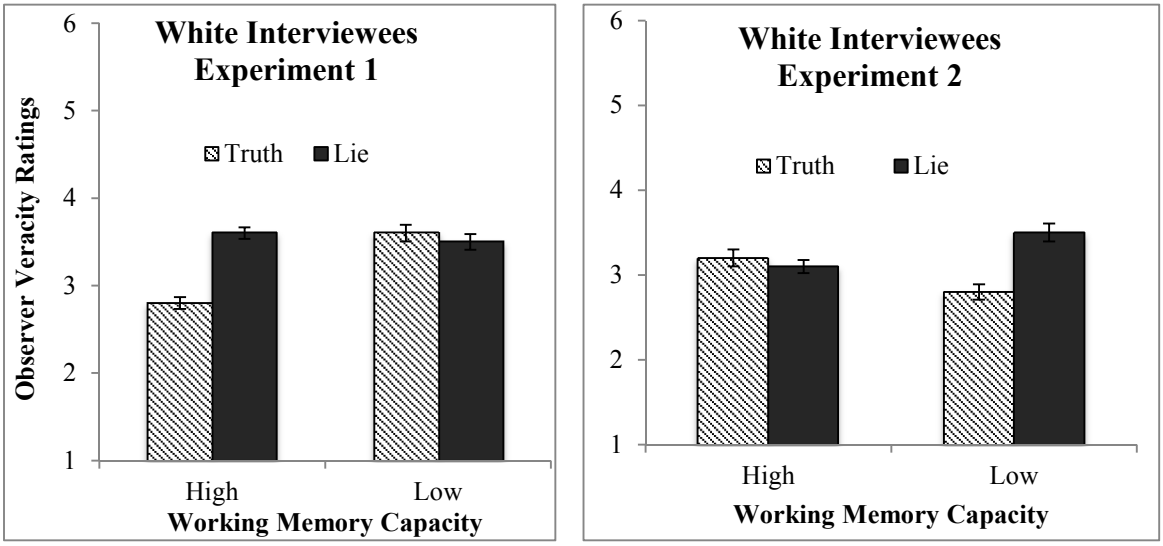

Figure 13. Observer veracity ratings of White Interviewees in Experiment 1 (left panel) and Experiment 2(right panel), presented as a function of interviewees working memory capacity and veracity.

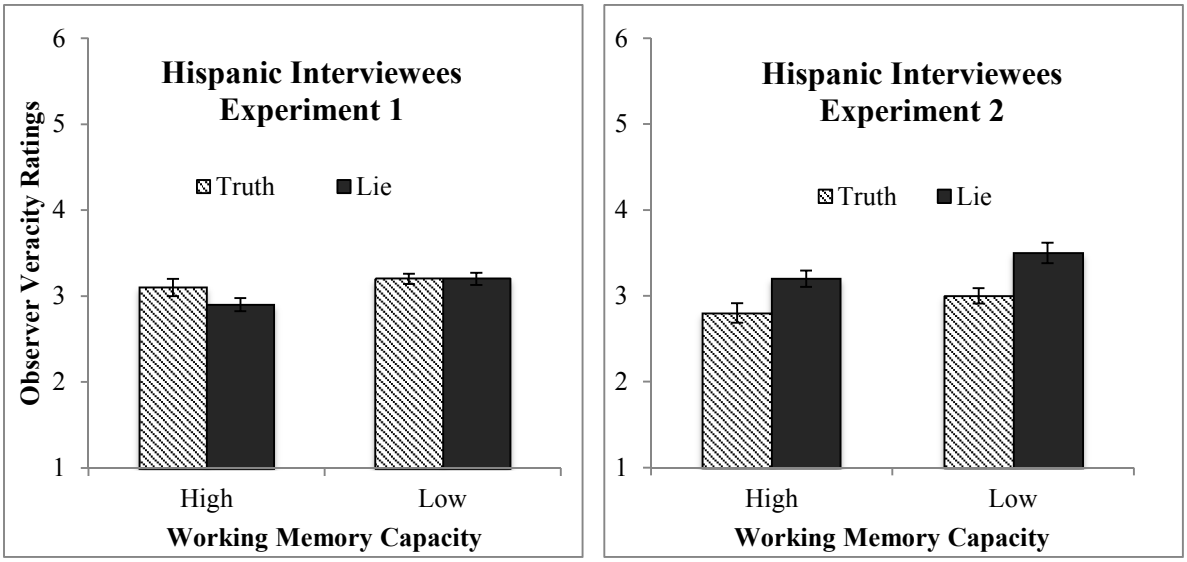

Figure 14. Observer veracity ratings of Hispanic Interviewees in Experiment 1 (left panel) and Experiment 2(right panel), presented as a function of interviewees working memory capacity and veracity.

These results align with explanations of working memory such as Unsworth and Engle (2007). According to Unsworth and Engle, individual differences in working memory are determined by (a) an individuals' ability to maintain goal-relevant information active in memory (referred to as "primary" memory), and an individuals' ability to apply effective memory strategies that aid retrieval of task-relevant information. Accordingly, in this study when the interview imposed higher cognitive demands on liars than truth-tellers, high working memory capacity liars may have been more easily detected than truth-tellers because their ability to use resource-intensive strategies that typically allow for effective detection evasion (e.g., impression management strategies such as keeping their story simple) was disrupted. Truth-tellers with low working memory capacity in a highly demanding interview (Experiment 1) may have also experienced more interference and been more distractible than truth-tellers with high working memory capacity. This is because all interviewees, including truth-tellers likely rely on working memory to search and retrieve information from memory that is task-relevant, control attention to the interviewers questions, and control responses on the secondary task (auditory detection task) used in this study (see Unsworth and Engle, 2007; Shipstead, Lindsey, Marshall, & Engle, 2014). For truth-tellers with lower working memory capacities, their working memory may have been overloaded more easily by these tasks than those with higher working memory capacities. Results of one line of research supports this explanation, such that individuals who have lower working memory capacity tend to have greater difficulty staying "on task" when attentional demands are high (see Kane and McVay, 2012 for a review). The results of these studies suggest that lower working memory capacity individuals are more prone to have off-task

thoughts and this distraction causes them to perform poorly on tasks that require a high degree of attentional control. Future studies may consider examining (a) the presence of task-unrelated thoughts in interviewees (especially truth-tellers), (b) strategy use of interviewees under high and low cognitive demands, (c) how cognitive load and arousal interact with working memory capacity in predicting observer discrimination accuracy, and (d) manipulating the level of cognitive load experienced by interviewees during an interview.

