UNCONSCIOUS THOUGHT AND STEREOTYPES: HOW POSTERIORI STEREOTYPE ACTIVATION BIASES UNCONSCIOUS THOUGHT

by

Lucca Aleksandr Reiter

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Psychological Science

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2019

ACKNOWLEDGEMENTS

I would like to thank my family for their love, support, and the amazing examples they have set for me. Being raised by and among such remarkable people has been an extraordinary privilege. I would also like to thank my friends for understanding when I needed to be left alone, and knowing when they had to drag me out to blow of steam. Most of all, I would like to thank my amazing girlfriend, for always being my number one cheerleader. You are my rock.

I would also like to thank Keith Hutchison, Michelle Meade, and Jessi Smith for their valuable input and incredible patience. Special thanks to my Mentor, Ian Handley, for his carefully guidance, and patience with my many revisions. Lastly, I would like to thank my fellow graduate students for their support and friendship throughout this process.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

ABSTRACT

Unconscious Thought Theory (UTT) proposes that people can actively process goal-relevant information when they are distracted from consciously thinking about it (Dijksterhuis & Nordgren, 2006). Further, unconscious thought (UT) is purported to process the information in an aschematic bottom-up manner (Bos, Dijksterhuis, & van Baaren, 2008). Supporting this claim is experimental evidence that UT is less susceptible to stereotype use compared to the top-down schema-driven process of conscious thought (CT; Bos & Dijksterhuis, 2011). Based on these findings, UTT has proposed that UT does not utilize stereotypes when forming impressions. However, other research suggests that participants may form biased impressions in experimental contexts arguably suitable to the operation of UT if a stereotype is non-consciously activated during distraction periods following information acquisition (van Knippenberg & Dijksterhuis, 1996). Further, goal-relevant information is actively integrated during UT, therefore stereotype activation during UT could lead to more biased impressions compared to conditions where negligible thinking occurs. The aim of this thesis was to compare how the active processes of UT and the passive process of being merely distracted are differentially impacted by stereotype activation. In the present experiment, participants were presented with information about a hypothetical person, 'Person 1,' some of which implied traits that are consistent and inconsistent with the stereotype of African-American men. Then, they either received an evaluation goal (UT) or no goal (MD) prior to being distracted for 3min. During distraction, some participants were primed with the stereotype of African-American men. Then, accessibility of the stereotype-consistent and inconsistent traits were measured using a lexical decision task (LDT). Lastly, participants provided their impressions' of 'Person 1' on various trait dimensions. Results supported the hypotheses that among participants in UT conditions, stereotype activation inhibited the accessibility of stereotype-inconsistent concepts relative to consistent concepts. However, this biased accessibility did not impact impressions of 'Person 1.' Possible limitations, implications, and future directions are discussed.

INTRODUCTION

Numerous studies demonstrate that people make superior, or at least comparable, decisions from complex information following a period of unconscious thought (UT) compared to conscious thought (CT; Dijksterhuis, 2004; Dijksterhuis & Nordgren, 2006; Strick et al., 2011). Unconscious Thought Theory (UTT) refers to the superior judgments resulting from UT relative to CT as the unconscious thought advantage (UTA), and suggests this difference results from the greater processing capacity of UT (Dijksterhuis & Nordgren, 2006). Dijksterhuis and Nordgren (2006) offer several principles in their explication of UTT, one of which involves the difference in how these modes of thought process available information. To elaborate, UTT suggests CT is well suited to make simple rule-based decisions, but struggles in making decisions from more complex information due to its limited capacity. Therefore, CT processes information in a top-down or schematic fashion (e.g., using stereotypes), focusing on only a subset of relevant information when forming evaluations. As a result, CT is prone to form evaluations that are biased by stereotypes and other schemas. Alternatively, UTT claims that UT has a vast processing capacity, and therefore employs bottom-up aschematic processes, using most or all relevant information when forming evaluations. Thus, UT should form evaluations that exhibit little to no bias.

