

The Effect of Cognitive Load on Gender Discrimination in Job Interviews

By

Sulan Kith

**A Thesis
Presented to
The University of Guelph**

**In partial fulfillment of requirements
For the degree of
Master of Arts
In
Psychology**

Guelph, Ontario, Canada

© Sulan Kith, January, 2018

ABSTRACT

THE EFFECT OF COGNITIVE LOAD ON GENDER DISCRIMINATION IN JOB INTERVIEWS

Sulan Kith
University of Guelph, 2018

Advisor:
Dr. Deborah Powell

Interviews are cognitively-demanding tasks; and the load they put on interviewers may be exacerbated by a number of extraneous situational factors. High cognitive load may compromise the resistance to discrimination that otherwise distinguishes structured interviews from other interview methods, leading to heuristically-driven or discriminatory hiring decisions. The present study examined whether cognitive load influences interviewers' likelihood to gender discriminate during job interviews, and whether the effect differs based on the measure of discrimination. In the study, participants completed an online interview simulation administered through a crowdsourcing platform. During the simulation, participants were randomly assigned to either a high cognitive load or low cognitive load condition. All participants were asked to evaluate the interview responses of two equally-qualified, but differently-gendered candidates who were applying for a stereotypically-masculine or stereotypically-feminine job. Participants provided ratings of each candidates' suitability, as well as a final hiring decision. The cognitive load manipulation was not found to affect the numeric ratings that participants gave the candidates; however, participants in the high cognitive load condition chose to hire the stereotypically gender-congruent candidate more frequently than participants in the low cognitive load condition. These results support the need for incorporating multi-step selection strategies that go beyond interviews, in order to help to prevent discrimination from influencing hiring decisions.

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Deborah Powell, for taking the time to read through countless drafts of this document, and for providing me with valuable feedback throughout our time working together. I would also like to thank Dr. Peter Hausdorf and Dr. Harjinder Gill for serving on my advisory committee, and for taking time out of their busy schedules to meet with me to help me improve my work. As well, I would like to thank my wonderful unofficial committee member, Scott Cassidy, for his gentle patience and incredible insight, which he never failed to offer up during our many discussions together.

I would also like to thank my amazing friends Jessica Christie, Maria Ning, and Ian Christie for their endless support, and for reminding me that I'm never alone. Furthermore, I would like to thank my family for their support (which they offered unconditionally, despite never fully understanding my program of study), and for ensuring that I was remained well-fed throughout my Master's program. Finally, I would like to thank all the wonderful people in the Guelph's Psychology program for making the many long nights I spent in Blackwood Hall that much brighter.

TABLE OF CONTENTS

| | |
|--|-----|
| ABSTRACT..... | ii |
| ACKNOWLEDGEMENTS..... | iii |
| TABLE OF CONTENTS..... | iv |
| Stereotypes and Discrimination..... | 1 |
| Evidence of Stereotyping and Discrimination in Interviews..... | 4 |
| Employment Interviews..... | 6 |
| Factors That Encourages Stereotyping..... | 10 |
| Stereotyping and Cognitive Load..... | 13 |
| Manipulating Cognitive Load..... | 14 |
| Measuring Cognitive Load..... | 15 |
| The Present Study..... | 16 |
| Method..... | 18 |
| Participants..... | 18 |
| Materials..... | 19 |
| Procedure..... | 22 |
| Measure..... | 25 |
| Results..... | 26 |
| Exclusion Criteria..... | 26 |
| Manipulation Check: Cognitive Load..... | 28 |
| Exploratory Outcome: Stereotype Activation..... | 29 |
| Data Aggregation..... | 30 |
| Hypothesis 1..... | 33 |
| Hypothesis 2..... | 36 |
| Discussion..... | 36 |
| The Effect of Cognitive Load on Gender Discrimination..... | 36 |
| Manipulation check: Cognitive Load..... | 39 |
| Implications for Research..... | 39 |
| Implications for Management..... | 40 |
| Limitations and Future Directions..... | 41 |
| Conclusion..... | 42 |
| References..... | 43 |

| | |
|---|----|
| Table 4 | 50 |
| Table 5 | 51 |
| Appendix A: Rating Scales, Questions and Competency Description | 52 |
| Critical Thinking | 52 |
| Monitoring | 53 |
| Final hiring decision | 53 |
| Appendix B: Introduction to Job Position | 54 |
| Appendix C: Study Overview | 56 |
| Appendix D: Interview Scripts | 57 |
| Critical Thinking Script A | 57 |
| Critical Thinking Script B | 58 |
| Monitoring Script A | 59 |
| Monitoring Script B | 60 |
| Appendix E: Hypothesis Results | 62 |
| Appendix F: Cognitive Load Manipulation Instructions | 63 |
| Appendix G: Cognitive Load Manipulation Secondary Task | 64 |
| Appendix H: Experimental checks | 66 |
| Hiring Experience | 66 |
| Job Position | 66 |
| Job Position Gender Stereotype | 66 |
| Candidate's Gender | 66 |
| Appendix I: Demographic Questions | 67 |
| Appendix J: Attention Check Questions | 68 |
| Appendix K: Cognitive Load Self-Rating Scale | 69 |
| Appendix L: Stereotype Activation | 69 |
| Appendix M: Recruitment Ad | 70 |
| Appendix N: Information Letter and Consent Form | 71 |
| Appendix O: Debriefing Form | 73 |

The Effect of Cognitive Load on Gender Discrimination in Job Interviews

Employment interviews are difficult, not only for interviewees, but interviewers as well. Interviews often require interviewers to actively focus on a number of tasks (e.g., taking notes, asking questions and listening actively) to accurately evaluate candidates. Insufficient preparation, or external distractions can further increase the mental strain on interviewers and, in turn, potentially impact the accuracy of their hiring decisions. The term “cognitive load” is used to describe this strain on the cognitive system produced from performing a specific task (Paas, Van Merriënboer, & Adam, 1994). Little research has investigated the effect of the interviewers’ cognitive load on their decision making. When faced with challenging hiring decisions, interviewers may attempt to save mental effort by unconsciously relying on a technique that eases the decision-making process: stereotyping. By definition, stereotyping is the automatic tendency to categorize people based on easily observable traits, such as gender, race, age or marital status (Tajfel, 1969). However, stereotyping can promote inaccurate perceptions of others. When stereotypes lead to discrimination, they can also block minority groups attempting to enter occupational fields (Cohen & Bunker, 1975; Frauendorfer & Schmid Mast, 2013). This marginalization of minority groups violates the Canadian Human Rights Act (*Canadian Human Rights Act*, 1977). To contribute to the accuracy and fairness of the selection process, my research investigates the impact of high (versus low) cognitive load on interviewers’ tendency to discriminate in job interviews.

Stereotypes and Discrimination

Stereotyping stems from the natural tendency to sort everyone, including ourselves, into either in-groups or out-groups (Fiske & Lee, 2008; He & Brown, 2013; Hornsey, 2008; Jetten, Spears, & Postmes, 2004). It is a process that is beneficial in saving mental effort. For example,

upon entering a store, an individual may categorize people as either customers or employees, thus allowing them to better direct their search for assistance. Under high cognitive load, stereotypes allow for quick evaluations without requiring a deep understanding of the individual being evaluated (Fiske & Lee, 2008; Frauendorfer & Schmid Mast, 2013). Stereotyping eases information processing by relying on already stored information instead of new incoming information (Hilton & von Hippel, 1996). An example of a stereotype is the conception that women are caring, warm and family-oriented. Stereotypes alone are not necessarily harmful; they hold a function in assisting in quick decision making. Stereotypes also act to provide meaning and understanding about oneself and one's experiences using one's group membership (Tajfel, 1981). For example, a kind and caring woman may use her gender as an explanation for her gentle manner. However, problems arise due to stereotypes' unavoidable relationships with discrimination.

Unlike stereotyping, discrimination tends to lead to harmful consequences.

Discrimination takes place when an individual's behaviour advantageously affects one group over another (Fiske & Lee, 2008). Discrimination can be conceptualized through the social identity approach. The social identity approach is based on a combination of three theories: a) social identification, b) self-categorization, and c) social comparison. Social identification suggests that an individual's self-image is based on their membership within different social groups (Hornsey, 2008; Tajfel, 1982). The self-categorization theory describes the tendency to sort everyone, including ourselves, into either in-groups or out-groups (Hornsey, 2008; Jetten et al., 2004). Self-categorization theory stems from the desire to view one's in-group favourably and to protect one's self-image and self-esteem, as social identity is linked with group membership (Hornsey, 2008; Tajfel, 1982). Given that these mentally created groups only

psychologically exist when defined in comparison to other groups, people will think and act in ways to achieve or maintain a positive image of their in-group, especially in comparison to other out-groups. This comparison between groups is known as social comparison. Behaviours that support or favour an individual's in-group will also in turn disadvantage people outside their group. Discrimination is the product of this disadvantage. An example of discrimination could include a male CEO not selecting a board member candidate due to her gender. In addition to discrimination between groups, discrimination within groups can occur. Intragroup members who deviate from perceived in-group similarities may threaten group cohesion and self-meaning. Intragroup discrimination (i.e., within-group discrimination) can occur due to in-group members' desire to uphold cohesion and meaning (Abrams & Hogg, 1988). An example of intragroup discrimination is a female manager not promoting a female subordinate to a position associated with more responsibility and greater time commitment due to the subordinate's gender. Discrimination can hinder both out-group members and those within who contradict their group norms.

Discrimination and stereotypes are unfortunately connected. Devine's (1989) dissociation model helps explain the connection between discrimination and stereotypes, as well as the different levels of discrimination enacted across individuals. According to Devine's (1989) dissociation model, there exists a clear division between knowledge and personal beliefs about a stereotype group (Devine, 1989; Gawronski & Creighton, 2013; Lepore & Brown, 1997). The dissociation model suggests that all people, regardless of their attitudes, are generally aware and knowledgeable of cultural stereotypes within their culture. However, a person's knowledge of stereotypes may not necessarily reflect their own personal beliefs. For example, an activist for women's rights would be aware of stereotypes associated with women, but believe these

stereotypes to be untrue. One key study that supports the dissociation model is Devine's (1989) study, which asked participants to list stereotypes associated with African Americans. Afterwards, participants completed the seven-item Modern Racism Scale to assess participant's racism. Participants listed similar stereotypes (i.e., their knowledge) regardless of their racism score (i.e., their beliefs). The Dissociation model also suggest that stereotypes are unavoidable and automatically activated in equal strength in all people, despite their personal beliefs and attitudes. Uninhibited, automatically activated stereotypes will lead both high and low discriminatory people to form stereotypical congruent responses. However, when automatic processes conflict with personal beliefs, controlled processes can hinder typically automatic processing responses. These controlled, non-discriminatory responses require attention, time and effort (Devine, 1989). Using the previous example about women, Devine's dissociation model would assume that all people have a general awareness that women are associated with traits such as caring, nurturing and strong family values. These stereotypes are automatically activated in all people. However, people who discriminate exhibit less suppression of the automatic association between women and female-related stereotypes (i.e., women are caring, nurturing and family oriented). Devine's dissociation model provides insight on the importance of suppressing stereotypes, as acts of discrimination can form from unsuppressed stereotypes. Preventing discrimination becomes especially critical during situations such as employment interviews, because hiring decisions have large implications for both the interviewee and the hiring organization.

Evidence of Stereotyping and Discrimination in Interviews

Because interviews have important consequences for both parties, problematic issues can arise when stereotyping leads to discrimination in employment interviews. Stereotyping allows

for interviewer's biases towards non-performance related factors (e.g. gender, marital status, sexual orientation, racial background and age) to impact hiring decisions (Ahmed, Andersson, & Hammarstedt, 2012; Cohen & Bunker, 1975; Huffcutt & Roth, 1998; Nadler & Kufahl, 2014; Przygodzki-lionet, Olivier, & Desrumaux, 2010). For example, racial and age discrimination have been shown to be factors that impact hiring decisions. In regard to racial discrimination, a meta-analysis of 31 studies found that Caucasian applicants received approximately a quarter standard deviation higher in interview ratings than African-American and Hispanic applicants, with a greater effect in low structured interviews ($d = .32-.71$) than high structured interviews ($d = .17-.23$) (Huffcutt & Roth, 1998). With the aging population, ageism in the workforce has also become a growing concern. In Ahmed et al.'s study (2012), employers were approximately three times more likely to offer an interview or job to a fictional 31 year old male applicant than to a 46 year old male applicant for a restaurant worker position, and four times more for a sales assistant position. In another study, Nadler & Kufahl (2014) revealed that interactions between sexual orientation and marital status in women may impact hiring decisions. They manipulated the marital status, gender and sexual orientation of interviewees in an interview simulation using Mechanical Turk. The study's results showed that ratings were influenced by the interaction between marital status and sexual orientation among women, such that single homosexual women obtained higher ratings than did married homosexual women. However, married heterosexual women obtained higher ratings than did single heterosexual women. These studies provide evidence of discrimination within the interview process and its potential effects.

The fit between candidates' gender and stereotypical norms surrounding the job position can also influence job recruiters' hiring decisions, as shown in Cohen & Bunker's (1975) study. Cohen and Bunker had job recruiters evaluate hiring materials (i.e., an interview transcription) of

a male or female job applicant for either a stereotypically male (personnel technician) or stereotypically female (editorial assistant) job position. Their results showed that the interaction between the applicant's gender and stereotypical norms surrounding the job position influenced the hiring decisions of the job recruiters, such that job recruiters recommended more male applicants for the personnel technician position, and more females for the editorial assistant position. Given the interaction between candidate's gender and job position stereotypes, greater discrimination may be observed in hiring decisions for job positions closely associated with gender stereotypes. Discrimination within the selection process results in the marginalization of minority groups. Given the illegal and unjust implications associated with stereotyping, it is necessary for organizations to take deliberate efforts to minimize stereotyping within hiring practices, especially in interviews.