These findings extend a growing body of literature that encourages researchers to consider the boundary conditions and underlying mechanisms that create an effective cognitively loading interview (see Blandón-Gitlin, et al., 2014). In this study, a Cognitive Load Approach interview was successful only to the degree that liars' and truth-tellers experienced distinct cognitive and emotional differences. In Experiment 1 and 2, when interviewees experienced stereotype threat and higher levels of arousal, even truth-tellers cognitive resources were overloaded to a similar degree as liars. As a consequence, the effectiveness of a Cognitive Load Approach technique was undermined. The cognitive load experienced by a truth-teller appears important to consider when deciding on the level of cognitive demand to impose on an interviewee. For interviewees susceptible to stereotype threat, a cognitively demanding interview could make lie-truth discrimination accuracy more difficult, increasing the misidentification rate of truth-tellers at the cost of catching a liar.

Social-Contextual Factors: Ethnicity of Interviewee and Interviewer Matters

Do certain ethnic groups experience stereotype threat in the context of a criminal investigation? The results of this study suggest the Hispanic, more than White, suspects

believed they were stereotyped as criminals when accused of a mock-crime. In Experiment 1a and 2a Hispanics reported experiencing threat in the context of the interview and generally in their lives more often than Whites. In Experiment 1, Hispanic interviewees experienced higher levels of arousal (Hispanics had higher MABP values than Whites) and cognitive load (Hispanics had longer response times on the secondary task than Whites) than White interviewees when interviewed by a White interviewer. Further, Hispanic interviewees were judged as displaying more deceptive and less truthful behaviors than White interviewees, regardless of actual veracity. Consequently, Hispanic liars and truth-tellers in Experiment 1 were discriminated at lower levels of accuracy than Whites. These results are consistent with the predictions of research on stereotype threat in the context of criminal investigations, suggesting that individuals who perceive that they are stereotyped as criminals will experience stereotype threat when accused of a crime (Najdowski, 2011). In a cognitively demanding interview, this increase in threat and arousal amplified the cognitive load of interviewees, reducing the behavioral differences between liars and truth-tellers and decreasing discrimination accuracy between the two as a consequence. For innocent or guilty suspects with stigmas surrounding perception of their criminality, a different-ethnicity interviewer triggered stereotype threat and reduced ability to appear truthful during a cognitively demanding interview.

In Experiment 2, the use of a Hispanic interviewer showed some promise for reducing stereotype threat for Hispanic participants. With a Hispanic interviewer, Hispanic and White interviewees experienced similar levels of arousal and load. Further, observer ratings of cognitive load were similar for Hispanic and White interviewees.

With a Hispanic interviewer, the magnitude of difference in deceptive behaviors between liars and truth-tellers was similar for Hispanics and Whites. Hispanic and White liars and truth-tellers were detected at similar levels of accuracy. These results provide partial support that a same-ethnicity interviewer may reduce overall experiences of threat for individuals who are susceptible to threat when accused of a crime. An alternative explanation is that in the Hispanic interviewer condition, a White interviewee's experiences of threat were increased to a similar level of a Hispanic interviewee. According to Najdowski's (2011) model, Whites do not tend to believe that they are the targets of stereotyping in the context of a criminal investigation, such as in this study. Further, Whites' scores on the stereotype threat scales suggest that they reported similar and somewhat lower levels of perceived threat in Experiment 2 (Hispanic interviewer experience: $M = 1.50$, $SD = .63$) than Experiment 1 (White interviewer experience: $M = 2.53$, $SD=.68$). A more plausible explanation for these results is that the presence of a Hispanic interviewer reduced threat for Hispanic interviewees (both liars and truth-tellers). At this reduced level of threat, truth-tellers' cognitive load was lower than liars', but presumably both groups experienced similar amounts of diminishment in cognitive load when compared to experiencing higher levels of stereotype threat. Future studies may consider testing these hypotheses further.

The results of this study elucidate some of the mechanisms underlying stereotype threat in the context of a criminal investigation. As predicted by Najdowski (2011), stereotype threat was hypothesized to reduce an interviewee's working memory capacity by increasing arousal and increasing attention (vigilance) to irrelevant cues that may be perceived as indicative of stereotyping on the part of the interviewer. The increases in

distraction to irrelevant cues accompanied by an increase in arousal were hypothesized to overload the working memory of individuals, reducing behavioral performance on the given task (in this study, the task was to appear truthful and innocent). Hispanic interviewees under high threat (with a White interviewer in Experiment 1) experienced higher levels of arousal than White interviewees. However, the results of this study were inconclusive regarding other cognitive processes hypothesized to contribute to the relationship between threat and detection appearance such as cognitive vigilance. The vigilance scale data were inconsistent in Experiment 1 and 2. Other measures of vigilance may be more sensitive to differences in stereotype threat in the context of a criminal investigation. Examples of possible alternative measures for use in future studies include physiological vigilance (Murphy et al. 2007), speed at completing a given task (Seibt & Foerster, 2004), or questionnaires that ask direct questions about off-task thoughts such as the Verbal Thoughts Questionnaire (Beilock et al., 2007) or Cognitive Interference Questionnaire (Coy et al., 2011).

Applied Importance of these Results

The results of this study suggest that considering ethnic differences in the context of forensic interviewing is important. Social-contextual conditions (in this study, an interethnic interviewer of individuals who felt stigmatized as criminals) increased levels of experience stereotype threat and reduced the accuracy of cognitive load interview approach techniques as a consequence. Applying Cognitive Load Approach techniques to a forensic interview may be less successful when interviewee's are prone to stereotype threat. One promising result of this study is that a Cognitive Load Approach was successful at eliciting above-chance levels of lie-truth discrimination accuracy, especially

in Experiment 1 with a White interviewer. With a Hispanic interviewer in Experiment 2, lie-truth discrimination accuracy for low working memory capacity White interviewees was above chance levels, whereas all other conditions were at about-chance. One critical difference between the results of Experiments 1 and 2 is that the objective measures of cognitive load and arousal were significantly different between liars and truth-tellers in Experiment 1 but not Experiment 2. Using these direct and objective measurements in conjunction with a Cognitive Load Approach interview may provide a more complete picture of the actual veracity of an interviewee.