Thus far, there is limited research investigating and supporting the notion that UT forms evaluations that are less biased by schemas (specifically, stereotypes) relative to CT. Further, the research supporting that UT reduces stereotype use have consciously activated relevant stereotypes before participants receive information on which to later

base their evaluations. However, stereotypes can be activated unconsciously, outside of conscious awareness (Devine, 1989), and at different stages of an evaluation task (Van Knippenberg & Dijksterhuis, 1996). In this thesis, I build the case that UT might produce evaluations biased by stereotypes if the stereotypes are activated unconsciously *during* a period of UT. This line of reasoning challenges the top-down versus bottom-up principle of UTT, and may reveal necessary refinements to that theory.

Unconscious Thought Theory

According to UTT, individuals can think about information either consciously or unconsciously to arrive at evaluations and decisions. Dijksterhuis (2004) defined conscious thought (CT) as "the cognitive and/or affective task-relevant processes one is consciously aware of while attending to a task" (p.586). Alternatively, he defined unconscious thought (UT) as "cognitive and/or affective task-relevant processes that take place outside conscious awareness" (p.586). Thus, the mode of thinking is distinguished simply by whether or not an individual is aware of the actual task-relevant information-processing they are using to arrive at an evaluation. Beyond this definitional difference, UTT suggests that the two modes of thought have different qualities. Critically, UTT forwards a capacity principle, suggesting that UT has a vast processing capacity, whereas CT has a limited processing capacity. Accordingly, individuals should form sound, information-based, evaluations from CT when information is simple (perhaps even relative to UT). However, as the complexity of the information grows, the capacity of CT will eventually become overwhelmed, resulting in suboptimal judgments based off of a

subset of the relevant information. Alternatively, because UT has a vast processing capacity, the evaluations formed through this thinking modality should not be affected by the complexity of the information (at least, to the same degree), leading to better evaluations from complex information than CT. Authors commonly refer to the better information-based evaluations formed via UT, relative to CT, as the unconscious thought advantage (UTA).

Although UTT primarily focuses on the distinction between UT and CT, it also suggests that UT is a real and active thinking modality. Thus, many experiments investigating UTT also include control conditions in which participants either have no time to think about, or no goal to process, the information after receiving it. Commonly, such experiments demonstrate that participants in UT conditions form better evaluations relative to participants in these control conditions as well. This effect of unconscious thought relative to "minimal thought" is sometimes referred to as the Unconscious thought Effect (UTE), and is useful in demonstrating the operation of UT even when its outputs are comparable to CT.

Classic UTT Paradigm

In the classic UTT paradigm, researchers present participants attribute information on four possible choices (i.e., cars, roommates, apartments, or laptops), and ask them to form an impression of the objects. Then, participants are instructed either to think consciously about the objects, or are actively distracted from consciously thinking about the objects for some time, before choosing the best object and/or rating the extent

to which each is desirable. For example, Dijksterhuis, Bos, Nordgren, and Van Baaren (2006) informed participants that they were going to see attributes describing four possible cars, and gave them the goal to form an overall impression of the cars. They were then presented with 48 attributes, one at a time, in a random order. Each car was described by 12 attributes, some of which were positive (i.e., The Nabusi has good handling), whereas others were negative (i.e., The Hatsdun has poor legroom). The compositions of attributes for the cars were designed so that one car had the most positive attributes (e.g., 8 positive and 4 negative) and was the "best," one had the least positive attributes (e.g., 4 positive and 8 negative) and was the "worst," and two other cars were described by an equal number of positive and negative attributes (e.g., 6 each). After receiving the information, participants were randomly assigned to thinking conditions. Some participants were instructed to spend 4min thinking consciously about the cars on which they will make judgments later. The other participants were also informed that they would make judgments about the cars later, but first completed a distraction task (e.g., solving anagrams) that prevented them from *consciously* thinking about the cars attributes, yet presumably allowed UT to actively integrate the information. After 4min of CT or UT, all participants were asked to choose the best car, and provide preference ratings for each car. Additionally, some studies include control conditions allowing negligible thinking, such as an immediate-decision (ID) condition in which participants make their judgments immediately after information presentation. The most common dependent variable in the classic UTT paradigm is the difference scores between participants' preference ratings of the worst and best objects (preference DV).