Employment Interviews

Interviews are often an important component of the hiring process. Employment interviews are defined "as a personally interactive process of one or more people asking questions orally to another person and evaluating the answers for the purpose of determining the qualifications of that person in order to make employment decisions" (Levashina, Hartwell, Morgeson, & Campion, 2014). There are various types of interviews, which can vary based on communication medium, structure and type of questions. In addition to the traditional in-person interviews, modern technology has allowed for interviews to take place through both immediate forms of communication (e.g. phone and video chat) as well as delayed interactions; for example, when interviews are recorded and the interviewer evaluates the interviewee at a later point in time (Levashina et al., 2014).

Interviews can also vary based on structure. Campion, Palmer, and Campion (1997) broadly define structure in interviews as any improvement to the interview process that increases psychometric properties, such as standardization, evaluations methods or better questions. Interviews with more structure tend to show greater reliability (Huffcutt, Culbertson, & Weyhrauch, 2013) and validity (McDaniel, Whetzel, Schmidt, & Maurer, 1994). For example, a meta-analysis using an employment interview sample revealed higher mean inter-rater reliabilities for structured interviews conducted separately ($M = .61$, $n = 95$, $SD = .11$, $CI_{80\%} = .53 - .69$) or in panels ($M = .78$, $n = 26\ 202$, $SD = .12$, $CI_{80\%} = .76 - .80$) compared to unstructured interviews conducted separately ($M = .40$, $n = 1\ 260$, $SD = .22$, $CI_{80\%} = .34 - .46$) or in panels ($M = .55$, $n = 1\ 506$, $SD = .13$, $CI_{80\%} = .50 - .60$) (Huffcutt et al., 2013). In terms of validity, McDaniel et al. (1994) showed higher validity for structured interviews ($\rho = .44$, $n = 12\ 847$, $\sigma_p = .27$) than unstructured interviews ($\rho = .33$, $n = 9\ 330$, $\sigma_p = .12$) in regards to predicting job performance. Adding structure tends to enhance interviews as a selection assessment tool.

Structured interviews tend to be better predictors of job performance ($r = .51$) than other selection tools, such as unstructured interviews ($r = .38$) (Schmidt & Hunter, 1998), because well-developed structured interviews assess candidates based on core competencies required for the job position. These core competencies are valid predictors of job performance, as they are determined through a job analysis. In order to evaluate candidates on these core competencies, most structured interviews employ either situational or behavioural questions (Taylor & Small, 2002). Situational questions involve asking the interviewee to speculate how they would react in imaginary situations that present a dilemma (Latham, Saari, Pursell, & Campion, 1980; Pulakos & Schmitt, 1995). These questions tend to be future oriented. Behavioural or experience-based interview questions probe into the interviewee's past experiences that relate to knowledge, skills

and ability relevant to the specific interviewing position (Janz, 1982; Motowidlo et al., 1992; Pulakos & Schmitt, 1995). For this study, I used a structured interview with delayed interactions and the following elements of structure: interview questions based on a job analysis, individually rated answers with rating scales, consistent interview questions for all candidates, past-behavioural questions and no discussion between interviews about the candidates or answers.

Gender Discrimination in Interviews. Previous research on the structured interview suggests that even structured interviews are vulnerable to stereotypes, although the effect is small (Levashina et al., 2014; McCarthy, Van Iddekinge, & Campion, Michael, 2010). Levashina et al. specifically examined the effects of gender and race on structured interview ratings. Their results showed both a small effect of gender ($k = 13, n = 125,720, d = .09, SD = .30, CI = .05, .13$) and race ($k = 16, n = 126,386, d = .13, SD = .21, CI = .09, .17$) in structured interviews, suggesting that demographics, such as race and gender, may slightly influence hiring decisions in structured interviews. McCarthy et al. also investigated the effect of gender and race, but McCarthy et al. distinguished between different types of structured interviews (situational and past-behavioural) using an applicant sample applying for a managerial track position within a U. S. government organization. McCarthy et al. analyzed the effect of race between White-Black, White-Asian, and White-Hispanic for both situational and behavioural interviews. Their results showed non-significant main effects for applicants' race, interviewer's race and interactions between applicant and interviewer's race, with effect sizes ranging from .00 to .02. Using a sample of 36,572, McCarthy et al. also revealed no main effect of interviewer's gender for both situational ($d = .00$) and behavioural interviews ($d = .01$), but a small significant main effect of applicant gender for both situational ($d = .03$) and behavioural ($d = .04$) interviews. However, neither Levashina & Campion nor McCarthy et al. accounted for potential gender stereotype norms

associated with the specific job role as part of their studies. Both studies assumed that females would experience more discrimination compared to men, without looking at the fit between applicants' gender and job gender stereotypes. Women applying for positions in the human resources and office administrations departments may not have experienced discrimination, as their gender would fit with the cultural stereotype of women being "supportive". By not accounting for the fit between the candidate's gender and job stereotypes, these studies may have diluted the effect of gender.

Previous research suggests that gender stereotypes associated with the job positions can impact ratings (Davison & Burke, 2000). Davison and Burke explored the effect of gender stereotypes associated with jobs (gender-type) on applicant ratings within simulated employment context as part of their meta-analysis. Davison and Burke classified jobs as either male gender-type, female gender-type or neutral gender-type and examined the difference between relative mean effect sizes. A positive correlation signified that participants rated male applicants higher, and a negative correlation signified that participants rated female applicants higher. Their results revealed that for female gender-type jobs (e.g. secretary, home economics teacher, day care center director) females received higher selection ratings than males ($M = -.13$, $N = 553$, $SD = .23$, $CI = -.21, -.05$). On the other hand, males received higher selection ratings for male gender-type jobs (e.g. heavy machinery sales manager and auto salesperson) ($M = .17$, $N = 1061$, $SD = .15$, $CI = .11, .23$). This effect of gender-type decreased when more job-relevant information was provided. These results support the idea that raters will rate applicants who are congruent to the job gender-type higher than applicants who are incongruent to the job gender-type (Davison & Burke, 2000). Thus, the effects of gender discrimination may be most prominent in positions tied to gender norms. Although structured interviews can reduce the effects of discrimination, they do

not eliminate discrimination completely (Levashina et al., 2014; McCarthy et al., 2010). Small effects are still important given the impact hiring decision can have on a person's life. My study examines gender discrimination within the context of a structured interview for a gender-typed job.

Factors That Encourages Stereotyping

In order to further reduce discrimination in interviews, better understanding of the factors that encourage stereotyping is necessary. Various factors, such as ambiguity (i.e., lack of relevant information), can compromise interviewers' into stereotyping (Hilton & von Hippel, 1996; McDermott, 1997). An example in which ambiguity, or lack of available relevant information, can promote stereotypes are elections. An American National Election Survey asking respondents to rate candidates' perceived ideology, ranging from very liberal to very conservative, found that respondents rated female candidates as more liberal compared to their male counterparts, regardless of their political party (Democratic or Republican) (McDermott, 1997). Data from the same survey from 1986 through 1994 revealed that among voters with no contact with the candidates (low information condition), the female Democratic candidate was favoured over their male counterparts. This effect diminished when voters had contact with the female candidate (high information condition). These results indicate that voters with less information favoured the female Democratic candidate based on gender. In order to prompt stereotyping in the interviewers, we added elements of ambiguity within this study.

Another study that exemplifies the effects of ambiguity is Davison and Burke's (2000) meta-analysis. Davison and Burke also examined the effects of ambiguity, specifically, how the availability of job-relevant information influenced ratings of different gender applicants. For their study, they evaluated and scored the amount of job-relevant and non-relevant information

present within each gender discrimination study. After, Davison and Burke correlated the job-relevant information score to each study's effect size. They also created a ratio between job-relevant and non-relevant information and correlated this ratio to each study's gender discrimination effect size. Negative correlations indicated that as less job relevant information were provided to raters, female applicants experienced greater discrimination. The results from both the job-relevant information score ($r = -.16, k = 38$) and job-relevant and non-relevant information ratio ($r = -.15, k = 35$) provided support that people tend to rely more on stereotypes in conditions with limited information or more ambiguity. Although the interview is supposed to provide an opportunity for interviewers to meet interviewees, the interview's short, and often stressful, nature may limit interviewers' ability to gather quality information about the interviewee.

Ambiguity's effects on discrimination may vary depending on the level of ambiguity. Dovidio and Gaertner (2000) investigated the effects of ambiguity on the hiring ratings and decisions for black and white candidates. In their study, participants read interview excerpts from either white or black candidate for a peer counseling program. Dovidio and Gaertner manipulated the strength of the candidate's qualification to create three qualification levels: clearly strong, ambiguous, and clearly weak. Participants answered if they would hire the candidate (yes or no) and rated the strength of their recommendation on a 10-point scale, with 1 (not at all) to 10 (very strongly). Results from the dichotomous hiring decision revealed an interaction between race and qualification of the candidate ($\chi^2 = 6.75, df = 2, N = 194, p < 0.035$). Participants hired the black candidate at a similar frequency compared to the white candidate for both the clearly strong (91% vs. 85%) and clearly weak (13% vs. 6%) conditions. However, the participants favoured the white candidate when candidate's qualifications were ambiguous (45%

vs 76%) ($\chi^2 = 6.35$, $df = 1$, $N=66$, $p < 0.012$). Participant's strength of recommendation also followed a similar pattern, with an interaction between qualifications and race of candidate, $F(2,182) = 6.08$, $p < .003$. Participants recommended black and white candidates at comparable strengths for both the clearly strong ($M_s = 7.18$ vs 6.52) and clearly weak ($M_s = 3.50$ vs 2.81) conditions. However, the white candidate was favoured over the black candidates in the ambiguous condition, $t(64) = 2.79$, $p < .001$. Dovidio and Gaertner's (2000) study suggest that discrimination occurs most strongly in ambiguous situations. For this reason, I incorporated ambiguity within this study in order to mimic a realistic interviewing environment where interviewers are vulnerable to discrimination, but keep the level of ambiguity in both cognitive load conditions consistent.

A second factor that can also promote the use of stereotypes in an interview is priming. Priming is the influence of prior experiences on current perspective and cognition. For example, men tend to engage in more sexual harassment behaviours after watching non-violent pornographic videos (McKenzie-Mohr & Zanna, 1990). Priming has also been shown to impact behaviour even a week after exposure (Smith, Stewart, & Buttram, 1992). For example, Smith et al. demonstrated that participants judged primed behaviours more quickly even if participants did not recognize that they had seen the behaviour seven days prior. A meta-analysis that examined the effects of racial and gender stereotypes on test performance suggests that priming is effective in stimulating stereotype threat (situations where social groups feel at risk of conforming to their associated stereotypes) (Nguyen & Ryan, 2008). It is difficult to avoid priming in interviews. For example, knowing the candidate's name and inferring gender before an interview could potentially prime an interviewer to use gender stereotypes.

Stereotyping and Cognitive Load

Cognitive load is an additional factor that can encourage stereotyping, and is one of the primary areas within this current study. Predictors of cognitive load includes task characteristics (e.g., structure, reward system, novelty and time pressure), environmental conditions of the task (e.g., temperature and noise), subject's characteristics (e.g., cognitive abilities, prior knowledge) and subject-task interactions (e.g., arousal and motivation). Cognitive load theory (CLT) (Chandler & Sweller, 1991; Paas, Tuovinen, Tabbers, & Van Gerven, 2003; Paas et al., 1994) suggests that every individual has a finite amount of cognitive processing capabilities. Based on the CLT, there exist three types of cognitive load: intrinsic, extraneous and germane (Paas et al., 2003). The two most relevant for the current study are intrinsic and extraneous cognitive load. Intrinsic cognitive load refers to the inherent nature of the task and the individual's expertise. For example, the type of interview (structured or unstructured), and the interviewer's experience interviewing will contribute to the amount of intrinsic load. Extraneous load is the additional load beyond intrinsic cognitive load that can be caused by poor execution or design. For example, confusing rating materials, unclear procedures or a noisy environment can increase extraneous load (Debie & van de Leemput, 2014; Paas et al., 2003). The effects of intrinsic and extraneous load are additive. In order to reduce total cognitive load, Paas et al. suggested focusing on the minimization of extraneous load. In order to investigate cognitive load in this study, I compared interviewers with increased extraneous load to interviewers under normal conditions.

The amount of cognitive load on interviewers is an important consideration when designing interviews, as it can influence the recall of information. In cognitively demanding situations, people tend to better recall information congruent with established expectations

(Bodenhausen & Lichtenstein, 1987; Macrae, Hewstone, & Griffiths, 1993; Stangor & Duan, 1991). For example, Macrae et al. (1993) had subjects view recordings of a conversation between two women, with one woman as the target confederate answering questions. In these recordings, the target provided either stereotypical or non-stereotypical information about herself. In the high cognitive load condition, the researcher gave participants the secondary task of remembering an eight-digit number, which participants performed concurrently to watching the recording. In the low cognitive load condition, participants were not given a secondary task. After watching the video, all participants were asked to recall information about the target confederate. Participants in the high cognitive load condition recalled more stereotypical consistent information than inconsistent information. However, participants in the low cognitive load condition remembered more stereotypical inconsistent information than consistent information. These findings support that under high cognitive load, people experience easier recall of information congruent to established stereotypes. The focus of the current study will concentrate on cognitive load and its relationship with discrimination in an interview context.

Manipulating Cognitive Load

Researchers require manipulations for cognitive load to create cognitively distinct groups. Task-based manipulations, such as the secondary task technique, are a reliable and sensitive method to determine cognitive load (Paas et al., 2003). In the secondary task method, participants perform an additional activity simultaneous to the primary task. This additional task acts to impose further strain on participants' cognitive resources. Typically, these are simple activities that involve sustained attention, such as remembering an eight digit number (Macrae et al., 1993), detecting a stimulus (Paas et al., 2003) or reading and recalling information from a passage (Logie, Baddeley, Mane, Donchin, & Sheptak, 1989). However, secondary tasks can

vary greatly. Logie et. al. (1989) explored the effects of various cognitive load manipulations on performance in a complex computer game. Cognitive load manipulations that affected performance included: asking participants to repeat the day of the week presented, stating the day of the week that follows the presented day, slowly tapping a foot pedal every three seconds, alternating tapping between two foot pedals, recalling word lists, and identifying the subject or object within sentences. In the present study, I used a secondary task I designed to imitate certain interviewing tasks.