In the context of a criminal investigation the negative effects of experiencing stereotype threat may extend beyond lie detection. According to Davis and Leo (2012), stereotype threat could trickle down through other parts of the investigation. If officers determine that a suspect is lying during their pre-interrogation interview, they are more likely to subject the suspect to an interrogation. Suspects experiencing stereotype threat may experience a more rapid depletion of executive functioning resources than non-stereotype-threatened suspects. Stereotype threatened suspects may then display more "guilty" behaviors, increasing the likelihood that coercive tactics will be used during the interrogation and increasing the likelihood of falsely confessing to crimes as a consequence. Applying an accusatory interview, common in American police interviews, might amplify the levels of stereotype threat as experienced by Hispanics in this study. That is, under the conditions of a high pressure, accusatory interview, liars and truth-tellers experiencing stereotype threat may be even more likely to appear deceptive than what was found in this study. Besides ethnicity of interviewee, there are likely to be many more forms of stigma or stereotype threat that occur in the context of a criminal

investigation. For example, Davis and Leo (2012) suggest that a stepfather accused of molestation may experience stereotype threat during a forensic interview. Future studies may wish to compare detection accuracy for observers viewing an accusatory versus information-gathering interview of threatened and non-threatened group members, and investigate how different forms of stereotype threat act on these interview techniques.

Future Directions and Limitations

Experiment 2 partially supports the point that a same-ethnicity interview may reduce the level of experienced stereotype threat. Another effective method for reducing threat may be through rapport building. Rapport building is the process of creating trust and liking between an interviewer and interviewee. It is a common practice in therapy settings, and more recently has been suggested as a critical component in interrogation and intelligence gathering interviews. A rapport-building interview may reduce the perception that an interviewer is stereotyping them as a criminal by increasing liking of the interviewer, while at the same time increasing the likelihood that a guilty person will confess (e.g., Goodman-Delahunty,Martschuk, & Dhami, 2014).

Observer ethnicity was not manipulated in this study. Using a random sample from Amazon's M-Turk, a majority of the observers in this study were White (Experiment 1d: 70%; Experiment 2d: 75%). The ethnicity of observer and interviewee may affect detection accuracy. For example, Vrij and Winkel (1994) found that Black interviewees displayed nonverbal cues that were sometimes judged as more stereotypically deceptive by White observers. Further, multiple studies suggest that there are cultural differences in recognizing emotions (for a review see Elfenbein & Ambady, 2003). A post-hoc analysis suggests that the ethnicity of interviewee did not impact

observer ratings of behavioral cues or veracity. Eight separate 2 (Interviewee Ethnicity) x 2 (Interviewee Veracity) x 2 (Observer Ethnicity: White vs. Other Ethnicity) mixed-design ANOVA's, with Observer Ethnicity as the between-subjects variable, were conducted on the three behavioral cues and veracity ratings in Experiment 1 and Experiment 2. Results suggest that there were no statistically significant effects of observer ethnicity on veracity or behavioral ratings (Experiment 1: F's < 2.34; Experiment 2: F's < 3.46. Future research may investigate this by manipulating the ethnicity of the observer.

Manipulating the cognitive load of the interview may be another important factor for future research. This is because the level of cognitive load may differentially impact overall lie-truth discrimination accuracy; there may be a curvilinear relationship between cognitive load and accuracy. At "highest" and "lowest" levels of cognitive load liars and truth-tellers are hypothesized to look similar (making observers worse at discriminating between liars and truth-tellers), whereas at middle levels of cognitive load liars and truth-tellers are hypothesized to appear significantly different (making observers better at discriminating between liars and truth-tellers). In this study there was not a "lowest" level of cognitive load -- that is, everyone was subjected to an interview designed to be cognitively demanding. Participants experiencing "lowest" levels of cognitive load in this study may be more equal to some type of "middle" or "moderate" level of cognitive load. It is possible that participants fell into only the "middle" to "upper end" of the hypothetical cognitive load distribution. This would result in observing only the negative relationship between cognitive load and deception such that as load increases, accuracy decreases. The results in Experiment 1 partially support this hypothesis, such that for

Hispanics who showed higher levels of load regardless of actual veracity, liars and truth-tellers were not discriminated as accurately as White interviewees (who showed significant differences in load depending on veracity).

Another important issue to address in future studies is the dispositional characteristics of interviewers. Only one White and one Hispanic interviewer were used in this study. While the researchers attempted to make the interviewers as similar as possible across multiple attributes (age and background knowledge of criminal investigations), it is likely that individual differences existed that could not be controlled experimentally. Future studies should include multiple interviewers of each ethnicity and cross interviewers on multiple demographic characteristics such as gender, ethnicity, age, or personality. Another related limitation is that the interviewees were not asked to identify the actual ethnicity of interviewers. However, a pilot study was conducted in which undergraduate students (n = 38) answered a series of questions about the interviewers, including the critical question "what is the ethnicity of this person?" A majority of students were able to identify the ethnicity of both interviewers. A total of 87% of students correctly identified the ethnicity of the Hispanic interviewer, and 76% of students correctly identified the ethnicity of the White interviewer. These proportions did not significantly differ (p = .105). Thus, it appears that most interviewees in both studies were able to identify the interviewers' ethnicity. However, future studies may consider measuring this more directly by asking interviewees about interviewers' ethnicity immediately following an interview.

Conclusions

The results of this study suggest that stigmatized group members who perceive that they are stereotyped as criminals appear more deceptive and more guilty, regardless of actual veracity. A cognitively demanding interview was marginally successful in improving detection accuracy beyond chance levels (61%) when liars experienced high levels of cognitive load and arousal (Experiment 1). These results suggest that groups who experience threat and depletion may be more likely to experience diminished cognitive resources during an interview and appear guilty as a consequence. The implications of this research suggest that law enforcement officials and intelligence interviewers should be sensitive to possible stigmas of suspects.

References

Abe, N., Fujii, T., Ito, A., Ueno, A., Koseki, Y., Hashimoto, R., . . . Mori, E. (2014). The neural basis of dishonest decisions that serve to harm or help the target. *Brain and Cognition, 90*, 41-49. doi: 10.1016/j.bandc.2014.06.005

Abumrad, J. (Narrator). (2013, June 27). The Voices Inside You [Radiobroadcast episode]. In Ellen Horne (Producer), Radiolab. Washington, DC: National Public Radio.

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556-559. doi: 10.1126/science.1736359

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*(1), 1-29. doi:10.1146/annurev-psych-120710-100422

Baddeley, A. (2003). Working Memory: Looking Back and Looking Forward. *Nature Reviews Neuroscience*, *4*(10), 829-839. doi:10.1038/nrn1201

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin*, *74*(2), 81-99. doi:10.1037/h0029531

Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 33*(6), 983-998. doi:10.1037/0278-7393.33.6.983

Beilock, S. L. & Carr, T. H. (2005). When high-powered people fail: Working memory and "choking under pressure" in math. *Psychological Science*, *16*, 101-105. doi: 10.1111/j.0956-7976.2005.00789.x

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, *136*(2), 256-276. doi:10.1037/0096-3445.136.2.256

Berkman, E.T., Burklund, L., & Lieberman, M.D. (2009). Inhibitory spillover: Intentional motor inhibition produces incidental limbic inhibition via right inferior frontal cortex. *NeuroImage*, *47*, 705–712.