Alternatively, some experiments ask participants to choose the best object (choice DV) instead of, or in addition too, the preference DV. Commonly, such experiments demonstrate that UT forms better ratings about, and choices for, the objects relative to CT (a UTA) and control conditions (e.g., ID; a UTE).

A straightforward comparison of outputs from CT and UT does not necessarily reveal that UT is an active process. Indeed, its possible that the UTA is the result of passive processes, where distraction disrupts non-productive thought process, and gives the participants a "fresh look" at the information. To address this issue, Bos, Dijksterhuis and Van Baaren (2008) devised a method to assess if UT was an active or a passive process. These researchers argued that UT does not process all information all of the time; rather, UT processes information that is relevant to an active goal. In other words, UT is goal dependent, and will not operate during a distraction period unless there is an active goal to process particular information. Following this logic, Bos et al. incorporated a new condition in their experiment in which participants experienced a distraction after receiving information—just like participants in a UT condition—but were informed that they were done with that information and will proceed to different tasks for the rest of the experimental session. Thus, participants in these "mere distraction" (MD) conditions held no goal to process the earlier information during the distraction task, whereas participants in the UT conditions did. Bos et al. showed that a period of UT did lead to superior judgments compared to a period of mere distraction (i.e., a UTE), consistent with UT being an active and goal-dependent process.

Furthermore, there is evidence that UT actively integrates and organizes information in memory. In their second experiment, Bos et al., (2008) presented participants with 18 sentences describing 'Jeroen' and asked them to form an overall impression of him. Some of the sentences suggested that Jeroen was athletic, others that he was politically left-wing, and others that he was intelligent. Following information presentation, participants were randomly assigned to either a period of UT or MD. After a distraction task, participants were asked to write down as much as they could recall about Jeroen. The DV of interest was the clustering of recall around the three traits implied by the descriptions. Researchers determine greater clustering of recall the higher the likelihood that a behavior describing one trait is recalled after a behavior describing the same trait. Higher clustering suggests greater organization of information in memory. The results showed that a period of UT lead to higher clustering of information compared to a period of MD, suggesting that UT better integrated and organized information in memory.

Despite numerous findings like the above, there remains controversy in the UTT literature with many experiments failing to find evidence of a UTA. Further, a recent well-powered replication experiment and meta-analysis found no evidence of superior choices following a period of UT compared to CT (i.e., no evidence of UTA for the choice DV; Nieuwenstein et al., 2015). However, a different meta-analysis utilizing both choice and preference DVs found that UT results in significantly more accurate evaluations of options compared to CT (UTA), as well as ID and MD (UTE); however effect sizes are small (Strick et al., 2011). It is possible that inconsistent results are due to

unknown moderators of the UTA and UTE, and well-powered experiments should continue to identify these possible moderators (Garrison & Handley, 2017).

Qualities of Conscious and Unconscious Thought

Two principles of UTT laid out by Dijksterhuis and Nordgren (2006) are of particular importance for understanding how CT and UT are susceptible to stereotypes. First, the capacity principle states that CT is restricted by its limited processing capacity. Thus, when encountering complex problems, CT is forced to focus on only a subset of the task-relevant information. Alternatively, UT has a much greater processing capacity, and therefore does not have to rely on a subset of the information when solving complex problems. Second, because of these capacity differences, Dijksterhuis and Nordgren (2006) propose that CT uses a top-down schematic process, whereas UT uses a bottom-up aschematic process. Thus, as the capacity of CT becomes overwhelmed, it increasingly relies on stereotypes to simplify the information. This results in evaluations that are biased by stereotypes. Alternatively, UT is not limited by its capacity, and therefore should be able to process information in a more holistic bottom-up process that does not utilize stereotypes, but allows processing of all information equally. This results in evaluations that should be free of bias from stereotypes.