Measuring Cognitive Load

In order to further investigate cognitive load in interviews, a measure of cognitive load was necessary. The two main types of cognitive load measures are self-rated scales and physiological techniques. Self-rated scales are by far the most popular measure of cognitive load (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). These subjective scales are based on the assumption that individuals can self-reflect on their cognitive processes and report on the amount of cognitive efforts expended (Gopher & Braune, 1984). Previous studies have shown that self-rated scales are reliable measures and are sensitive to small changes in cognitive load (Paas et al., 1994). For example, Paas et al. examined training and cognitive load in students learning and solving mathematical problems. Paas et al. provided students with a set of mathematical problems ranging in complexity. Simple problems reflected problems practiced in the prior instruction phase; complex problems had a different format, and required additional problem solving skills not provided in the instruction phase. Before the study and after each question, students rated their cognitive load on a 9-point scale, which ranged from "very, very low mental effort" to "very, very high mental effort". Afterwards, Paas et al. examine the reliability and sensitivity of the cognitive load rating scale. Reliability was determined based on the internal

consistency of the ratings, while sensitivity was determined by the scale ability to distinguish the problems' complexity. Student's subjective ratings of cognitive load reflected the problem complexity, and reveal reliabilities (Cronbach α) of 0.82 – 0.90, suggesting that people are capable of self-evaluating cognitive load.

Physiological techniques are another means of measuring cognitive load. Physiological techniques measure changes in physical variables (i.e. heart, brain and eye activity) with the assumption that changes in cognitive functions will manifest in physiological changes (Paas et al., 1994; Paas et al., 2003). Although less subjective, physiological techniques can be difficult to use outside of traditional laboratory research environments. Researchers interested in cognitive load can use either of these measures. In the current study, I only included a self-reported measure of cognitive load as I administered the study using an online platform.

The Present Study

Interviews are cognitively demanding tasks. However, factors such as the interviewer's experience, interviewer's preparedness, the interviewee's delivery, or the interview itself can cause variability in terms of the cognitive load experienced by the interviewer. Currently, there is a gap in the literature directly investigating the effects of cognitive load on discrimination within an employment interview setting. High cognitive load may jeopardize the structured interview's resistance to discrimination. The field requires more research on this topic in order to improve the accuracy of interviews and hiring decisions. Although previous research supports the relationship between cognitive load and the use of stereotypes (Fiske & Lee, 2008; Hilton & von Hippel, 1996; Macrae et al., 1993), limited research has investigated the influence of cognitive load on interviewers specifically. Interviews are a cognitively demanding situation with large

financial and personal implications for both sides, especially interviewees. For the above reasons, I examined whether cognitive load influences interviewers' likelihood to discriminate during job interviews. Particularly for positions associated with strong stereotypical norms. I investigated whether increased cognitive load of hiring managers impacts their decision when selecting between two equally qualified, but gender different candidate, such that each candidate differs in gender congruency to the job. Congruency refers to the fit between the candidate's gender and the job stereotypes. With the rationale that high cognitive load promotes stereotypes, and that stereotyping is necessary for discrimination, I hypothesize the following:

H1a: There will be an interaction between congruency (congruent/incongruent) and cognitive load (high/low) such that: interviewers in the high cognitive load condition will rate incongruent candidates lower than will interviewers in the low cognitive load condition (Appendix E, comparing cell C to D).

H1b: Interviewers in the high cognitive load condition will not rate the congruent candidates differently than interviewers in the low cognitive load condition (Appendix E, comparing cell A to B).

H2: Interviewers in the high cognitive load condition will, as a whole, hire a lower frequency of incongruent candidates, compared to interviewers in the low cognitive condition (Appendix E).

I tested these hypotheses by manipulating participants' cognitive load and the congruency between the candidate's gender and the job position's stereotypes. Specifically, I used a 2 (cognitive load) x 2 (congruency) between-within study design. Participants were placed in either a high or low cognitive load condition. Every participant rated both a congruent and

incongruent candidate through an online survey administrated through a crowdsourcing platform. Olian, Schwab, & Haberfeld (1988) revealed potential advantages of using a within-subject design and a professional sample in their meta-analysis, through their exploration of the effects of applicant's gender on ratings. Their results showed a stronger effect of gender for within-subject design ($d = 0.69, k = 8, n = 674, SD = .35, CI = .01, 1.37$) than between-subject design ($d = 0.22, k = 11, n = 1168, SD = .10, CI = .01, .43$). An online platform also provided me the opportunity to conduct my study within a limited timeframe with a professional sample with hiring experience. Olian et al.'s results also showed greater homogeneity within the student sample ($d = 0.36, k = 6, n = 643, SD^2 = .02, CI = .08, .64$) compared to the profession sample ($d = .40, k = 12, n = 1139, SD^2 = .15, CI = -.36, 1.16$), as the profession sample showed a greater corrected variance and a wider confidence interval despite having almost twice the sample size. Employing a between-within design and a sample with hiring experience was helpful in leveraging a potentially greater effect size and a more representative sample.

Method

Participants

The target participants are managers with previous interviewing experience, as they are likely to participate in real recruitment and hiring decisions. We included 366 participants (194 female, 170 male, 1 undeclared) from Mechanical Turk. Only participants who had previously completed over 500 Mturk tasks with a 95% success rate were allowed to participate. Participants previously conducted an average of 36.75 interviews ($SD = 70.68$). Average age of participants was 40.33 ($SD = 11.89$). Participants took on an average of 19.82 minutes to complete the survey.

The occupational composite of Mechanical Turk's participant pool appears comparable to the Cooperative Congressional Election Survey (CCES) data (Huff & Tingley, 2015). Management represents 11.94% of participants for Mechanical Turk in comparison to the 15.69% observed in CCES. Given the data collection format, possible selection prejudice towards respondents with skills to operate survey tools, ease with internet survey and access to computer network may have occurred (Zhang, 2000). The advantage of surveying current managers from a diverse range of fields justifies the use of a crowdsourcing internet marketplace for this study. I provided participants who passed the initial selection and attention checks questions with compensation of \$0.75 USD for their participation in the study. Payment on Mechanical Turk (MTurk) rarely exceeds \$1.00 (Wu, Corney, & Grant, 2014) and compensation of \$0.50 has been shown to be sufficient for a 30 minute long survey (Buhrmester, Kwang, & Gosling, 2011). Given that the experiment requires skilled participants with hiring experience, I set the compensation on the higher end of the spectrum.

To determine target sample size for this study, I performed a safe guard power analysis based on the studies by Kulik, Perry, & Bourhis, (2000), Macrae et al., (1993) and Wigboldus, Sherman, Franzese, & van Knippenberg (2004). The confidence interval for the weighted average of the studies' Cohen's F^2 was found ($F^2 = 0.071$), and a safe guard ANOVA power analysis on the lower bounds suggests an $n = 110$ per condition, for a total $n = 440$ (110 participants per cognitive load condition (high/low), per job position (electrical power-line apprentice/childcare worker)).

Materials

Mock interview development. I evaluated the presence of discrimination in hiring decisions by having participants listen to scripted and pre-recorded interviews. I provided

participants with a job description and a list of general requirements for the hypothetical position. Participants hired either for a stereotypically feminine position (childcare worker) or stereotypically masculine position (electrical power-line apprentice). I selected childcare worker and electrical power-line apprentice as the hypothetical positions, because these stereotypically gendered, entry level positions shared two core competencies of equal importance (National Center for O*NET Development, n.d.).

Participants listened to and evaluated interview voice recordings of two potential candidates. Based on their job experience and voice, both candidates appeared to be of the same perceived age range (early 20s). To further encourage gender stereotyping, I named the candidates common names for the candidates' intended age range. I named the female candidate "Jessica" and the male candidate "Matthew". Jessica was the most common name in the 1990s in the USA with 303,079 girls named Jessica. "Matthew" was also a highly common name in 1990s with 351,596 boys named Matthew (Social Security Administration (US), n.d.).

Script development. During the mock interview, each candidate answered the same two non-technical behavioural questions for both job positions (Appendix A). O*Net's (National Center for O*NET Development, n.d.) job descriptions rated the competencies of monitoring and critical thinking as equally important for both occupations. Thus, each candidate answered one question regarding monitoring and another on critical thinking. I created two sets of answers for each position based on real examples gathered through interviews of individuals who match the target age group of the hypothetical candidates. Afterward, a focus group of four subject matter experts (SME) helped refine both sets of scripts. Next, I tested the scripts with a second group of 19 SMEs to ensure equal quality (Appendix D). During the script testing, SMEs reviewed both scripts in text-format online. SMEs provided quantitative evaluations of the scripts' quality using

a 10-point scale, qualitative feedback about the script's quality, and selected one candidate for hire for either an electrical power-line apprentice or childcare worker position. Out of the 19 SMEs, 9 SMEs selected the candidate associated with Script A, and 10 selected the candidate associated with Script B. SMEs also rated both scripts of similar quality.

Voice recording development. After creating the scripts, I hired four voice actors, two male and two female, in the targeted age range from the University of Guelph drama department. The recordings from the four voice actors were pilot tested. A sample of 26 heterogeneous participants were drawn using a snowball sampling approach over social media for the voice pilot test. Based on the results, I selected one male and one female voice actor who provided similar delivery of the scripts with similar quality. Participants also rated the actor's voice in terms of perceived ethnicity, age, sexual orientation, social economic status and attractiveness. Participants largely perceived both candidates to be of North American or European descent, around the age of 22-23, heterosexual, middle to upper middle social economic status and reasonably attractive.

To simulate an interview-like situation, participants listened to voice recordings of pre-recorded interview responses to structured behavioural questions. In order to help participants take on the role of the interviewer (as opposed to someone simply watching an interview), voice recordings included only the candidate's voice. The interview questions appeared in writing on the screen in order to prevent participants from inferring the presence of a separate interviewer.

Stereotype facilitation elements. I added various other components to encourage stereotyping within the experiment. Job descriptions were presented before the mock interview and included statements to help prime candidates to utilize traditional gender stereotypes (Appendix B). For example, the job description informed participants that the company is

interested in hiring either a caring, empathic and kind (communal) candidate or a competent, efficient and independent (agentic) candidate. (Hilton & von Hippel, 1996; Phelan, Moss-Racusin, & Rudman, 2008). We did not allow participants to rewind or pause the interview voice recordings in order to increase ecological validity, stimulate stereotyping, and further raise cognitive load. To create a more challenging hiring decision, we added ambiguity to the task through the candidates' qualifications. Candidates met all basic technical requirements, however their perceived age suggested inexperience in the role. Priming and demands on participants' cognitive load should have provoked participants to rely on stereotyping during their hiring decision.

Cognitive load manipulation. I manipulated participants' cognitive load using a secondary task that reflects the physical action of note taking--a typical task hiring managers often perform during interviews (Appendix F). Participants in the high cognitive load condition were asked to type out an unrelated passage, while performing the primary task. The passage was presented as a JPEG image to prevent participants from copying and pasting the text. I set the passages' length such that only a skilled professional typist could complete the secondary task within the given timeframe. Participants also self-rated their cognitive load on a ten-point scale immediately after completing the final hiring decision. Participants in the low cognitive load (control) condition did not perform a secondary task.

Procedure

Table 1

Experimental Procedure for the Proposed Study.

| Description |
|-------------------------------------|
| 1. Introduction & consent |
| 2. Task & job description (priming) |

| |
|---|
| 3. Instructions for the secondary task (High cognitive load group only) |
| 4. Interview 1 |
| 5. Attention check 1 and job position experimental check |
| 6. Interview 2 |
| 7. Attention check 2 |
| 8. Final hiring decision |
| 9. Self-rated cognitive load |
| 10. Optional written explanation for decision |
| 11. Stereotypic activation measure |
| 12. Gender experimental check |
| 13. General demographic survey |
| 14. Debrief & conclusion |

The study began with an introduction about the experiment. Participants were informed that the purpose of the experiment was to understand the impacts of different interview formats, specifically phone interviews (Appendix C). Participants answered an initial question regarding their experience with interviewing. Participants who reported previous interviewing experience were allowed to continue. Participant's type of interviewing experience was not considered, as we did not provide participants with behaviorally anchored rating scales or require participants to take notes, and so their familiarity with these tools became irrelevant. Next, participants listened to two mock interviews (one male, one female) for either an electrical power-line apprentice or childcare worker position. I randomized the order and candidates' scripts among participants. For example, participant 1 may have heard candidate A answering using script A and candidate B using script B, but participant 2 may have hear candidate A reading script B and candidate B answering with script A. Details about the mock interview were as described in the previous section. Before viewing the second candidate's response, participants had to answer two basic attention check questions based on the content of the candidate's answers and select the job position both candidates were currently applying for, to ensure conscious awareness of the job

position and interview recordings. The survey ended for participants who incorrectly answered the attention check questions or selected the wrong job position. Participants answered a second set of attention check questions after the second candidate's response. After providing their final hiring decision, an optional section appeared asking participants to explain their reasoning for their decision. Next, participants self-rated their cognitive load during the interview. As experimental checks, participants were asked to speculate about the purpose of the study, rate both candidates on gender traits to determine stereotype activation, identify the candidate's gender and estimate the portion of females in the occupation. Estimating the portion of females in the occupation helped to ensure participants associated the intended stereotypes with each job position. Finally, participants answered a general demographic questionnaire.

Manipulation condition. We informed participants in the manipulation (high cognitive load) condition that listening and evaluating candidates was their primary task, and transcribing the passage was their secondary task. We presented participants with instructions about the secondary task before beginning the primary task (listening to and rating a mock interview). I added a countdown clock to create greater pressure. Participants in the low cognitive condition did not perform this additional activity or have a countdown clock. As a manipulation check, all participants rated their cognitive load after the interview on a ten-point scale.