Blandón-Gitlin, I., Arrieta, B., Gombos, V., Mayberry, E. & Gerkens, D.R.(manuscript under review). A repeated questioning interview strategy induces cognitive load and improves lie detection. Manuscript submitted for publication

Blandón-Gitlin, I., Echon, R., & Pineda, E. (June, 2013). *Detecting Deception: The benefit of depleting executive control in liars.* Paper presented at the Bi-annual meeting of the Society for Applied Research in Memory and Cognition, Rotterdam, Amsterdam.

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

Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science, 12,* 225–229. doi: 10.1111/1467-9280.00340

Bond, G. D. (2012) Focus on basic cognitive mechanisms and strategies in deception research (and remand custody of 'wizards' to Harry Potter movies). *Journal of Applied Research in Memory and Cognition, 1*(2), 128–130. doi:10.1016/j.jarmac.2012.02.004.

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

Bond**,** C. F., & 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

Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology, 40,* 247–255. doi:10.1016/S0022-1031(03)00099-

Brown, B & Benedict, W. (2002). Perceptions of the police. *Policing* 25:543–80.

Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science*, *32*(1-2), 115-132. doi:10.1023/B:TRUC.0000021812.96911.c5

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk : A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science, 6*(1), 3-5. doi: 10.1177/1745691610393980

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

Carter, D. L. (1983). Hispanic interaction with the criminal justice system in Texas: Experiences, attitudes, and perceptions. *Journal of Criminal Justice*, *11*(3), 213-227. doi:10.1016/0047-2352(83)90115-0

Carter, D. L. (1985). Hispanic perception of police performance: An empirical assessment. *Journal of Criminal Justice*, *13*(6), 487-500. doi:10.1016/0047-2352(85)90078-9

Christ S.E., Van Essen, D.C., Watson, J.M., Brubaker, L.E., & McDermott, K.B. (2008). The contributions of prefrontal cortex and executive control to deception: Evidence from activation likelihood estimate meta-analysis. *Cerebral Cortex, 19,* 1557-66. doi: 10.1093

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786. doi: 10.3758/bf03196772

Costanzo, M. & Krauss, D. (2015). Forensic and legal psychology: Psychological science applied to the law. New York: Worth Publishers.

Coy, B., O'Brien, W. H., Tabaczynski, T., Northern, J., & Carels, R. (2011). Associations between evaluation anxiety, cognitive interference and performance on working memory tasks. *Applied Cognitive Psychology*, *25*(5), 823-832. doi:10.1002/acp.1765

Davis, D., & Leo, R. A. (2012). Interrogation-related regulatory decline: Ego depletion, failures of self-regulation, and the decision to confess. *Psychology Public Policy and Law, 18*(4), 673-704. doi: 10.1037/a0027367

Debey, E., Verschuere, B., & Crombez, G. (2012). Lying and executive control: An experimental investigation using ego depletion and goal neglect. *Acta Psychologica*, *140*(2), 133-141. doi:10.1016/j.actpsy.2012.03.004

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74-118. doi:10.1037/0033-2909.129.1.74.

Elfenbein, H. A., & Ambady, N. (2003). Universals and cultural differences in recognizing emotions. *Current Directions In Psychological Science (Wiley-Blackwell)*, *12*(5), 159. doi:10.1111/1467-8721.01252

Espinoza, R. E., & Willis-Esqueda, C. (2008). Defendant and defense attorney characteristics and their effects on juror decision making and prejudice against Mexican Americans. *Cultural Diversity and Ethnic Minority Psychology*, *14*(4), 364-371. doi:10.1037/a0012767

Evans, J. R., Houston, K.A., & Meissner, C.A. (2012) A positive, collaborative, and theoretically-based approach to improving deception detection. *Journal of Applied Research in Memory and Cognition, 1*(2), doi:10.1016/j.jarmac.2012.04.007

Evans, J.R., Meissner, C.A., Brandon, S.E., Russano, M.B., & Kleinman, S. M. (2010). Criminal versus HUMINT interrogations: The importance of psychological science in improving interrogative practice. *Journal of Psychiatry & Law, 38*(1/2), 215-249.

Evans, J. R., Michael, S. W., Meissner, C. A., & 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*(1), 33-41.

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191. doi:10.3758/BF03193146

Fenn, E., Blandón-Gitlin, I., Coons, J., & Pineda, C. The Inhibitory Spillover Effect: Controlling the Bladder Makes Better Liars. Manuscript submitted for publication.

Fenn, E., Mcguire, M., Langben, S., & Blandón-Gitlin, I. (March, 2014). *Imposing Cognitive Load Reduces Accuracy When Detecting Intentions.* Paper presented at the Annual Meeting of the American Psychology-Law Society, New Orleans, LA.

Frantz, C. M., Cuddy, A. C., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The race implicit association test as a stereotype threat experience. *Personality and Social Psychology Bulletin*, *30*(12), 1611-1624. doi:10.1177/0146167204266650

Gamer, M., Bauermann, T., Stoeter, P., & Vossel, G. (2007). Covariation among fMRI, skin conductance and behavioral data during processing of concealed information. *Human Brain Mapping, 28*(12), 1287–1301. doi:10.1002/hbm.20343.

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

Goff, P. A., Steele, C. M., & Davies, P. G. (2008). The space between us: Stereotype threat and distance in interracial contexts. *Journal of Personality and Social Psychology, 94*, 91-107. doi: 10.1037/0022-3514.94.1.91.

Gombos, V.A. (2006). The cognition of deception: The role of executive processes in producing lies. *Genetic, Social, and General Psychology Monographs. 132*, 197-214. doi:10.3200/MONO.132.3.197-214

Goodman-Delahunty, J., Martschuk, N., & Dhami, M. K. (2014). Interviewing High Value Detainees: Securing Cooperation and Disclosures. *Applied Cognitive Psychology*, *28*(6), 883-897. doi:10.1002/acp.3087

Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using receiver operating characteristic analysis. *Current Directions in Psychological Science*, *23*(1), 3-10. doi:10.1177/0963721413498891

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

Hu, X., Chen, H., & Fu, G. (2012). A repeated lie becomes a truth? The effect of intentional control andtraining on deception. *Frontiers in Psychology*, *3*: 488 doi:10.3389/fpsyg.2012.00488

Hu, X., Evans, A., Wu, H., Lee, K., & Fu, G. (2013). An interfering dot-probe task facilitates the detection of mock crime memory in a reaction time (RT)-based concealed information test. *Acta Psychologica*, *142*(2), 278-285. doi:10.1016/j.actpsy.2012.12.006

Hauch, V., Blandon-Gitlin, I., Masip, J., Sporer, S. (2014). Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and Social Psychology Review* Advance online publication. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25387767.

Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C (2005). *Essentials of the Reid Technique: Criminal Interrogations and Confessions*. Jones & Bartlett Publishers: Sudbury, MA.

Kane, M. J., & McVay, J. C. (2012). What Mind Wandering Reveals About Executive-Control Abilities and Failures. *Current Directions in Psychological Science, 21*(5), 348-354. doi: 10.1177/0963721412454875

Kassin, S. M., & Gudjonsson, G. H. (2004). The psychology of confessions: A review of the literature and issues. *Psychological Science in the Public Interest, 5*(2), 33-67. doi: 10.1111/j.1529-1006.2004.00016.x

Kleinman, S.M. (2006). KUBARK Counterintelligence Interrogation: Observations of an Interrogator. In Destro, R., et al. (Eds) *Educing Information: Interrogation Science and Art* (pp. 95-140). CreateSpace Independent Publishing Platform.

Lane, S. M., & Vieira, K. M. (2012). Steering a new course for deception detection research. *Journal of Applied Research in Memory and Cognition, 1*(2), 136-138. doi:10.1016/j.jarmac.2012.02.004.

Lane, S., Martin, S., Elliott, E., & Mennie, K. (March, 2014). *The impact of cognitive load and WMC on information elicitation*. Paper presented at the Annual Meeting of the American Psychology-Law Society, New Orleans, LA.

Masip, J., Barba, A., & Herrero, C. (2012). Behaviour Analysis Interview and common sense. A study with novice and experienced officers. *Psychiatry, Psychology and Law, 19*(1), 21-34. doi: 10.1080/13218719.2010.543402

Masip, J., Herrero, C., Garrido, E., & Barba, A. (2011). Is the Behaviour Analysis Interview just common sense? *Applied Cognitive Psychology, 25*, 593-604. doi: 10.1002/acp.1728

Masip, J., & Herrero, C. (2013). 'What would you say if you were guilty?' Suspects' strategies during a hypothetical Behavior Analysis Interview concerning a serious crime. *Applied Cognitive Psychology*, *27*(1), 60-70. doi:10.1002/acp.2872

Masip, J., & Herrero, C. (2014). Police detection of deception: Beliefs about behavioral cues to deception are strong even though contextual evidence is more useful. *Journal of Communication, 56*(1), 125-145. doi: 10.1111/jcom.12135

Meissner, C. A., Russano, M. B., & Narchet, F. M. (2010). The importance of a laboratory science for improving the diagnostic value of confession evidence. In G. Lassiter & C. A. Meissner (Eds.) *Police interrogations and false confessions: Current research, practice, and policy recommendations* (pp. 111-126). Washington, DC: American Psychological Association. doi:10.1037/12085-007

Menjívar, C., & Bejarano, C. (2004): Latino immigrants' perceptions of crime and police authorities in the United States: A case study from the Phoenix Metropolitan area. *Ethnic and Racial Studies, 27(*1), 120-148. doi: 10.1080/0141987032000147968

Mirande, A. (1980). Fear of crime and fear of the police in a Chicano community. *Sociology & Social Research*, *64*(4), 528-541.

Morgan III, C. A, Mishara, A., Christian, J., & Hazlett, G. (2008): Detecting deception through automated analysis of translated speech: Credibility assessments of Arabic speaking interviewees. *Journal of Intelligence Community Research and Development*, 1-22.

Murphy M. C., Steele, C. M. & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science, 18*(10), 879-885. doi: 10.1111/j.1467-9280.2007.01995.x

Mustard, D. B. (2001). Racial, ethnic, and gender disparities in sentencing: Evidence from the U.S. Federal Courts. *Journal of Law and Economics*, *44*(1), 285-314.

Najdowski, C. J. (2011). Stereotype threat in criminal interrogations: Why innocent Black suspects are at risk for confessing falsely. *Psychology, Public Policy, and Law, 17*(4), 562-591. doi: 10.1037/a0023741

Najdowski, C.J. (March, 2013) Stereotype threat in police encounters: Why African Americans are targeted as criminal suspects. Paper presented at the annual American Psychology-Law Society conference, Portland, OR.

Newton, J. (2006, April 17). The changing face of the LAPD. The Los Angeles Times. Retrieved from http://articles.latimes.com/2006/apr/17/local/me-lapd17

Niemann, Y. (2001). Stereotypes about Chicanas and Chicanos: Implications for counseling. *The Counseling Psychologist*, *29*(1), 55-90. doi:10.1177/0011000001291003

Paas, F., Tuovinen, J.E., Tabbers, H., & Van Gerven, P.W.M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63-71. doi: 10.1207/S15326985EP3801_8\

Reaves, B. A. (2010). *Local Police Departments, 2007*. Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics.

Richeson, J. A., & Shelton, J. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science*, *14*(3), 287-290. doi:10.1111/1467-9280.03437

Richeson, J. A., & Shelton, J. N. (2007). Negotiating interracial interactions: Costs, consequences, and possibilities. *Current Directions in Psychological Science, 16*, 316-320. doi: 10.1111/j.1467-8721.2007.00528.x

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J., & Müller, M. (2011). "pROC: an open-source package for R and S+ to analyze and compare ROC curves". *BMC Bioinformatics*, 12, p. 77. DOI: 10.1186/1471-2105-12-77

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*(2), 336-356. doi:10.1037/0033-295X.115.2.33

Seibt, B., & Forster, J. (2004). Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology, 87*(1), 38-56. doi: 10.1037/0022-3514.87.1.38

Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language, 72*, 116-141. doi: 10.1016/j.jml.2014.01.004

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. Behavior Research Methods, Instruments & Computers, 31(1), 137-149. doi:10.3758/BF03207704

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797-811. doi:10.1037/0022-3514.69.5.797

Tedeschini, J. (2012). Overcoming roadblocks to reform. *Journal of Applied Research in Memory and Cognition, 1*(2), 134–135. doi: 10.1016/j.jarmac.2012.02.004

Tierney, J. (2014, March 23). At airports, misplaced faith in body language. *The New York Times.* Retrieved from: http://www.nytimes.com/2014/03/25/science/in-airport-screening-body-language-is-faulted-as-behavior-sleuth.html?