Bos and Dijksterhuis (2011) tested this prediction in two experiments on impression formation. In their first experiment, participants were initially informed that they would receive information about a Surinamese man, Mr. Jesserun, and that they

should form an impression of him. Thus, all participants presumably had their stereotype of the Surinamese activated from the outset of the experiment. Then, participants were presented with 24 behavioral descriptions of Mr. Jesserun, some of which implied stereotype-congruent traits, others stereotype-incongruent traits, and still others that implied traits that were irrelevant to stereotypes of the Surinamese. Following information presentation, participants were randomly assigned to either a 7min period of CT or UT. After this 7min period, all participants provided their impressions of Mr. Jesserun on trait dimensions, some of which were congruent, others incongruent to the stereotype implied by the stimulus materials, and two other stereotype-irrelevant traits that were not implied by the materials. Lastly, participants were given 2min to recall as much information as they could about Mr. Jesserun.

Given participants received both stereotype congruent and incongruent trait-information about Mr. Jesserun, impressions formed without bias from stereotypes would result in high ratings on both types of traits relative to the irrelevant traits. Therefore, Bos and Dijksterhuis (2011) subtracted participants' averaged ratings of the irrelevant trait dimensions from the averaged ratings on the congruent trait dimensions in order to score how high participants rated Mr. Jesserun on the congruent trait dimensions. The same process was repeated to get a score of the incongruent trait dimensions. There was no difference between thought conditions on the ratings of stereotype-congruent trait dimensions. However, participants in the UT condition rated Mr. Jesserun significantly higher on the stereotype-incongruent trait dimensions than did those in the CT condition. This implies that participants in the CT condition were biased by stereotypes when

forming impressions of Mr. Jesserun, and this inhibited their processing of his stereotype-incongruent behavior. Likewise, participants in the UT condition were either not biased or less biased, and used the inconsistent as much as the consistent information when forming their impressions. Additionally, when looking at participants' free recall, there was no difference between thought conditions in the recall of stereotype-congruent information, but those in the UT condition recalled significantly more stereotype-incongruent information. Again, this inhibition of stereotype-incongruent information for those in the CT condition would suggest that they were biased by stereotypes when forming an impression. Furthermore, given UT lead to better recall of incongruent traits suggests that UT either prevented, or greatly reduced, the biased use of stereotypes.

The second experiment from Bos and Dijksterhuis (2011) was very similar to the first, except that they measured the accessibility of stereotype congruent and incongruent traits using a lexical decision task (LDT) after a period of CT or UT. In a LDT, participants are shown letter strings and must respond as quickly as possible indicating whether each string is a word or not a word. The logic behind this task is that people will respond more quickly to concepts that are highly accessible versus less accessible. There was no difference in response times between the congruent and incongruent traits for participants in the UT condition. However, participants in the CT condition responded significantly slower to the incongruent traits compared to the congruent traits. This decreased accessibility of the stereotype-incongruent traits—a phenomenon referred to as lateral inhibition—suggests that CT was biased by the stereotypes. As in their first experiment, UT either prevented, or greatly reduced, the biased use of stereotypes.

It is informative to look at the results from Bos and Dijksterhuis (2011) through the process of stereotype activation and application. When participants in Bos and Dijksterhuis's experiments were first told that Mr. Jesserun was a Surinamese man, they were engaged in no other tasks and were therefore under low cognitive load. According to Gilbert and Hixon (1991), people easily activate stereotypes when they experience low cognitive load and encounter individuals from a particular group (e.g., a Surinamese man). Thus, participants' stereotype regarding Surinamese men was presumably activated from the initial information in Bos and Dijksterhuis's experiments. However, although the stereotype was likely activated, that does not guarantee the stereotype should bias later evaluations of the man. Rather, individuals use or apply activated stereotypes to form evaluations when they experience high cognitive load (i.e., processing capacity is limited) versus low cognitive load (i.e., processing capacity is available; Gilbert & Hixon, 1991). According to UTT, CT has a limited processing capacity, and it is therefore likely that participants who thought consciously about complex information in Bos and Dijksterhuis's experiments experienced high cognitive load. As a result, participants in CT conditions applied the active stereotype of Surinamese men to form evaluations of Mr. Jesserun, which exhibited bias. However, given UT has a vast processing capacity according to UTT, it is likely that participants who thought unconsciously about complex information in Bos and Dijksterhuis's experiments experienced low cognitive load (albeit, unconscious cognitive load in this case). As a result, participants in UT conditions did not apply the active stereotype of Surinamese men to form evaluations of Mr. Jesserun, resulting in less bias.