Congruency condition. In order to randomize gender congruency, I selected two job positions for participants to hire. Participants hired for either an electrical power-line apprentice or childcare worker position. Electrical power-line apprentice represented the male congruent and female incongruent job position, as men represent 98% of labour force for this occupation according to the Canadian National Household Survey (2011). Childcare worker represented the

female congruent and male incongruent job position as 97% of childcare workers are women (“National Household Survey,” 2011).

Measure

Candidate evaluation. Participants evaluated candidates by providing (1) individual question ratings for critical thinking and monitoring, (2) the overall ratings for both the congruent and incongruent candidate and (3) a final hiring decisions. For these ratings (1, 2), participants rated candidates on a ten-point scale (Appendix A). For the final hiring decision (3), participants selected the candidate that they would hire for the job position.

Discrimination. Discrimination is behaviourally advantaging one group over another (Fiske & Lee, 2008). In line with this definition, discrimination was theorized to manifest as the preference for the gender-congruent job candidate. Presence of discrimination was determined based on two aspects: candidate evaluation and congruency. Discrimination was observed by any of the three measures, such that the congruent candidate (1) received higher question ratings for critical thinking and monitoring, (2) received higher overall ratings or (3) were hired more often over the incongruent candidate.

Participants’ individual ratings for each competency and overall ratings were analyzed under hypothesis 1. Lower ratings for the incongruent candidate would represent discrimination. This discrepancy would represent an inaccurate assessment of candidates’ perceived quality due to candidate’s gender. In contrast, comparable ratings between congruent and incongruent candidates would represent an accurate assessment of candidates’ quality (i.e. no discrimination).

The final hiring decision acted as an additional measure of discrimination. Forcing participants to select from two equally qualified candidates enhances the difficulty and ambiguity

of the decision, which in turns placed pressure on participants to stereotype. Enabling participants to stereotype was necessary to produce a spectrum of discrimination. Final hiring of both candidates at similar frequency would represent no discrimination. Discriminating in the final hiring decision, despite similar ratings, would signify a reliance on stereotyping for ambiguous decisions, despite accurate assessment of candidates. The final hiring decision provides deeper insight into the interviewers' thought process.

Exploratory Outcome: Stereotype activation. I included an exploratory outcome of stereotype activation to ensure that participants were indeed activating and applying gender stereotypes during the interview. Using a 7-point scale, participants rated both the female and the male candidate on 6 traits: 2 gender neutral (observant and responsible), 2 masculine (competent and decisive) and 2 feminine (warm and caring). Higher ratings for the female candidate for the stereotypically feminine trait, higher ratings for the male candidate for the stereotypically masculine trait, and similar ratings for the neutral trait would support the activation and application of gender stereotypes during the interview.

Results

Exclusion Criteria

I excluded participants from both the control and manipulation conditions for failing basic technical or attention checks. The survey ended for participants who incorrectly answered the sound check question after three consecutive attempts (15) or reported no hiring experience (350). The survey ended for participants who incorrectly answered the first set of attention check questions after listening to the first candidate, and any participants who indicated hiring for the wrong job position (740). To ensure that the participants did in fact listen to both candidates' responses, I excluded participants who incorrectly answered two or more attention check

questions out of the total four attention checks (5). Awareness of the job position was necessary in order to discriminate towards or against stereotypical norms surrounding the job.

I had planned to exclude participants who incorrectly identified the gender of either candidate (Jessica as female and Matthew as male); however no participants were excluded at this step as all participants correctly identified the candidates' gender. To demonstrate awareness of gender stereotypes around each job norm, participants estimated the percentage of women in the occupation the participant was hiring for. I excluded 5 participants who estimated less than 50% of childcare workers were women and 6 participants who estimated more than 50% of electrical power-line apprentices were women. Over average, participants estimated 80.52% ($SD = 9.93$) of childcare workers and 17.75% ($SD = 9.75$) of electrical power-line worker were women. I also excluded participants with multiple attempts. I determine this based on the Mturk worker ID participants provided at the beginning of the survey. I excluded 40 participants for multiple attempts.

Participants who did not meet the minimum threshold for performance on the secondary task had their data omitted. This threshold was determine based on two criteria: number of character and presence of key phrases. I set the minimum threshold for characters as 1 character per 2.41 second. This was determined by a subject matter expert that had previous hiring experience, but minimal technological knowledge and English as their fourth language. We also conducted a search for the first two words in each passage as a measure to verify that participants did indeed transcribe the provided passage. Participants who met the minimum character threshold, but failed the search for the first two words had their passages manually checked. From this pool, we included participants if their third and fourth words were present. We did not evaluate spelling and grammar as part of this second round of key phrases evaluation. We

excluded 15 participants for not meeting the minimum threshold for performance on the secondary task.

We also excluded 9 participants for speculating that the purpose of the study was related to gender research. 366 participants remained after all of the data cleaning. Of these 366 participants, 96 participants hired for the childcare worker position and 88 participants hired for the electrical power-line worker position in the control condition. For the manipulation condition, 91 participants hired for the childcare worker position and 91 hired for the electrical power-line worker position.

Manipulation Check: Cognitive Load

As a check to confirm that the secondary task did increase cognitive load, I used participants' self-ratings of cognitive load to compare participants in the control condition ($M = 4.52$, $SD = 2.83$) to participants in the manipulation condition ($M = 4.84$, $SD = 2.54$). The confidence interval was long and thus inconclusive of any self-reported increase in cognitive load for the participants in the manipulation condition ($t = 1.12$, $df = 360.22$, $d = -0.12$, $CI = -0.32, 0.09$, $p = 0.26$).

As an exploratory analysis, I examined the number of participants that I removed from each cognitive load condition (high/low). Specifically, I compared the number of participants that were removed in each condition due to multiple attempts at the study, failing attentions checks, or failing technical checks. I speculated that more participants in the high cognitive load condition were removed because their task was more difficult compared to the low cognitive load condition. The sum of square deviation suggested that more participants in the high cognitive load condition were removed than the low cognitive load condition ($\chi^2 = 9.37$, $df = 1$). This suggest that the high cognitive load condition was indeed more cognitively demanding.

Exploratory Outcome: Stereotype Activation

As an exploratory outcome, I tested for stereotype activation by comparing ratings from the female candidate's gender traits to the male candidate's gender traits. I combined the two related items for each gender trait, such that decisive and competent ratings were combined to form the masculine trait, caring and warm ratings were combined to form the feminine trait, and observant and responsible ratings were combined to form the neutral trait. I performed these analyses separately for both low and high cognitive load conditions (Table 2). If participants rated the female candidate higher for the feminine trait, the male candidate higher for the masculine trait, and rated both candidates similarly for the neutral trait, this would support the activation and application of gender stereotypes during the interview.

For the low cognitive load condition, participants rated the female candidate higher than the male candidate for the feminine trait ($t = 4.02$, $CI = .21, .62$, $d = -0.30$, $CI_D = -0.53, -0.12$) and rated the female and male candidates similarly for both the neutral ($t = -1.82$, $CI = -0.24, 0.01$, $d = 0.13$, $CI_D = -0.08, 0.33$) and masculine traits ($t = -1.74$, $CI = -0.28, 0.02$, $d = 0.13$, $CI_D = -0.08, 0.33$). The confidence interval for the feminine trait was fairly narrow and indicated a small to moderate effect. The confidence interval for the neutral and male traits were longer and inconclusive, as it supported a small effect, no effect and a slight detriment.

For the high cognitive load condition, participants rated the female candidate higher than the male candidate on the feminine trait ($t = 4.34$, $CI = .22, .58$, $d = -0.32$, $CI_D = -0.55, -0.14$), the male candidate higher than the female candidate on the masculine trait ($t = -3.79$, $CI = -0.47, -0.15$, $d = 0.28$, $CI_D = 0.06, 0.47$), and the male candidate higher than the female candidate on the neutral trait ($t = -2.75$, $CI = -0.33, -0.05$, $d = 0.20$, $CI_D = 0.00, 0.41$). The confidence intervals were moderate in length, supporting a small to moderate effect. Descriptive statistics for the trait

evaluation for each candidate is listed in Table 2. Overall, the results suggest that stereotype activation occurred for the feminine trait in both cognitive load conditions and for the masculine trait in the high cognitive load condition.

Table 2

Descriptive Statistics for Stereotype Activation

| Candidate | Traits | Low cognitive load | | High cognitive load | |
|-----------|-----------|--------------------|----------------|---------------------|----------------|
| | | Mean | Std. Deviation | Mean | Std. Deviation |
| Female | Neutral | 2.13 | 0.77 | 1.94 | 0.79 |
| | Masculine | 1.95 | 0.90 | 1.79 | 0.93 |
| | Feminine | 2.00 | 0.96 | 1.88 | 0.91 |
| Male | Neutral | 2.25 | 0.69 | 2.13 | 0.76 |
| | Masculine | 2.08 | 0.81 | 2.10 | 0.80 |
| | Feminine | 1.58 | 1.16 | 1.48 | 1.02 |

Note. Low cognitive load $n=184$; high cognitive load $n = 182$

Data Aggregation

We collected our data using two script versions, two gendered occupations, and with participants of different genders. In order to aggregate the data, and examine congruency of candidate's gender to stereotypes of each job position, I performed analyses on the effect of the script version, participant's gender and job position.

Effect of script. To further ensure that both versions of the scripts were of equal quality, I analyzed the candidate's ratings separately for each gender candidate using t-tests. I compared the ratings for each script versions separately for each candidate (male, female). For this comparison, I collapsed cognitive load and job position conditions. Participants rated the male candidate reading script A ($M = 7.66$, $SD = 1.72$, $n = 171$) lower than his reading of script B ($M = 8.26$, $SD = 1.69$, $n = 195$, $t = 3.32$, $d = 0.35$, $CI = 0.14, 0.55$, $p = .00$). Participants also rated the

female candidate reading script A ($M = 7.43$, $SD = 2.00$, $n = 195$) lower than her reading of script B ($M = 8.08$, $SD = 1.51$, $n = 171$, $t = -3.50$, $d = -0.36$, $CI = -.57, -.15$, $p = 0.00$). The confidence interval for both candidates' script readings were moderate in size, indicating that the effect of script ranged from small to moderate. This suggests that the scripts were not of equal quality. However, as a precaution to circumvent the effects of unequal scripts, we counterbalanced the scripts between the cognitive load condition, hiring job position conditions, the gender candidates and the order in which the script was read. Any effect of scripts would have been diluted due to this counterbalancing.

Effect of participant's gender. I tested for an effect of participants' gender and the interaction between participant's gender and candidate's gender using a 2 (participant's gender) x 2 (candidate's gender) within-between subject ANOVA with participant's gender (male, female) as the between factor and candidate's gender (male, female) as the within factor on candidate's ratings. I performed this analysis separately for each job position as gender congruency depended on the job position, but collapsed across cognitive load conditions. Table 5 includes descriptive statistics on the effect of participant's gender. Results for the childcare position revealed little to no interaction between participant's gender and candidate's gender (Wilks' $\lambda = 1.00$, $F = .030$, $\eta^2 = .00$, $CI = .00, .02$, $p = .86$). The gender of the participant had little to no effect on ratings ($F = 5.377$, $\eta^2 = .028$, $CI = .00, .09$, $p = .02$). The gender of the candidate had little to no effect on the ratings (Wilks' $\lambda = .99$, $F = 1.67$, $\eta^2 = .01$, $CI = .00, .05$, $p = .20$). Results for the electrical power-line apprentice position revealed little to no interaction between participant's gender and candidate's gender (Wilks' $\lambda = 1.00$, $F = .40$, $\eta^2 = .00$, $CI = .00, .04$, $p = .53$). The gender of the participant had little to no effect on ratings ($F = 1.53$, $\eta^2 = .01$, $CI = .00, .05$, $p = .22$). The gender of the candidate had little to no effect on the ratings

(Wilks' $\lambda = .98$, $F = 3.18$, $\eta^2 = .02$, $CI = .00, .07$, $p = .08$). The confidence intervals for the gender of participant effects, gender of candidate effects and their interaction were all very short, signifying a very slight to no effect. Thus, the gender of the participant either did not or very slightly impact the degree of gender discrimination, as observed in the ratings. Participants rated both the candidate congruent and incongruent to their gender similarly for both job position conditions. These results align with the current literature (Frauendorfer & Schmid Mast, 2013; McCarthy et al., 2010), which supports no interaction between interviewer's and candidate's gender on ratings. Given that no interaction was observed, we collapsed participants' gender when analyzing the main hypotheses.

Effect of job position. I explored gender discrimination as the preference for the gender-congruent candidate using two stereotypically gender different job positions. I examined the effect of job position to determine whether gender congruency had a similar effect when hiring for two stereotypically gendered job positions. I analyzed the two cognitive load conditions separately, as participants may have been more likely to discriminate in the high cognitive load condition. In order to combine the data from both job positions when examining the main hypotheses, I examined the differential effects of the two gendered job position (childcare worker and electrical power-line apprentice) on hiring decisions using a 2 (job position) x 2 (candidate's gender) within-between subject ANOVA with job position as the between factor and candidate's gender as the within factor for each cognitive load condition (high/low) separately.

In the low cognitive load condition, ratings were similar for both the male ($M = 8.01$, $SD = 1.54$, $N = 88$) and female ($M = 7.91$, $SD = 1.64$, $N = 88$) electrical power-line apprentice candidates as well as the male ($M = 7.96$, $SD = 1.79$, $N = 96$) and female ($M = 7.90$, $SD = 1.72$,

$N = 96$) childcare worker candidates. The confidence intervals for the effect of job position ($F = .027$, $\eta^2 = .00$, $CI = .00, .02$, $p = .87$), gender of candidate ($F = .321$, $\eta^2 = .00$, $CI = .00, .03$, $p = .57$) and the interaction between candidate's gender and job position (Wilks' $\lambda = 1.00$, $F = .019$, $\eta^2 = .00$, $CI = .00, .02$, $p = .89$) were all very short, indicating very slight to no effects.