Trawalter, S., Richeson, J. A., & Shelton, J. (2009). Predicting behavior during interracial interactions: A stress and coping approach. *Personality and Social Psychology Review*, *13*(4), 243-268. doi:10.1177/1088868309345850

Trawalter, S., & Shapiro, J. R. (2010). Racial bias and stereotyping: Interpersonal processes. In B. Gawronski, B. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 375-391). New York, NY: Guilford Press.

Tuk, M.A., Trampe, D., & Warlop, L. (2011). Inhibitory spillover: Increased urination urgency facilitates impulse control in unrelated domains. *Psychological Science, 22*(5), 627-633. doi: 10.1177/0956797611404901.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1), 104-132. doi: 10.1037/0033-295x.114.1.104

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498-505. doi:10.3758/BF03192720.

U.S. Census Bureau. (2014, March 27a) *State & county quick facts: Los Angeles County, C.A*. Retrieved April 9, 2014, from http://quickfacts.census.gov.

U.S. Census Bureau. (2014, March 27a) *State & county quick facts: Orange County, C.A*. Retrieved April 9, 2014, from http://quickfacts.census.gov.

U.S. Census Bureau. (2014, March 27b). *State & county quick facts: USA*. Retrieved April 9, 2014 from http://quickfacts.census.gov.

Van Bockstaele, B. B., Verschuere, B. B., Moens, T. T., Suchotzki, K., Debey, E., & Spruyt, A. (2012). Learning to lie: Effects of practice on the cognitive cost of lying. *Frontiers in Psychology*, *3*(526). doi:10.3389/fpsyg.2012.00526

Verschuere, B., Spruyt, A., Meijer, E. H., & Otgaar, H. (2011). The ease of lying. *Consciousness and Cognition: An International Journal*, *20*(3), 908-911. doi:10.1016/j.concog.2010.10.02

Visu-Petra, G., Varga, M., Miclea, M., & Visu-Petra, L. (2013). When interference helps: Increasing executive load to facilitate deception detection in the concealed information test. *Frontiers in Psychology, 4* :146. doi: 10.3389/fpsyg.2013.00146

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities (2nd ed.)*. New York, NY: John Wiley & Sons Ltd.

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

Vrij, A., & Granhag, P.A. (2012a). Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition, 1*(2), 110–117. doi: 10.1016/j.jarmac.2012.02.004

Vrij, A., & Granhag, P.A. (2012b). The sounds of critics: New tunes, old tunes, and resistance to play. *Journal of Applied Research in Memory and Cognition, 1*(2), 139–143. doi: 10.1016/j.jarmac.2012.05.001

Vrij, A., Granhag, P.A., Mann, S., & Leal, S. (2011) Outsmarting the liars: Toward a cognitive lie detection approach. *Current Directions in Psychological Science, 20*(1), 28-32. doi: 10.1177/0963721410391245

Vrij, A., Leal, S., Granhag, P.A., Mann, S., Fisher, R.P., Hillman, J., & Sperry, K. (2009). Outsmarting the liars: The benefit of asking unanticipated questions. *Law and Human Behavior, 33*(2), 159-166. doi: 10.1007/s10979-008-9143-y

Vrij, A., Mann, S., & Fisher, R. P. (2006). An empirical test of the behaviour analysis Interview. *Law and Human Behavior*, *30*(3), 329-345. doi:10.1007/s10979-006-9014-3

Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior*, *32*(3), 253-265. doi:10.1007/s10979-007-9103-y

Vrij, A., Mann, S., Leal, S., & Fisher, R. (2010). 'Look into my eyes': Can an instruction to maintain eye contact facilitate lie detection? *Psychology, Crime & Law*, *16*(4), 327-348. doi:10.1080/10683160902740633

Vrij, A., & Winkel, F. (1994). Perceptual distortions in cross-cultural interrogations: The impact of skin color, accent, speech style, and spoken fluency on impression formation. *Journal of Cross-Cultural Psychology*, *25*(2), 284-295. doi:10.1177/0022022194252008

Walczyk, J. J., Igou, F. P., Dixon, A. P., & Tcholakian, T. (2013). Advancing lie detection by inducing cognitive load on liars: A review of relevant theories and techniques guided by lessons from polygraph-based approaches. *Frontiers in Psychology, 4(14)*, 1664-1078 doi:10.3389/fpsyg.2013.00014.

Walczyk, J. J., Mahoney, K. T., Doverspike, D., & Griffith-Ross, D. A. (2009). Cognitive lie detection: Response time and consistency of answers as cues to deception. *Journal of Business and Psychology, 24*, 33-49. doi:10.1007/s10869-009-9090-8

Whelan, C. W., Wagstaff, G., & Wheatcroft, J. M. (2015). High stakes lies: police and non-police accuracy in detecting deception. *Psychology Crime & Law, 21*(2), 127-138. doi: 10.1080/1068316x.2014.935777

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika*, *70*(1), 203–212. doi:10.1007/s11336-003-1119-8

Appendices

Appendix 1:
Mock-Crime Phase Instructions and Description

During the ***crime phase*** interviewees will participate in a scenario that mimics a situation relevant to real-world contexts in which eliciting accurate information is as critical as detecting deception. The scenario involves entering a professor's office to retrieve a packet of journal articles to be delivered to another location. For analyses purposes we will divide the event into three intervals within the crime phase – each designed to ensure that liars and truth tellers spend equal amounts of time in the room. Interval Two is designed as the "critical" interval whereby liars and truth tellers' actions will differ. The latter ensures that the memories for liars and truth tellers will be unique in at least one, critical aspect (the lie).

"*Here is an office key to Professor Gitlin's office. Your mission is to work as a research assistant to retrieve journal articles from Professor Gitlin's office, to give to Professor Fenn. Please take this office key and backpack to Professor Gitlin's office in room 710j. When your reach the door, knock on her door. If Professor Gitlin answers, introduce yourself and let her know that you are there to pick up the three journal articles for Professor Fenn*." (note: no one will be in the office)

(Interval One)
"If she doesn't answer the door after a couple knocks, use the office key to let yourself in. ***Close the door****. Find the journal articles stored inside the desk drawer located under the printer. They are in a manila folder, open it to be sure that there are articles in there* (GRA shows participants sample article). *Take the journal articles and put them in your backpack.*

(Interval Two)
Innocent Suspects: *"Once the articles are in your backpack, wait for someone to come to the office to pick you up. The person will arrive in a few minutes and you can leave then. Do NOT leave before this person arrives. Instead you can take this time to look around. If you choose to move anything, make sure you put it back where it belongs. When the person arrives and knocks, tell her or him to wait a minute. At this time, please return to the professor's desk and write a note saying you took the articles. There are post-its and a pen on the desk. You may stick your note onto the professor's computer screen. After this, leave the room with the GRA. Here is a little cheat sheet overviewing your mission.* (GRA hands respective cheat sheet to participant)*"*