It is important to note that the effects of stereotypes on memory depend on a variety of factors, such as motivation to not use stereotypes (Scarabis & Florack, 2008), self-enhancement goals (Sinclair, 1999), cognitive load during information acquisition (Macrae, Hewstone, & Griffiths, 1993), individual differences in need for closure (Dijksterhuis, Van Knippenberg, Kruglanski, & Schaper, 1996), and whether the stereotype itself is activated in an overt or subtle way (Heider et al., 2007). Indeed, plenty of research suggests that stereotype activation leads to *improved* memory for stereotype-inconsistent information relative to consistent information (Rojhan & Petigrew, 1992; Sherman, Stroessner, Conrey & Azam, 2005), but other research suggests the opposite (Fyock & Stangor, 1994; Dijksterhuis & Van Knippenberg, 1995, Heider et al., 2007). For example, Heider et al. found that activating stereotypes in overt ways (e.g., having participants read a paragraph that repeatedly mentioned how stereotypical "Tyrone's" behavior is of an African-American man) increased memory for subsequently presented stereotype-inconsistent information relative to consistent information. However, when the stereotype was activated in a subtler manner (e.g., simply by providing the stereotypical name "Tyrone"), they found a memory advantage for stereotype-consistent information. The method of stereotype activation employed by Bos and Dijksterhuis (2011) is consistent with this subtler method of stereotype activation, as the researchers merely noted that the target individual was a Surinamese man.

Similarly, Van Knippenberg and Dijksterhuis (1996) found that stereotype activation impaired memory for inconsistent information only if the inconsistent information loaded onto a different trait dimension than the stereotype-consistent

information. For example, activation of the professor stereotype inhibited recall of aggressive behaviors (i.e., stereotype-inconsistent information that does not load onto the same trait dimension as stereotype-consistent trait intelligent), but did not impair recall of unintelligent behaviors (i.e., stereotype-inconsistent information that loaded directly onto the stereotype-consistent trait dimension of intelligence). In the experiments by Bos and Dijksterhuis (2011) the stereotype-consistent information (i.e., sociable and lazy) and inconsistent information (i.e., intelligent and nonmusical) loaded unto different trait dimensions. Thus, their findings that CT leads to worse memory for stereotype-inconsistent information, more stereotypical impressions, and inhibits the accessibility of stereotype-inconsistent traits suggest that CT was biased by stereotypes.

Stereotypes and UT: More to the Story?

Consistent with UTT, Bos and Dijksterhuis (2011) concluded that UT processes either prevent stereotype bias or greatly reduce the use of stereotypes. However, this conclusion could be an overgeneralization; perhaps UT only reduces the use of stereotypes in particular contexts. Specifically, research suggests that providing an expectancy (e.g., a stereotype) prior to encoding increases memory for expectancy consistent and inconsistent information compared to irrelevant information (Srull, Lichtenstein & Rothbart, 1985). Srull et al. proposed that this occurs because participants compare new information against the expectancy during encoding, resulting in more in-depth processing of information relevant to the expectancy (i.e., consistent and

inconsistent information) compared to information that is irrelevant to the expectancy. Thus, the consistent and inconsistent information should have been highly salient for all participants in the Bos and Dijksterhuis experiments. However, once the deliberation phase began, the limited capacity of CT likely lead participants to focus on the information that was most easily accessible (i.e., the consistent information). That is, participants in the CT conditions likely relied on their stereotypes to simplify information processing given the overwhelming volume of the information (Macrae, Stangor & Milne, 1994). Alternatively, given UT is not limited by its capacity, it could have focused on both the consistent and inconsistent information, explaining why both types of information influenced participants ratings and memory equally. In reality, Bos and Dijksterhuis showed that UT is able to better process stereotype consistent and inconsistent information, compared to CT, if that information is highlighted prior to, and during, information encoding. Thus, although UTT broadly suggests that UT engages in bottom-up processing and is not biased by stereotypes, the evidence to date only supports that conclusion in contexts where stereotypes are explicitly active *prior* to information acquisition and periods of UT. It is therefore important to fully test the bottom-up principle proposed by Dijksterhuis and Nordgren (2006), and assess whether stereotypes activated implicitly, and during *thinking periods*, bias the evaluations of UT. Further, to fully test if stereotype activation can affect UT, it would be useful to employ a control condition like MD. Comparing UT to a condition like MD, which involves a negligible amount of thought, allows for a pure investigation of whether stereotypes can bias the active-integration processes that occur during UT. Beyond suggesting more thorough

investigations of the bottom-up principle, however, there is reason to suggest that stereotypes can at times bias evaluations from UT.