For the high cognitive load condition, the effects of job position ($F = 3.861$, $\eta^2 = .02$, $CI = .00, 0.08$, $p = .051$), gender of the candidate (Wilks' $\lambda = .969$, $F = 5.72$, $\eta^2 = .03$, $CI = .00, .10$, $p = .02$) and the interaction between candidate's gender and the job position (Wilks' $\lambda = 1.00$, $F = .050$, $\eta^2 = .00$, $CI = .00, .02$, $p = .82$) also revealed very short confidence intervals, signifying very slight to no effects.

These results reveal little to no effect of job position for either of the cognitive load conditions, which suggests that the job position participants hired for did not impact their hiring decisions. The results were also consistent with previous research (Davison & Burke, 2000; Frauendorfer & Schmid Mast, 2013), which suggests that interviewers tend to favour the candidate congruent to the job stereotype in both stereotypically male and female occupation. For this reason, we collapsed both job positions when analyzing the main hypotheses.

Hypothesis 1

Hypothesis 1 proposed that there will be an interaction between congruency (congruent/incongruent) and cognitive load (high/low) on interviewers' ratings of the candidates, such that interviewers in the high cognitive load condition will rate incongruent candidates lower than will interviewers in the low cognitive load condition, and interviewers in the high cognitive load condition will rate congruent candidates higher than will interviewers in the low cognitive load condition. I analyzed this interaction using three separate 2 (cognitive load) x 2 (congruency) within between subject ANOVAs, with cognitive load as the independent between

factor and candidate's congruency to job stereotypes as the independent within factor. This interaction was examined separately using three continuous dependent variables: critical thinking rating, monitoring rating and overall ratings. A summary of descriptive statistics can be found in Table 3, and a summary of the observed power can be found in Table 4. Higher ratings for the congruent candidate over the incongruent candidate would represent discrimination. An interaction between cognitive load and congruency, with a larger effect under high cognitive load than low cognitive load, would support a positive relationship between cognitive load and discrimination, such that greater cognitive loads may be associated with greater discrimination.

Table 3

Descriptive Statistics for Hypothesis 1

| | Cognitive Load Condition | Mean | Std. Deviation | N |
|-----------------------------|--------------------------|------|----------------|-----|
| Congruent Overall Ratings | Low | 7.95 | 1.63 | 184 |
| | High | 7.79 | 1.92 | 182 |
| | Total | 7.87 | 1.78 | 366 |
| Incongruent Overall Ratings | Low | 7.93 | 1.72 | 184 |
| | High | 7.75 | 1.83 | 182 |
| | Total | 7.84 | 1.78 | 366 |

For the overall rating, results revealed a short confidence interval for the effect of congruency, (Wilks' $\lambda = 1.00$, $F = .058$, $\eta^2 = .00$, $CI = .00, .011$, $p = .81$), a short confidence interval for cognitive load ($F = 1.432$, $\eta^2 = .00$, $CI = .00, .03$, $p = .23$) and a very short confidence interval for the interaction between congruency and cognitive load (Wilks' $\lambda = 1.00$, $F = .010$, $\eta^2 = .00$, $CI = .00, .01$, $p = .92$). These results are consistent with congruency having no effect up to a small effect on overall ratings.

For the critical thinking ratings, results also revealed a very short confidence interval for congruency (Wilks' $\lambda = 1.00$, $F = .806$, $\eta^2 = .00$, $CI = .00, .02$, $p = .37$), a very short confidence interval for cognitive load ($F = 3.014$, $\eta^2 = .08$, $CI = .00, .04$, $p = .08$) and a very short confidence interval for the interaction between congruency and cognitive load (Wilks' $\lambda = 1.00$, $F = .806$, $\eta^2 = .00$, $CI = .00, .02$, $p = .37$). These narrow confidence interval ranges signifies no effect up to a very slight effect of congruency on critical thinking ratings.

For the monitoring ratings, results revealed a very short confidence interval for congruency (Wilks' $\lambda = 1.00$, $F = 1.412$, $\eta^2 = .00$, $CI = .00, .03$, $p = .24$), a very short confidence interval for cognitive load ($F = 2.567$, $\eta^2 = .007$, $CI = .00, .03$, $p = .110$) and a very short confidence interval for the interaction between congruency and cognitive load (Wilks' $\lambda = 1.00$, $F = 4.187$, $\eta^2 = .01$, $CI = .00, .04$, $p = .04$). These very narrow confidence interval ranges suggest that congruency had no effect up to a very slight effect on monitoring ratings.

H1a: Hypothesis 1a proposed that there will be an interaction between congruency (congruent/incongruent) and cognitive load (high/low) such that: interviewers in the high cognitive load condition will rate incongruent candidates lower than will interviewers in the low cognitive load condition. Interviewers in the high cognitive load condition ($M = 7.75$, $SD = 1.83$) did not rate incongruent candidates lower than interviewers in the low cognitive load condition ($M = 7.93$, $SD = 1.72$). The confidence interval was fairly large in length and inconclusive, as it supported a small detriment, no effect and a small effect. Thus, H1a was not supported ($F = 1.02$, $d = .10$, $CI = -.10, .31$, $p = .31$).

Hypothesis 1b proposed that interviewers in the high cognitive load condition will not rate congruent candidates differently than interviewers in the low cognitive load condition. Interviewers in the high cognitive load condition ($M = 7.79$, $SD = 1.92$) rated congruent

candidates similarly to interviewers in the low cognitive load condition ($M = 7.95, SD = 1.63$). However, the confidence interval was fairly large in length and inconclusive, as it supported a small detriment, no effect and a small effect. Thus, H1b was weakly supported ($F = .79, d = .09, CI = -.11, .30, p = .37$).

Hypothesis 2

H2 stated that interviewers in the high cognitive load condition will hire fewer incongruent candidates, compared to interviewers in the low cognitive condition. I tested H2 by comparing the frequency of incongruent hires in the high cognitive condition ($n = 74$) to the frequency of incongruent hires in the low cognitive condition ($n = 95$) with a chi-square test. A summary of the observed power can be found in Table 4. The sum of square deviation supported H2 ($\chi^2 = 4.00, df = 1, p = 0.04$).

Discussion

The Effect of Cognitive Load on Gender Discrimination

In this study, question ratings acted as one of the measures of gender discrimination. Hypothesis 1 focused on the ratings and was not supported; the narrow CIs from the 2 (cognitive load) x 2 (congruency) within between subject ANOVAs for critical thinking, monitoring and overall rating suggest that participants did not discriminate, or very slightly discriminated towards the congruent candidate in either the high or low cognitively load conditions. Although a small interaction between congruency and cognitive load was observed for the monitoring ratings, it accounted for only 1% of the variance and the confidence interval includes zero; other main effects or interactions appeared non-existing up to very slight. Participants' overall average of candidate's ratings for both the congruent ($M = 7.97, SD = 1.78, n = 366$) and incongruent (M

= 7.84, $SD = 1.78$, $n = 366$) candidate seemed comparable to SMEs evaluations of the candidate's gender neutral written scripts ($M = 7.42$, $SD = 1.35$, $n = 19$). Thus, this suggests that participants correctly evaluated both candidates.

These results may also emphasize the structured interview's resistance to discrimination (Levashina et al., 2014; McCarthy et al., 2010). Past research has revealed only minimal effects of gender discrimination in structured interviews for all job positions (Levashina et al., 2014; McCarthy et al., 2010). However, gender discrimination has been shown to be more prominent when selecting for jobs associated with strong gender norms (Davison & Burke, 2000). These results suggest that structured interviews may continue to resist discrimination even for highly "gendered" job positions. However, further investigations directly comparing unstructured to structured interviews in hiring for highly gendered job positions is needed to support this explanation.

Structured interviews likely cause greater cognitive load than unstructured interviews for the interviewer. In comparison to unstructured interviews, structured interviews should require greater intrinsic load due to the additional components that promote structure within the interview (e.g. note taking, evaluation candidates against pre-establish bars, separately rating each individual question). These additional elements of structure likely cause extra strain on interviewers' mental resources beyond the strain needed for unstructured interviews. People tend to stereotype more under high cognitive load (Bodenhausen & Lichtenstein, 1987; Macrae et al., 1993; Stangor & Duan, 1991). During times of limited cognitive resources, stereotypes' efficiency becomes especially beneficial. However, even while performing an unrelated secondary task, the ratings from the current study's structured interview revealed no evidence of discrimination. The fairly comparable ratings between the cognitive load conditions suggest that

the benefits from imposing structure outweigh the increased cognitive load cost in terms of probability to discriminate.

Hypothesis 2 focused on this study's second measure of discrimination, the final hiring decision. Unlike the ratings, the results from the final hiring decisions did support hypothesis 2; participants hired fewer incongruent candidates in the high cognitive load condition than did participants in the low cognitive load condition. Despite both being measures of discrimination, the final hiring decision and the ratings presented two different perspectives. The inherent difference between each measure's decision making processes could be cause for the observed discrepancy. Compared to rating a candidate on a ten-point scale, selecting an individual candidate to hire may have demanded greater intrinsic cognitive load, due to the ambiguity in the decision. The final hiring decision, unlike the ratings, did not provide participants with the opportunity to accurately assess the candidates on an individual level by providing equivalent scoring. It may have been challenging for participants to select only one candidate, as both candidates were similarly qualified and limited information about the candidates were provided. As a result, there was more of a tendency to stereotype for the final hiring decision, as compared to the individual question ratings.

In regards to the different cognitive load conditions, discrimination was observed only in the high cognitive load condition. This provides evidence that reaffirms Devine's (1989) dissociation model, as all of the participants were aware of the gender stereotypes, as confirmed by the job and stereotype activation exploratory outcome. However, participants in the high cognitive load condition probably experienced difficulty suppressing the stereotype due to the strain on their cognitive resources. As a result, greater discrimination was observed in the high cognitive load condition. These results further confirmed the relationship between stereotypes

and cognitive load (Bodenhausen & Lichtenstein, 1987; Macrae et al., 1993; Stangor & Duan, 1991).

Manipulation check: Cognitive Load

Despite having to complete an additional task, participants in the high cognitive load condition reported experiencing the same cognitive load as the participants in the low cognitive load condition. These results appear to contradict previous research as self-ratings are a common measurement of cognitive load. The literature on cognitive load supports people's capability to self-rate their experienced cognitive load (Paas et al., 2003, 1994). However, past studies often employed self-rated measures of cognitive load in the context of a within-subject design (Brünken, Steinbacher, Plass, & Leutner, 2002; Paas et al., 1994; Révész et al., 2015). In this study, we evaluated the manipulation of cognitive load in a between-subjects design. Perhaps people can make comparative judgments regarding cognitive load, but struggle with making absolute ratings in the absence of specific anchors. An anchored rating scale may help to improve the reliability and validity of self-rated measurements of cognitive load.

Implications for Research

The present study builds on the current understanding of structured interviews and discrimination. The study supports the utility of structured interviews by providing initial evidence of structured interviews' resistance to discrimination, even when selecting for a highly gendered job position. The fit between candidates and gender stereotype associated to job positions has been showed to skew raters towards the congruent candidate (Davison & Burke, 2000; Koch, D'Mello, & Sackett, 2015); however this congruency has been rarely considered in previous studies examining structured interviews (Levashina et al., 2014; McCarthy et al., 2010). This study also highlights the distinction between ratings and hiring decisions. The selection

literature heavily relies on ratings as a proxy for hiring decision; however, providing ratings may be less cognitive loaded and ambiguous than selecting an actual candidate. The current study illustrates the differences between ratings and final hiring decision. Previous studies that rely on ratings as a proxy for hiring decision may not reflect hiring decisions made in an applied setting. Small effects, such as discrimination, may appear more prominent when evaluated using measures such as a final hiring decisions.

Implications for Management

Beyond research, the present study suggests some practical implications for management. One key implication is the need for multi-step strategies for minimizing discrimination within the selection process. Although the current study demonstrates further support of structured interviews resilience against discrimination, it also stresses the need to incorporate additional strategies to prevent discrimination from contaminating the final hiring decision. All selection processes require a final hiring decision to determine the selected candidate. Without a structured protocol, the final hiring decision begins to parallel a holistic, overall evaluation that Campion et al. (1997) tends to advise against. These holistic evaluations are more susceptible to discrimination.

This study also highlights the importance of reducing the cognitive load in interviewers. Providing enough resources for interviewers to prepare for interviews, such as time for practicing and training, could act as one method of reducing cognitive load. Previous research suggests that trained interviewers demonstrate better reliability, validity and report more positive users' experiences (Campion et al., 1997). Relying on interviewers to seek out assistance may prove ineffective. Interviewers may struggle with accurately assessing their cognitive load, as indicated by participants' failure to objectively assess their cognitive load. Therefore, organizations should

take strives to objectively minimize the cognitive load of their interviewers in order to discourage discriminatory decision making. Minimizing cognitive load and introducing subsequent selection procedure, can help promote greater organizational diversity.

Limitations and Future Directions

The platform for the current study lends itself to a few limitations in terms of generalizability. The study uses an online simulation to mimic an employment interview. Unlike a real employment interview where interviewers select a candidate to join their organization, participants did not anticipate any future engagement with the candidates. The lack of future engagement limited the amount of risk involve in the ratings and hiring decision. Participants also did not belong to the hiring organization. Interviewers may be more likely to demonstrate discriminatory actions if they feel that hiring a gender incongruent candidate would violate their accustomed organizational norms. An applied sample may reveal greater discrimination than observed in this study.

Due to the online format, we also included attention checks for each response to ensure a general awareness of the candidate's responses. We removed more participants in the high cognitive load condition than in the low cognitive load condition. Participants in the high cognitive load condition likely needed more cognitive resources to pass the attention check than the participants in the low cognitive load condition. Participants in the high cognitive load condition may have possessed greater initial cognitive load resources than participants in the low condition load condition. Thus, the participant in the high cognitive load condition may have been better equipped to suppress gender stereotypes and not discriminate. Reproducing the effects in an applied setting is needed to support the study's generalizability.