Guilty Suspects: *"Once the articles are in your backpack, wait for someone to come to the office to pick you up. The person will arrive in a few minutes and you can leave then. Do NOT leave before this person arrives. Instead of leaving, steal a folder with sensitive information. This folder is stored in the black filing cabinet. Above this cabinet, there will be a small, blue chest. The key to unlock the filing cabinet will be in one of the drawers in this chest. After unlocking the filing cabinet, look for a manila folder, titled,*

"Faculty Information." Inside this folder should be a list of faculty information such as office room numbers, extensions, as well as information about other sensitive subjects. You should put this folder in your backpack, in addition to the journal articles. This is CONFIDENTIAL. Do not disclose taking the folder or its contents to anyone. If you are asked, you should provide an alternative story for what you did in the office during this time. When the person arrives and knocks, tell her or him to wait a minute. At this time return to the professor's desk. There are post-its and a pen on the desk. Please write a note saying you took the articles. You may stick this note onto the professor's computer screen. After this, leave the room with the person. Here is a little cheat sheet overviewing your mission. (GRA hands respective cheat sheet to participant)*"*

(Interval Three)
This interval is the part where innocent and guilty suspects write a note that they had taken the journal articles. This occurs after the knock at the end of the "mission," while GRA is waiting outside.

Appendix 2:
Interview instructions: Cognitive Load Approach Interview

1. Initial greeting and questions
   a. You are later than we expected you to be. Go on in and get setup, and I will enter when you are ready.
   b. Some files with sensitive information were stolen from the professor's office, and we need to gather as much information as we can to figure out who stole the information. We are asking you about this information because we know you were in the building and you went to the professor's office. Please speak loudly throughout the interview.
   c. Secondary Task Instructions
      a. While you are being interviewed, you will also be asked to complete a secondary task using a computer screen. There will be a keypad on the table that you are to use to respond to high or low tones that you hear. You should press the button when you hear the high/low tone (participants will hear a high/low tone at this point) and avoid pressing the button when you hear the low/high tone (participants will hear a low/high tone at this point).
   d. Demographics

   (These questions are asked to allow the participant time to become acquainted with the physiological measures and secondary task while concurrently creating awareness of social identity (question c)
      a. What is your ethnicity
      b. Tell me where you work
      c. What is your ethnicity

2. General Prompt: Please describe everything you did from the time you entered the professor's office until now. Please be as detailed as possible and include any details no matter how trivial they may seem.
3. Unanticipated Questions (Sensory/Spatial)
   a. If I was next to you and I could see but not hear, tell me what I would see around the office and describe the layout of the office.
   b. If I was next to you and I could hear but not see, what would I hear?
   c. How were your feelings when you were in the office? What emotions, if any, were you experiencing?

4. Unanticipated Questions (Temporal)
   a. Please tell me (more) about the first couple of things that you did during your time in the office.
   b. Please tell me (more) about the next couple of things that you did during your time in the office.

c. Please tell me (more) about the last couple of things that you did during your time in the office.

5. Repeated Question: I know this seems tedious, but I want to make sure I have your story straight. It is very important that we know everything that happened. Please tell me again about the events that occurred from the time you entered the professor's office until now.
6. Reverse Order Question: Now I want you to tell me what happened in reverse chronological order. Start by telling me what happened last and move toward the beginning from there.
7. Closing Question: Is there anything that you've left out or would like to correct about what you have reported during this interview?

Appendix 3:
Murphy 15 item memory scale:

This 15-item scale was developed based on the methods of Murphy, Steele, and Gross (2007), pg. 881.

Memory for Interview Scale

Thank you for participating in our interview! Please answer the following 15 questions regarding your experience and memory for what happened to you during the interview. If you were asked to lie during the interview, please tell the truth when answering the following questions. It is very important that you answer these questions as honestly as possible.

1. Where did your interview take place?
    a. 6th
    b. 8th
    c. 5th
    d. 7th

2. Why were you interviewed? *Please write your response in the space below.*

Please circle your response to the next questions.

3. Were you being videotaped during your interview?
    a. Yes
    b. No

4. If yes, where was the video camera located in the room?
    a. In front of me, close to my left
    b. In front of me, close to my right
    c. Far in the back of the room, centered in front of me
    d. Far in the back of the room, on the ceiling in front of me

Please write your response to the next question in the blank space provided

5. What did the interviewer say when you arrived to the interview?

Please indicate your agreement with the following statement by circling your response on a scale of 1 (strongly disagree) to 7(strongly agree).

6. The interviewer was pleased when I arrived to the interview

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

7. The interviewer asked questions about my emotions

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

Please answer by circling whether the following statements are "true" (a) or "false" (b).

8. The interviewer asked me about my ethnicity.
    a. True
    b. False

9. The interviewer repeated some questions during the interview.
    a. True
    b. False

10. I was asked to describe the layout of the office.
    a. True
    b. False

Please circle, or write in, your response to the following questions regarding your interviewer.

11. What was the race of the interviewer?
    a. White
    b. Hispanic
    c. Black
    d. Asian
    e. Other : _____ (please write your best guess)

12. What color were the interviewer's eyes?
    a. Brown
    b. Blue
    c. Green/hazel

13. What color was the interviewer's hair?
    a. Brown
    b. Blonde
    c. Red

Please write your response to the following questions in the blank space provided.

14. What was the interviewer wearing?

15. What was the final question that you were required to answer?

Appendix 4:
Self-report measure of stereotype threat--based on pg. 806 of Steele and Aronson (1995), and based on Najdowski (2013).

Please indicate how much you agree or disagree with the following statements by circling your response. A response of 1 indicates that you strongly disagree and a response of 7 indicates that you strongly agree with the statement. Thank you for your responses!

1. ***Some people perceive me as guilty because of my race***

   1 2 3 4 5 6 7
   Strongly Disagree Strongly Agree

2. ***Appearing innocent during an interview may be easier for people of my race***

   1 2 3 4 5 6 7
   Strongly Disagree Strongly Agree

3. ***The interviewer expected me to be guilty because of my race***

   1 2 3 4 5 6 7
   Strongly Disagree Strongly Agree

4. ***In interviews, people of my race often face biased evaluations***

   1 2 3 4 5 6 7
   Strongly Disagree Strongly Agree

5. ***My race does not affect people's perception of my guilty or innocence***

   1 2 3 4 5 6 7
   Strongly Disagree Strongly Agree

For the next items, please think about the interview that you just participated in. Please indicate how strongly you disagree (1) or agree (7) with the following statements regarding your experience with the interviewer.