Future studies may also wish to explore structured interviews' individual components and their effects on discrimination. The incremental value of each component of structure may differ. Determining the value added by each component has both theoretical and practical implications, as both researchers and practitioners rarely use all 15 components of the structured interview (Levashina et al., 2014). Understanding which components of structure provides the most value in terms of reducing discrimination would help inform researchers and practitioners which components of structure to include in order to maximize situations with limited resources. Better understanding the components of structure is a potential avenue for future research in selection.

Conclusion

Research investigating the relationship between cognitive load and discrimination in job interviews holds many potential benefits in promoting organizational diversity. The current study suggests that interviewers may be at an increased risk of discriminating when under higher cognitive load. Providing interviewers sufficient resources for interview preparation may help alleviate unnecessary cognitive load. Including supplementary strategies may also facilitate structured interviews' effectiveness in reducing discrimination and promoting organizational diversity. While many organizations may include diversity statements within their organizational mandate, their actual selection of candidates is often unreflective of such goals (Kang, Decelles, Tilcsik, & Jun, 2016). Reducing stereotyping tendencies within the selection process is essential in minimizing barriers impeding minority group entry and promoting actual long-term organizational diversity. In today's changing global market, organizations need flexibility and diversity to survive.

References

- Ahmed, A. M., Andersson, L., & Hammarstedt, M. (2012). Does age matter for employability? A field experiment on ageism in the Swedish labour market. *Applied Economics Letters*, *19*(4), 403–406. <http://doi.org/10.1080/13504851.2011.581199>
- Bodenhausen, G. V., & Lichtenstein, M. (1987). Social stereotypes and information-processing strategies: the impact of task complexity. *Journal of Personality and Social Psychology*, *52*(5), 871–880. <http://doi.org/10.1037/0022-3514.52.5.871>
- Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology*, *49*(2), 109–119. <http://doi.org/10.1027//1618-3169.49.2.109>
- 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. <http://doi.org/10.1177/1745691610393980>
- Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology*, *50*, 655–702.
- Canadian Human Rights Act (1977). Parliament of Canada.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), 293–332.
- Cohen, S. L., & Bunker, K. A. (1975). Subtle effects of sex role stereotypes on recruiters’ hiring decisions. *Journal of Applied Psychology*, *60*(5), 566–572. <http://doi.org/10.1037/0021-9010.60.5.566>
- Davison, H. K., & Burke, M. J. (2000). Sex discrimination in simulated employment contexts: A meta-analytic investigation. *Journal of Vocational Behavior*, *56*, 225–248.

<http://doi.org/10.1006/jvbe.1999.1711>

Debue, N., & van de Leemput, C. (2014). What does germane load mean? An empirical contribution to the cognitive load theory. *Frontiers in Psychology, 5*, 1–12.

<http://doi.org/10.3389/fpsyg.2014.01099>

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5–18. <http://doi.org/10.1037/0022-3514.56.1.5>

Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science, 11*(4), 315–319. <http://doi.org/10.1111/1467-9280.00262>

Fiske, S. T., & Lee, T. L. (2008). *Diversity at work*. (A. P. Brief, Ed.). New York: Cambridge University Press.

Frauendorfer, D., & Schmid Mast, M. (2013). Hiring gender-occupation incongruent applicants. *Journal of Personnel Psychology, 12*(4), 182–188. <http://doi.org/10.1027/1866-5888/a000095>

Gawronski, B., & Creighton, L. a. (2013). *Dual Process Theories*. (D. E. Carlston, Ed.). New York: Oxford University Press.

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors: The Journal of the Human Factors and Ergonomics Society, 26*(5), 519–532. <http://doi.org/10.1177/001872088402600504>

He, H., & Brown, A. D. (2013). Organizational identity and organizational identification: A review of the literature and suggestions for future. *Group & Organization Management, 38*(1), 3–35. <http://doi.org/10.1177/1059601112473815>

Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology, 47*, 237–71.

- Hornsey, M. J. (2008). Social identity theory and self-categorization theory: A historical review. *Social and Personality Psychology Compass*, 2, 204–222. <http://doi.org/10.1111/j.1751-9004.2007.00066.x>
- Huff, C., & Tingley, D. (2015). “Who are these people?” Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics*, 2(3), 1–12. <http://doi.org/10.1177/2053168015604648>
- Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2013). Employment interview reliability: New meta-analytic estimates by structure and format. *International Journal of Selection and Assessment*, 21(3), 264–276. <http://doi.org/10.1111/ijsa.12036>
- Huffcutt, A. I., & Roth, P. L. (1998). Racial group differences in employment interview evaluations. *Journal of Applied Psychology*, 83(2), 179–189. <http://doi.org/10.1037/0021-9010.83.2.179>
- Janz, T. (1982). Initial comparisons of patterned behavior description interview versus unstructured interviews. *Journal of Applied Psychology*, 67(5), 577–580. <http://doi.org/10.1037//0021-9010.67.5.577>
- Jetten, J., Spears, R., & Postmes, T. (2004). Intergroup distinctiveness and differentiation: A meta-analytic integration. *Journal of Personality and Social Psychology*, 86(6), 862–879. <http://doi.org/10.1037/0022-3514.86.6.862>
- Kang, S. K., Decelles, K. A., Tilcsik, A., & Jun, S. (2016). Whitened résumés: Race and self-presentation in the labor market. *Administrative Science Quarterly*, 1–34. <http://doi.org/10.1177/0001839216639577>
- Koch, A. J., D’Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied*

Psychology, 100(1), 128–161. <http://doi.org/10.1037/a0036734>

- Kulik, C. T., Perry, E. L., & Bourhis, A. C. (2000). Ironic evaluation processes: effects of thought suppression on evaluations of older job applicants. *Journal of Organizational Behavior*, 21, 689–711. [http://doi.org/10.1002/1099-1379\(200009\)21:6<689::AID-JOB52>3.0.CO;2-W](http://doi.org/10.1002/1099-1379(200009)21:6<689::AID-JOB52>3.0.CO;2-W)
- Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology*, 65(4), 422–427. <http://doi.org/10.1037/0021-9010.65.4.422>
- Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology*, 72(2), 275–287. <http://doi.org/10.1037/0022-3514.72.2.275>
- Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: narrative and quantitative review of the research literature. *Personnel Psychology*, 67, 241–293. <http://doi.org/10.1111/peps.12052>
- Logie, R., Baddeley, A., Mane, A., Donchin, E., & Sheptak, R. (1989). Working memory in the acquisition cognitive skills. *Acta Psychologica*, 71, 53–87.
- Macrae, C. N., Hewstone, M., & Griffiths, R. J. (1993). Processing load and memory for stereotype-based information. *European Journal of Social Psychology*, 23(1), 77–87. <http://doi.org/10.1002/ejsp.2420230107>
- McCarthy, J. M., Van Iddekinge, C. H., & Campion, Michael, A. (2010). Are highly structured job interviews resistant to demographic similarity effects? *Personnel Psychology*, 63, 325–359.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied*

Psychology, 79(4), 599–616.

- McDermott, M. L. (1997). Voting cues in low-information elections: Candidate gender as a social information variable in contemporary United States elections. *American Journal of Political Science*, 41(1), 270–283.
- McKenzie-Mohr, D., & Zanna, M. P. (1990). Treating women as sexual objects: look to the (gender schematic) male who has viewed pornography. *Personality and Social Psychology Bulletin*, 16(2), 296–308.
- Motowidlo, S. J., Carter, G. W., Dunnette, M. D., Tippins, N., Werner, S., Burnett, J. R., & Vaughan, M. J. (1992). Studies of the structured behavioral interview. *Journal of Applied Psychology*, 77(5), 571–587. <http://doi.org/10.1037/0021-9010.77.5.571>
- Nadler, J. T., & Kufahl, K. M. (2014). Marital status, gender, and sexual orientation: implications for employment hiring decisions. *Psychology of Sexual Orientation and Gender Diversity*, 1(3), 270–278.
- National Center for O*NET Development. (n.d.). National Center for O*NET Development. Retrieved August 2, 2016, from <https://www.onetonline.org/>
- Nguyen, H.-H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93(6), 1314–1334. <http://doi.org/10.1037/a0012702>
- Olian, J. D., Schwab, D. P., & Haberfeld, Y. (1988). The impact of applicant gender compared to qualifications on hiring recommendations. *Organizational Behavior and Human Decision Processes*, 41, 180–195.
- Paas, F., Tuovinen, J., 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. <http://doi.org/10.1207/S15326985EP3801>

Paas, F., Van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, 79(1), 419–430.

<http://doi.org/10.2466/pms.1994.79.1.419>

Phelan, J. E., Moss-Racusin, C. A., & Rudman, L. A. (2008). Competent yet out in the cold:

Shifting criteria for hiring reflect backlash toward agentic women. *Psychology of Women Quarterly*, 32(4), 406–413. Retrieved from

<http://search.ebscohost.com/login.aspx?direct=true&db=pbh&AN=35139466&site=ehost-live>

Przygodzki-lionet, N., Olivier, J., & Desrumaux, P. (2010). The effects of facial attractiveness, sex, internality of applicants on hiring decisions for managerial and non-managerial jobs.

Studia Psychologica, 52(1), 53–57.

Pulakos, E. D., & Schmitt, N. (1995). Experience-based and situational interview questions:

Studies and validity. *Personnel Psychology*, 48, 289–307.

Révész, A., Michel, M., Gilabert, R., Gao, J., Mazgutova, D., Moresi, F., ... Vasylets, L. (2015).

Measuring cognitive task demands using dual-task methodology, subjective self-ratings, and expert judgments. *Studies in Second Language Acquisition*, 38, 1–35.

<http://doi.org/10.1017/S0272263115000339>

Schmidt, F., & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings.

Psychological Bulletin, 124(2), 262–274.

Smith, E. R., Stewart, T. L., & Buttram, R. T. (1992). Inferring a trait from a behavior has long-

term, highly specific effects. *Journal of Personality and Social Psychology*, 62(5), 753–759.

<http://doi.org/10.1037/0022-3514.62.5.753>

Social Security Administration (US). (n.d.). Top names of the 1990s. Retrieved March 20, 2017, from <https://www.ssa.gov/oact/babynames/decades/names1990s.html>

Stangor, C., & Duan, C. D. (1991). Effects of multiple task demands upon memory for information about social groups. *Journal of Experimental Social Psychology*, 27(4), 357–378. [http://doi.org/10.1016/0022-1031\(91\)90031-Z](http://doi.org/10.1016/0022-1031(91)90031-Z)

Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Biosocial Science*, 1(S1), 173–191.

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, 33(1), 1–39. <http://doi.org/http://dx.doi.org/10.1146/annurev.ps.33.020182.000245>

Taylor, P. J., & Small, B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational and past behaviour employment interview questions. *Journal of Occupational and Organizational Psychology*, 75(3), 277–294. <http://doi.org/10.1348/096317902320369712>

Wigboldus, D. H. J., Sherman, J. W., Franzese, H. L., & van Knippenberg, A. (2004). Capacity and comprehension: spontaneous stereotyping under cognitive load. *Radboud Repository*, 22(3), 292–309. <http://doi.org/10.1063/1.3033202>

Wu, H., Corney, J., & Grant, M. (2014). Relationship between quality and payment in crowdsourced design. *Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 499–504. <http://doi.org/10.1109/CSCWD.2014.6846895>

Zhang, Y. (2000). Using the Internet for survey research: A case study. *Journal of the American Society for Information Science*, 51(1), 57–68. [http://doi.org/10.1002/\(SICI\)1097-4571\(2000\)51:13.0.CO;2-W](http://doi.org/10.1002/(SICI)1097-4571(2000)51:13.0.CO;2-W)

Table 4*The Observed Power for the Two Main Hypotheses*

| | Target n | Actual Total n | Observed η^2/d (95% CI) | Observed Power |
|--------------------------------------|----------|-------------------|---------------------------------|-------------------|
| H1 (Overall Ratings) | 440 | 366 | .00 (.00, .01) | 0.05 |
| H1 (Critical Thinking ratings) | 440 | 366 | .00 (.00, .02) | 0.15 |
| H1 (Monitoring Ratings) | 440 | 366 | .01 (.00, .04) | 0.53 |
| H2 | 349 | 366 | | 0.52 |

Table 5*Effect of Participant's Gender*

| Hiring Job Position | Candidate's Sex | Participant's Sex | Mean | Std. Deviation | N |
|------------------------------|-----------------|-------------------|------|----------------|-----|
| Childcare worker | Female | Female | 8.05 | 1.71 | 100 |
| | | Male | 7.61 | 1.94 | 87 |
| | | Total | 7.84 | 1.83 | 187 |
| | Male | Female | 8.29 | 1.67 | 100 |
| | | Male | 7.79 | 1.77 | 87 |
| | | Total | 8.06 | 1.73 | 187 |
| Electrical power line worker | Female | Female | 7.68 | 1.79 | 95 |
| | | Male | 7.52 | 1.82 | 83 |
| | | Total | 7.61 | 1.80 | 178 |
| | Male | Female | 8.06 | 1.91 | 95 |
| | | Male | 7.70 | 1.50 | 83 |
| | | Total | 7.89 | 1.74 | 178 |

Note. The effect of participant's gender are calculated based on candidates overall hiring ratings

Monitoring

Definition: Assessing performance of yourself, other individuals, or organizations to make improvements or take corrective action.

Interview: Tell me about a time you had to assess the performance of another task/person/organization.

Please rate the quality of Jessica's/Matthew's answer based on Jessica's/Matthew's demonstration of the competency **critical thinking**.

| | | | | | | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Poor | | | | Average | | | | Excellent | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| <input type="radio"/> |

Please rate the quality of Jessica's/Matthew's answer based on Jessica's/Matthew's demonstration of the competency **monitoring**.

| | | | | | | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Highly Unlikely | | | | Neutral | | | | Highly Likely | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| <input type="radio"/> |

Final hiring decision

Please select which candidate you would hire for this position.

- Matthew
- Jessica

In the box below, please describe your reasoning for choosing your selected candidate.

Appendix B: Introduction to Job Position

Job Position: ELECTRICAL POWER-LINE APPRENTICE

For this study, we ask that you take on the role of the interviewer for a power company looking to hire an **electrical power-line apprentice**. This is an entry-level position.

Electrical power-line apprentices are often responsible for installing or repairing cables or wires used in electrical power or distribution systems. They may also erect poles and light or heavy duty transmission towers.

Given the nature of the work, a competent, efficient, and fairly independent candidate is preferable. You will listen to two interview recordings of two potential candidates for the electrical power-line apprentice position. Both candidates have the basic technical knowledge for this role. In the recordings, each candidate will answer two questions regarding two additional essential competencies required to be successful as an electrical power-line installer. We ask that you use the candidate's answer to assess them on these two competencies. These competencies are critical thinking and monitoring skills.

Listed below are the two interview questions for the **electrical power-line apprentice** position:

Competency: Critical Thinking

Describe a time when you had to work through a challenging task or situation. Please also describe your thinking process.

Competency: Monitoring

Tell me about a time you had to assess the performance of another task/person/organization.

Job Position: CHILDCARE WORKER

For this study, we ask that you take on the role of the interviewer for a new daycare company looking to hire a **childcare worker**. This is an entry-level position.

Childcare workers are often responsible for attending to children at schools, businesses, private households, and childcare institutions. They may perform a variety of tasks; such as dressing, feeding, bathing, and overseeing play.

Given the nature of the work, a warm, kind, and fairly nurturing candidate is preferable. You will listen to two interview recordings of two potential candidates for the childcare worker position. Both candidates have the basic technical knowledge for this role. In the recordings, each candidate will answer two questions regarding two additional essential competencies required to be successful as a childcare worker. We ask that you use the candidate's answer to assess them on these competencies. These two competencies are critical thinking and monitoring skills.

Listed below are the two interview questions for the **childcare worker** position:

Competency: Critical Thinking

Describe a time when you had to work through a challenging task or situation. Please also describe your thinking process.

Competency: Monitoring

Tell me about a time you had to assess the performance of another task/person/organization.

Appendix C: Study Overview

This study will consist of two sections.

For the first section, you will act as an interviewer for a hiring process. No technical expertise is necessary. We will provide you with a job description for the position and the competencies/skills you will be assessing in the candidates. Afterwards, you will listen to voice recordings of two real interview responses from two actual candidates. Please listen carefully and evaluate each candidate on the competencies/skills using the rating scales provided.

After listening to each of the candidate's responses, you will be asked two basic questions about the content in the candidate's responses.

The second section will consist of a demographic survey that will take place at the end of the study.

Please click the arrow below to begin.

Appendix D: Interview Scripts

Critical Thinking Script A

So, two years ago I worked at the front desk at a fitness centre. During the evening shift, one of my responsibilities was to make sure maintenance washed the towels. During one of the evening shifts, the maintenance staff decided to go on strike, so none of the maintenance staff showed up that night. I was the only one at the front desk, and I was tasked with having to control the front desk, which entails checking the security cameras, letting people in, checking the fitness equipment and dealing with customers' complaints and now also washing the towels. The reason why this was difficult was because of the way the gym was set up; the front doors are always locked. So any time gym goers wanted to come in, I would have to buzz them in. The problem with that was the towel baskets were on a different floor, so I had to run up and down between the different floors. I was worried at first, because I couldn't be in two locations at the same time, and a lot of customers were complaining about the towel situation.

The towel situation was something I knew I needed to address. Due to my past experience on the job, and our discussions in team meetings, I knew that it was definitely the number one source of customers' complaints, so I knew that not addressing the towel situation was going to be the wrong thing to do.

After I gathered my thoughts, the first thing I did was make an announcement to all the gym goers letting them know about the situation, and that I was alone trying to balance everything. I thought that if customers understood my situation, they might be empathetic and more patient with me. After I made that announcement, I asked one of the personal trainers, who was a friend and a co-worker, to help me out with the front door so I could handle this other pressing issue, because I needed immediate help to deal with the rush hour period, when customers were coming in for fitness classes or straight after work.

The trainer was ok with it. We had a really good professional relationship, so I knew he would be able to support me. There wasn't any hesitation, and he kept reassuring me that everything was okay; that he was happy to help.

In that moment, it was the best I could do.

It worked well. One of the things that cued me was that no one was really coming to the front desk to complain anymore. This was a big difference compared to the very beginning, before I had taken any of those solutions. Also my supervisor, after the fact, let me know that it was the right thing.

I was able to finish off the shift; and it was only during the rush hour period that I needed the personal trainer's help anyways, because that was the time when we had the most customers. For the end of the shift, I didn't need to go between floors so much. I was able to stay, and have a steady influx of towels to provide the clients. All in all, it ended pretty well.

Critical Thinking Script B

So, a time when I had to work through a challenging task or situation at work was one morning I came in to the pool, where I was the pool coordinator for a summer camp, and the pool was super cloudy; we couldn't even see the bottom. I taught swimming for 5 years and I've never seen this before. This was an issue because clarity is really important, just in case there is something or even someone at the bottom of the pool. So, having not ever encountered this before, I was a little bit confused and concerned. I decided to go talk to maintenance and see if they had any suggestions on how to improve the pool clarity.

Maintenance said that the cloudy water was most likely due to the chemicals used in the pool and the pool being too hot. We decided the best thing to do was lower the temperature of the pool a little bit and keep more life guards on duty to make sure we could clearly see everything we needed to see, because it was going to take a while for the cloudiness of the water to go down. The way it works when you are a lifeguard is you need to constantly be able to supervise everyone in the pool. Typically, our pool had about 70 kids in it, so we needed 3 guards on deck. We always have four guards on duty. While three of them are guarding, one of them goes on break to rest their eyes. Due to the visibility issue, it made sense to have more people closer to the water to observe from all areas in the deep end. I placed guards where we could make sure all of the bottom of the pool could be seen even with the decreased visibility. I figured that out by walking around the pool and looking at the bottom, and seeing where I had to stand to be able to see all of it from different angles.

By adding the fourth guard, we had to take away breaks while we were guarding. This meant that we had to vacate the pool a little bit earlier, so that the guards could rest their eyes before the next group came in. So instead of a 25 minutes swim, we gave them 20 minutes. This gave us a 10 minute break before the next group of kids came in, instead of 5.

The life guards were really great. They understood that the kids' safety was the most important thing, so even though their breaks were reduced by 5 minutes, they understood why.

I think I made the right decision. I think that, again, it's the safety of the children that is our primary goal. I believe my decision to add more guards on before the water cleared up was the safest thing to do.

There were no major incidents that day. The extra guard made sure our swimming session ran smoothly throughout the morning. Then, by the time we got back from our lunch break, the pool had cleared up, and so we could go back to our regular rotation.

This was the only time this happen during the entire summer.

Monitoring Script A

Ok, I'm going to pick an example where I monitored a person. I worked as a piano teacher for five years. Every week I would come in, and I would have to assess the performance of my students. I would assess them on whether they practiced well or not, whether they listened to the feedback from the week before, and whether they had motivation to continue practicing a certain piece or technique.

I had a high school student who was particularly stubborn, in that he only liked to play jazz. He was really good at jazz. He would learn jazz pieces in two or three days. He was really remarkable, but he wasn't interested in learning anything else. This included the basic techniques, like scales, which would help him play jazz better.

He would come in, warm up with scales, which is the best way to assess whether they have been practicing, because scales are very cut and dry. You either know it or you don't. Then he would move on to some of the more complex pieces and then from there, it was easy to assess his use of dynamics, his tempo, and his rhythm.

I know the piece myself, so I know how it was supposed to sound. If he deviated from that too much, then that meant he wasn't doing something right.

I used my assessments to create timelines for him. So, I would assess him on when he was going to be ready for an examination, when he was ready to perform, when he was ready to compete, etc. I wouldn't let him do any of that unless he was ready. After the second lesson, I could tell he wasn't incorporating my feedback in his performance, so I wouldn't let him move on to the next step.

I ended up having to repeat that lesson to him. And, in my eyes, that was not a good performance. It made me feel frustrated, because I spent a lot of time and effort teaching him, and he was not taking the time to internalize what I was teaching him.

I definitely wanted to remain calm, because I wanted to be professional. I tried to figure out if this was going to be a reoccurring problem. Why was this happening? I wanted to empathize with him a little. I also didn't want to discourage him, and make him hate playing piano.

The motivation in this particular student was very apparent. I would say a big portion of it was his non-verbal features. So when I spoke with him, he was looking away, disinterested, and I could tell he was not motivated to learn. Sometimes when I teach, I would like to tell a bit of background information on a particular piece and if he followed up with questions, or if he looked at me and nodded, that signaled that he was getting the information, and I would know that he was motivated. He didn't.

Apparently, he loved jazz so much that he would not play anything else other than jazz. So I would assess him very poorly based on this. Then, because of this poor assessment in one area and strengths in another area, I would try to improve his motivation and interest by incorporating recordings of well-known pianists from YouTube. He liked practicing those more. I also assessed him on his learning styles. I found out that he learned better from imitation than sight reading. So I would play something, then he would play something. I started incorporating his learning style more into my teaching. After changing my teaching style, he did improve a little, not as quickly as I would like, but he got better.

Monitoring Script B

I'll tell you about a time when I had to assess another person's performance. I was working in a restaurant, as one of the main leads of the takeout portion of the restaurant. As a lead, I was responsible, not only training new takeout people, but also assessing how they are doing in the job.

So, there was one individual who got hired for the job, and I had a feeling from the beginning that it was going to take him a while to sort of get the job.

The thing about the job, for takeout in particular, is that you need to work at a fast pace, because the restaurant I was working at was very busy. You really needed someone who hit the ground running. I had the feeling that this one person wasn't putting in a lot of effort.

I could tell from the body gestures, his facial expressions that he didn't want to be there.

For example, his body movements were pretty slow. I was trying to tell him or instruct him to do something, but he wouldn't be making eye contact, and it would take a while for it to register because he wasn't listening.

For example, I would be showing him the different types of dips we offered. I would be naming each one, and he wouldn't be paying attention. Often, our new takeout employees that we hire bring a notebook and a pen to write notes. I even offered those materials to him, but he didn't take them. So for me, those were the types of things that initially cued to me that he wasn't maybe taking this as seriously as he should have been.

So I talk to him about it at the end of the night.

It was part of my role as a takeout lead to have those kinds of conversation with the trainees after their first few shifts, just to check in on how they're doing.

I didn't want to come across too harsh. I pretty much just took him aside and sat him down, cuz I didn't want to have a big power differential between the two of us. I started off by asking him how he felt about the day. I wanted to gauge where he was coming from. I was thinking maybe he was not aware that this is how he was coming across. That was why I took a causal approach at first.

After, I ask him to be honest about how he really felt about the job.

He told me he didn't really want to be working. He was particularly young, like high school. He had let me know that his mom had just encourage him to get a job, but he didn't think he was ready to get a job, especially for this type of role. So then, I was able to confirm my assessment of his performance. After that conversation I told him, "listen, in order for you to thrive in this role you really need to want to be here, because we work here as a team, and your team members rely on you".

I also asked him what he wanted. Do you want to stay in this role? I can give you some one-on-one guidance? Or do you want to try to find a different role for you? Or maybe it would be better off for you to even leave, if that what you want? So I gave him options. And he end up choosing not to continue with the job.

I think, he felt relieve from the conversation.

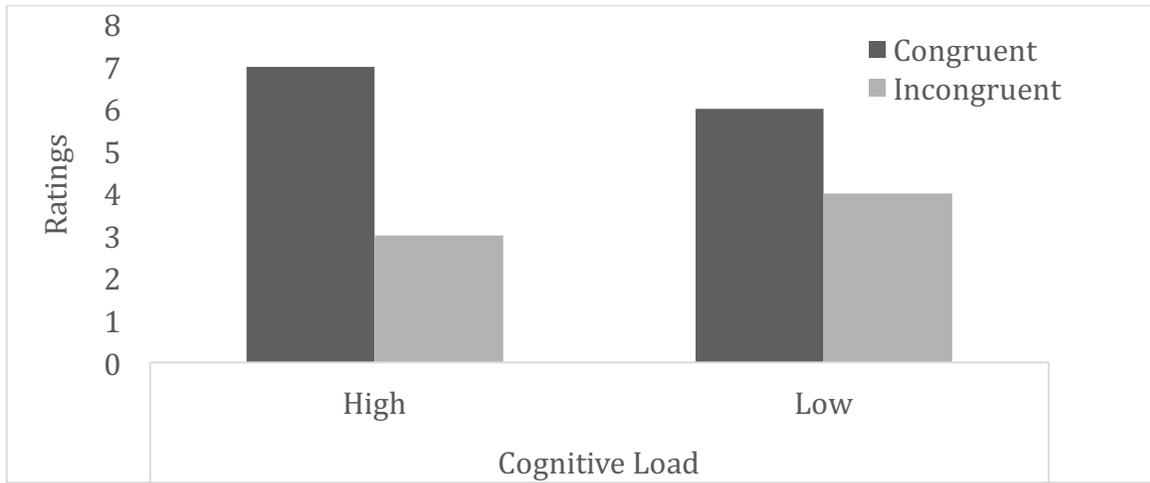
It was awkward at first, but I also wanted to get straight to the point. It was very obvious that customers were picking up on it. Even his co-workers were mentioning things to me. It was something I needed to address. It was awkward at first, but then I was able to hear his story and I felt good that we were able to together find a solution that was best for him, in his situation.

I feel like that an important part with assessing people is giving them feedback that sometimes can be hurtful to them, which is difficult. But, I try to be objective about everything, and not get my personal biases to affect me assess somebody and their work.