1. ***I was worried that the interviewer might stereotype me as a criminal because of my race***.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

2. ***I was worried that something I did might be misinterpreted as suspicious by the interviewer because of my race***.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

3. ***I was worried that the interviewer's perceptions of me might be affected by my race.***

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

4. ***I was worried that the interviewer would suspect me of having committed a crime just because of my race.***

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

5. ***I was worried that my anxiety about confirming the stereotype about my race and crime would negatively influence my interaction with the interviewer.***

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

Appendix 5: Experiment 1 Supplementary Tables

Table A. Experiment 1a Cognitive Vigilance Scores.

	Truth (n = 7)		Lie (n = 8)	
	M	***SD***	***M***	***SD***
Hispanic	18.16	4.11	20.50	2.68
White	23.72	2.85	21.41	3.04

Table B. Experiment 1a Participant-Interviewee Average RT and Error on the Secondary Task.

	Hispanic				White			
Error	**Truth (n = 8)**		**Lie (n = 8)**		**Truth (n = 8)**		**Lie (n =7)**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Error Baseline	.06	.08	.08	.14	.06	.11	.05	.10
Error Interview	.11	.06	.13	.14	.12	.10	.25	.13
RT (ms.)								
RT Baseline	433.97	222.96	494.33	190.24	234.94	106.77	272.38	130.84
RT Interview	503.61	186.98	436.40	239.45	290.38	79.56	291.20	39.95

Table C. Experiment 1a Participant-Interviewee Average Self-Report of Cognitive Load

	Hispanic				White			
Behavioral Cues	**Truth**		**Lie**		**Truth**		**Lie**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
How difficult was the auditory task of pressing a button when hearing a high or low tone?	3.63	2.39	3.00	2.00	2.63	1.77	4.13	2.03
How difficult was it to answer the interview questions?	4.00	1.31	3.00	1.51	3.25	1.39	2.5	1.20
How difficult was it to perform the tone task and answer the interview questions simultaneously?	5.19	2.00	5.00	1.31	4.63	2.00	5.13	.99
Overall, how difficult was the interview experience?	4.19	1.60	3.88	.83	4.00	1.69	3.25	1.39

Note. All ratings were on a 1(not at all) – 7 (very) scale.

Table D. Experiment 1a Participant-Interviewee Average Pulse.

Pulse	Hispanic				White			
	Truth (n = 7)		Lie (n = 8)		Truth (n = 8)		Lie (n =7)	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Pulse Pre-Interview	79.36	13.28	79.94	18.01	68.63	8.31	79.71	11.70
Pulse Post-Interview	76.29	12.58	16.82	12.51	69.19	9.40	78.79	11.68

Table E. Experiment 1a Participant-Interviewee State-Trait-Anxiety-Index (STAI) Ratings.

Time	Hispanic				White			
	Truth (n = 8)		Lie (n = 8)		Truth (n = 8)		Lie (n = 8)	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Pre - Mock Crime	35.35	9.63	35.38	6.44	33.25	4.83	35.25	5.26
Pre - Interview	38.38	14.66	38.88	9.01	36.88	10.64	42.06	13.66
Post - Interview	34.25	10.72	29.88	6.22	27.75	6.23	30.50	7.48

Appendix 6: Experiment 2 Supplementary Tables

Table A. Experiment 2a Cognitive Vigilance Scores.

	Truth (n = 7)		Lie (n = 8)	
	M	***SD***	***M***	***SD***
Hispanic	23.09	3.63	23.34	2.31
White	20.13	3.44	19.81	3.81

Table B. Experiment 2a Participant-Interviewee Average Error and RT on the Secondary Task.

	Hispanic				White			
Error	**Truth (n = 8)**		**Lie (n = 8)**		**Truth (n = 8)**		**Lie (n =7)**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Error Baseline	.05	.08	.02	.03	.07	.06	.02	.02
Error Interview	.24	.15	.17	.10	.18	.09	.29	.15
RT (ms.)								
RT Baseline	347.19	198.33	368.78	251.86	323.57	202.03	479.34	344.24
RT Interview	400.91	149.01	322.62	164.66	366.57	101.98	370.32	96.24

Table C. Experiment 2a Average Self-Report of Cognitive Load

	Hispanic				White			
Behavioral Cues	**Truth**		**Lie**		**Truth**		**Lie**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
How difficult was the auditory task of pressing a button when hearing a high or low tone?	2.89	1.96	2.25	1.49	3.63	1.30	2.38	1.51
How difficult was it to answer the interview questions?	3.75	2.05	2.89	1.55	3.38	1.30	4.00	2.00
How difficult was it to perform the tone task and answer the interview questions simultaneously?	4.50	1.07	4.38	1.92	4.5	1.31	5.00	1.60
Overall, how difficult was the interview experience?	3.75	2.12	4.00	1.31	3.75	1.49	4.88	1.46

Note. All ratings were on a 1(not at all) – 7 (very) scale.

Table D. Experiment 2a Hispanic Interviewer Average Mean Arterial Blood Pressure (MABP) and Pulse.

	Hispanic				White			
MABP	**Truth (n = 8)**		**Lie (n = 8)**		**Truth (n = 8)**		**Lie (n =8)**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
MABP Pre-Interview	92.54	12.25	89.79	12.03	94.71	8.81	97.15	10.90
MABP Post-Interview	87.73	10.30	90.54	14.50	91.83	9.58	94.10	10.74
Pulse								
Pulse Pre-Interview	87.38	15.56	88.31	14.98	80.31	14.02	82.56	14.73
Pulse Post-Interview	83.38	11.53	87.88	13.61	80.31	10.63	81.25	10.80

Table E. Experiment 2a Participant-Interviewee State-Trait-Anxiety-Index (STAI) Ratings.

Time	Hispanic				White			
	Truth (n = 8)		Lie (n = 8)		Truth (n = 8)		Lie (n = 8)	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Pre - Mock Crime	40.25	5.12	36.63	5.4	37.38	9.47	39	4.24
Pre - Interview	47.13	5.94	36.38	10.33	38.13	13.91	43.5	6.26
Post - Interview	31.38	7.41	29.63	4.17	32.13	12.92	35	6.05

Table F. Experiment 2c Observer (N = 134) Veracity Ratings

Interviewee Ethnicity	Veracity			
	Truth		Lie	
	M	***SD***	***M***	***SD***
Hispanic	2.91	.82	3.16	.86
White	2.97	.83	3.25	.83

Note: Observer ratings on 1(definitely truth) to 6 (definitely lying) scale.