Appendix E: Hypothesis Results

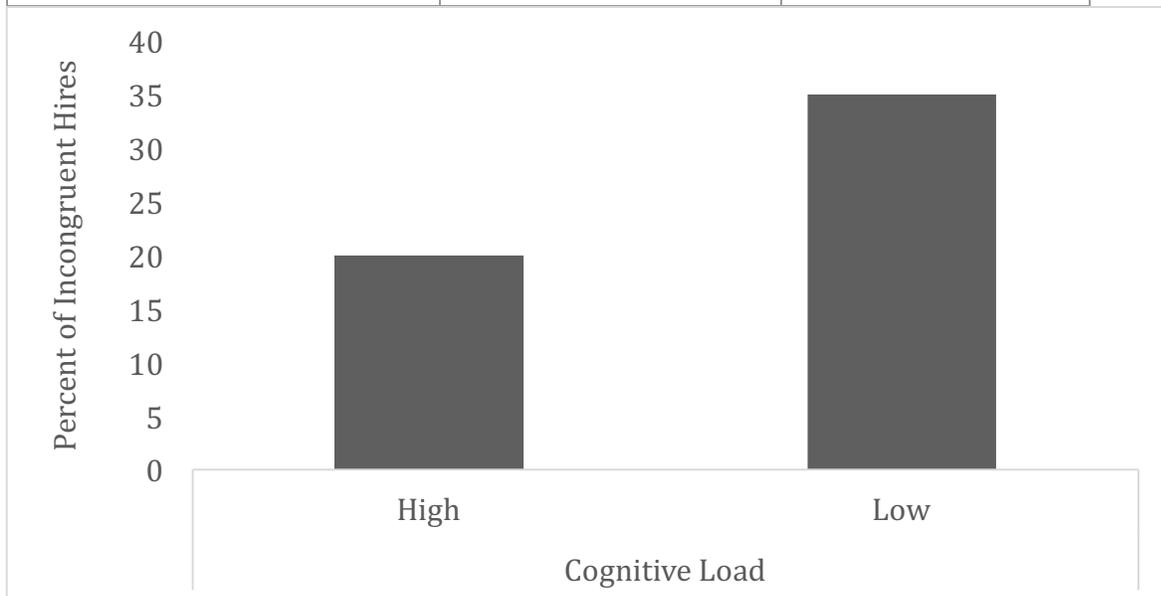
H1

| | Cognitive Load | | |
|-------------|----------------|-----|-----|
| | High | Low | |
| Congruent | A | B | H1b |
| Incongruent | C | D | H1a |



H2

| | Cognitive Load | |
|------------------------------|----------------|-----|
| | High | Low |
| Percent of Incongruent Hires | E | F |



Appendix F: Cognitive Load Manipulation Instructions

While interviewing the candidate, we ask that you complete the additional secondary task of typing out the presented paragraph. This task is used to simulate the action of note taking. **Your primary goal is to listen and evaluate the candidate; however, you do need to perform the secondary task simultaneously.** Please try your best to transcribe the entire paragraph.

Appendix G: Cognitive Load Manipulation Secondary Task

While interviewing, please type out the paragraph below in the provided space.

Wind is the movement of air from an area of high pressure to an area of low pressure. In fact, wind exists because the sun unevenly heats the surface of the Earth. Wind power is the ability to capture the wind in a way to propel the blades of wind turbines. When the blades rotate, this movement is switched into electrical current with the help of an electrical generator. In older windmills, wind energy turned mechanical machinery to do the physical work like crushing grain to make bread or pumping water to get water. Wind towers are built on wind farms, and usually there are several towers built together. In 2005, the worldwide use of wind-powered generators was less than 1% of all of the electricity use combined. There are several advantages of this energy source: there is no pollution, it never runs out, farming and grazing can still take place on the same land as the wind turbines, and wind farms can be built anywhere. As well, since the wind is free, operational costs are nearly zero once a turbine is erected. Mass production and technology advances are also making turbines cheaper. One disadvantage is that you need a consistent wind to get enough power. If the wind speed decreases, less electricity is produced. Large wind farms can also have a negative visual effect for people who live nearby.

While interviewing, please type out the paragraph below in the provided space.

Hydroelectric energy is the power that comes from the potential energy of water that is dammed up to supply energy to a water turbine and generator. Potential energy in the retained water is transferred into kinetic energy by water flowing through at high speeds. Another example of this energy is to make use of tidal power. Today, electric generators can be powered by hydro power that can run backwards as a motor to pump water for later use. Hydroelectric power provides almost all the energy for some nations. Norway, Brazil, and the Democratic Republic of Congo all get more than 90 percent of their electricity from hydroelectric power plants. An advantage is that you can control the use of the energy by controlling the water. You can also generate water all the time as there are no outside forces that prevent this from happening. Furthermore, there is no pollution in using this type of energy. In fact, you can reuse the water that is used for hydroelectric power. Using water as a source of energy is generally a safe environmental choice, however it is not perfect. Dams are expensive to build and maintain. There also needs to be a powerful enough supply of water in the area to produce energy. These man-made dams may also be obstacles for fish trying to swim upstream. Some dams have installed fish ladders to help fish migrate.

While interviewing, please type out the paragraph below in the provided space.

Geothermal means “earth heat”. This energy captures the heat energy under the Earth. Hot rocks under the ground help to heat water to produce steam. If holes are dug in this area of the ground, then the steam shoots up and is purified and used to drive turbines, which in turn gives power to electric generators. Geothermal energy is generated in over 20 countries. Geothermal energy has been used for thousands of years in some countries for cooking and heating. The first geothermally generated electricity was produced in Italy, in 1904. Geothermal energy can also be used to grow plants in greenhouses. The advantages of this type of energy is that there are no

harmful by-products, it is self-sufficient once the geothermal plant is built, and the plants are generally small so there is no negative visual effect on the area surrounding the plant. Unlike solar and wind energy, geothermal energy is always available, 365 days a year. One disadvantage is that using geothermal energy can be expensive.

While interviewing, please type out the paragraph below in the provided space.

Solar energy is used for heating, cooking, making electricity, and even taking salt out of saltwater so the water can be drinkable and used for additional purposes that do not need the salt. Solar power uses sunlight that hits the solar thermal panels to convert the sunlight to heat either air or water. Other methods of using solar power include simply opening up blinds or shades and letting the sunlight pass into the room or using some type of mirror to heat water and produce steam. One advantage of solar power is that it is renewable. As long as there is sunlight, you will be able to harness the power from it. There is also no pollution and it can be used efficiently to heat and light things. You can see the benefits of solar energy in heating swimming pools, spas, and water tanks in many cities across the country. It is an unlimited fuel source that is often noise-free. Solar energy use has increase at about 20 percent a year over the past 15 years. However, one disadvantage is that solar energy is expensive and require a lot of area to gather the sun's energy at rates useful to lots of people.

Appendix H: Experimental checks

Hiring Experience

Do you have hiring experience?

Please estimate the number of interviews you have performed in the past.

Job Position

Currently, you are interviewing for which of the following job position:

- Electrical power-line apprentice
- Childcare worker
- Dental Assistant
- Taxi Driver

Job Position Gender Stereotype

What percentage of electrical power-line apprentice do you think are women?

What percentage of childcare workers do you think are women?

Candidate's Gender

What gender was Candidate E?

What gender was Candidate J?

Appendix I: Demographic Questions

Please enter your age.

Please enter your sex.

Please select your occupational industry.

- Agriculture, plantations, other rural sectors
- Construction and infrastructure
- Mining (coal, other mining)
- Informal economy
- Food, drink, tobacco
- Forestry, wood, pulp and paper
- Oil and gas production, oil refining
- Basic metal production
- Chemical industries
- Mechanical and electrical engineering
- Transport equipment manufacturing
- Textiles, clothing, leather
- Health services
- Public Services
- Utilities
- Education and research
- Commerce
- Financial services, Professional services
- Hotels, catering, tourism
- Media, culture, graphical
- Postal and telecom services
- Shipping, ports, fisheries, inland waterways
- Transport (including civil aviation, railway, road)

Please check off all ethnicities which you identify with

- Aboriginal
- North American
- European
- Caribbean
- Latin, Central and South American
- African
- West Central Asian and Middle Eastern
- South Asian
- East and South East Asian
- Other Asian
- Oceania
- Other

Appendix J: Attention Check Questions

Based on the content in the candidate's answer, please select the correct answer.

At the **fitness center**, what problem did the maintenance staff strike cause?

- Towel shortage
- Flooded bathroom
- Missing snacks
- Rodent Infestation
- Injured clients

In the response about the **piano teacher**, what genre of music did the piano student want to play?

- Country
- Electric dance music
- Rap
- Jazz
- Hip hop

Based on the content in the candidate's answer, please select the correct answer.

In the response about the **pool coordinator**, what was discovered after arriving at the pool?

- Live animal
- Electrical power shortage
- Cloudy pool
- Broken computer
- Drained pool

In the response about the **take-out restaurant**, what was the new trainee decision about the job?

- Quit
- More training
- Increased hours
- Continue working
- Moved to a new city

Appendix M: Recruitment Ad

You will earn \$0.75 in approximately 30 minutes

Previous Hiring experience is required for this job. We will be asking you some questions about your hiring experience.

(Compensation is dependent on passing basic attention checks)

This job **must** be completed continuously, **without interruptions**.

This job must be completed from a **desktop** or **laptop computer** (No smartphones).

For this job, you will need the following:

- High-speed internet connection
- Sound Capabilities on your computer
- Adobe Flash Player
- Cookies must be enable

Your participation will help investigate the impacts of different interview formats, specifically phone interviews.

This job involves:

- Acting as interviewers for the non-technical component of the hiring process for a specific job position.
- Listening to voice recordings of two real interview responses from two actual candidates.
- Evaluating each candidate.
- Completing a general demographic survey.

Appendix N: Information Letter and Consent Form**INFORMATION LETTER****AND CONSENT FORM****INVESTIGATORS**

You may withdraw your consent at any time and discontinue participation without penalty. You are not waiving any legal claims, rights or remedies because of your participation in this project. This project has been reviewed and received clearance through a Project Ethics Board that adheres to the American Psychological Association guidelines for ethical project with human participants. If you have questions regarding your rights as a participant, contact:

Principal Investigator

Deborah *Powell*

(519) 824-4120, ext. 52167

dpowell@uoguelph.ca

Sulan Kith

(519) 824-4120, ext. 58931

skith@uoguelph.ca

Director, Project Ethics
(519) 824-4120, ext. 56606
sauld@uoguelph.ca

PURPOSE OF THE STUDY

The purpose of the experiment is to explore the impacts of different interview formats, specifically phone interviews. Data collected will help contribute to research about the selection process and the development of better hiring practices.

STUDY DESCRIPTION

You will be asked to listen to voice recordings of two interviewee candidates, both applying for the same specific job position. Each candidate will answer two interview questions based on two job competencies relevant to a job position. Following each interview question, you will evaluate the candidate based on the specific competency. After each candidate, you will provide an overall evaluation based on their suitability for the specific job position. Finally, you will be presented with a general questionnaire with demographic questions.

PARTICIPATION AND WITHDRAWAL

Participation in this study is voluntary. You may refuse to participate, refuse to answer any questions or withdraw from the study at any time during the survey. At the end of the survey, you are free to choose whether or not to participate and exercise your option of removing your data from the study. The investigator may withdraw you from this project if circumstances arise that warrant doing so.

CONFIDENTIALITY AND ANONYMITY

Every effort will be made to ensure confidentiality of any identifying information that is obtained in connection with this study. Any information you will be asked to provide will be linked to the task authentication code given to you by Mechanical Turk (MTurk) and the task completion code given to you at the end of this task. Only MTurk will have the ability to match these codes with your MTurk worker ID. MTurk will not have access to the responses that you make throughout your participation in this project, but they will be informed about the quality of the work that you complete. Members of the project team will not be able to link your responses to your MTurk worker ID or any other personal identifying information.

PAYMENT FOR PARTICIPATION

You will be paid \$0.80 USD for your participation.

POTENTIAL RISKS AND DISCOMFORTS

No more than minimal risks are involved in this project.

This research project will not involve any sensitive or incriminating questions, nor manipulation of participants' behaviour beyond the range of daily life.

Answering some of the interview questions may have a negative impact on your reputation or self-image if your identity were disclosed. The potential for your identity being disclosed is minimal.

CONSENT TO PARTICIPATE IN THE PROJECT

By entering the survey, I indicate that I have read the information provided and agree to participate.

Appendix O: Debriefing Form



DEBRIEFING FORM

Thank you for participating in our study. You have just participated in a research study conducted by Dr. Deborah Powell and Sulan Kith (Master's student), from the Psychology Department at the University of Guelph.

The results will contribute to a Master's thesis. You can find out more about the results by accessing Sulan Kith's Master's thesis in the University of Guelph's library atrium once it is submitted in the summer of 2017.

If you have any questions or concerns about the research, please feel free to contact Dr. Deborah Powell at dpowell@uoguelph.ca or (519) 824-4120, ext. 52167.

PURPOSE OF THE STUDY

The purpose of this study was not to investigate different interview formats, but to research the influence of cognitive load on discrimination within interviewers during job interviews. Specifically whether increased cognitive load on hiring managers will impacts their decision when selecting between two equally qualified candidates (of different genders) for positions associated with strong stereotypical norms. We apologies for the dishonesty and any discomfort we may have caused. Some deception was necessary for the purpose of this study. Again, we apologies for the deception.

CONFIDENTIALITY

Every effort will be made to ensure confidentiality of any identifying information that was obtained in connection with this study. Any information you provided will be linked to the task authentication code given to you by Mechanical Turk (MTurk) and the task completion code given to you at the end of this task. Only MTurk will have the ability to match these codes with your MTurk worker ID. MTurk will not have access to the responses that you make throughout your participation in this project, but they will be informed about the quality of the work that you complete. Members of the project team will not be able to link your responses to your MTurk worker ID or any other personal identifying information.

PARTICIPATION AND WITHDRAWAL

Participation in this study is voluntary. You may refuse to participate, refuse to answer any questions or withdraw from the study at any time during the survey. You are also free to choose whether or not to participate and exercise your option of removing your data from the study. The investigator may withdraw you from this project if circumstances arise that warrant doing so.

RIGHTS OF RESEARCH PARTICIPANTS

If you have questions regarding your rights as a research participant, contact:

**Research Ethics Officer
University of Guelph
437 University Centre
Guelph, ON N1G 2W1**

**Telephone: (519) 824-4120, ext. 56606
Email: sauld@uoguelph.ca
Fax: (519) 821-5236**