First, there is evidence suggesting that UT organizes information (e.g., Bos et al., 2008), and that experts benefit more from UT then novices (Dijksterhuis, Bos, Van der Leij & Van Baaren, (2009), both of which suggest that UT uses schemas to some extent. Second, there is evidence that *conscious* activation of stereotypes *after* encoding can affect participants' memory. For instance, Snyder and Uranowitz (1978) had participants read a narrative about a woman, then provided participants information suggesting she was either heterosexual, homosexual, or provided no information about her sexual orientation. They found that activation of stereotypes related to sexual orientation affected participants' memory of the target woman. Specifically, participants who were lead to believe that she was a lesbian falsely recognized more stereotypically lesbian behaviors compared to participants who were lead to believe that she was heterosexual or who did not receive any information about her sexual orientation. This suggests that conscious stereotype activation after encoding can bias memory in favor of stereotype-consistent information. Third, stereotypes activated *unconsciously* after encoding and before recall can also affect participants' memory. For example, Van Knippenberg and Dijksterhuis (1996) gave participants behavioral descriptions that were consistent and inconsistent with the stereotype of a soccer hooligan. Then, half of participants had the stereotype of soccer hooligans activated during a seemingly unrelated distractor task, before having them recall the earlier behavioral descriptions. Van Knippenberg and Dijksterhuis found that activation of the stereotype impaired memory for stereotype

inconsistent information relative to consistent information, whereas there was no difference between consistent and inconsistent recall for those who did not have the stereotype activated. This suggests that stereotype activation after encoding, and during a distraction task, can bias memory to the detriment of stereotype inconsistent information.

Of note, the procedures utilized by Van Knippenberg and Dijksterhuis (1996) closely resemble those of an UT condition in a person-impression experiment. As in typical UTT experiments, participants in Van Knippenberg and Dijksterhuis's research had the goal to form an overall impression of the described individual. Further, after participants received the information, they were distracted for 5min before reporting their impressions of that person. However, during distraction period, some of the participants had a stereotype non-consciously activated, whereas others did not. Given the distraction period in this experiment is highly similar to distraction periods used in UT conditions, it is reasonable to suggest that all participants in this experiment may have engaged in unconscious thought during the distraction period. Still, participants who were primed with the stereotype, relative to those who were not, remembered less stereotype-inconsistent information, suggesting they were biased by the stereotype. Thus, drawing a parallel between this experiment and UT conditions in other experiments, it seems stereotypes may indeed bias UT. Of course, Van Knippenberg and Dijksterhuis's experiment was not explicitly designed to test UT, and there are likely slight methodological differences that make it unclear whether UT was operating in their experiment and was biased by stereotypes. Further, their experiment did not include a control condition in which participants had no goal to process the behavioral information,

making it unclear to what extent participants in these UT-like conditions formed biased impressions. It is possible, for example, that the biased impressions formed by participants in these UT-like conditions are less biased—although still biased—relative to participants in control conditions. Thus, careful research testing UT's susceptibility to bias from stereotype activation is warranted, even though Van Knippenberg and Dijksterhuis's results are suggestive. Indeed, slight modifications to Van Knippenberg and Dijksterhuis's (1996) experiments can directly test a possible stereotype bias from UT. In particular, including definitive UT and MD conditions within such an experiment would allow for a test of stereotype bias on UT relative to minimal thought.

If Dijksterhuis and Nordgren (2006) are correct that UT is globally not biased by stereotypes (the impermeable hypothesis), then stereotype activation during UT should not affect participants' evaluations and stereotype consistent and inconsistent information should be weighted equally (cf., Bos & Dijksterhuis, 2011). Therefore, the influence of a stereotype on UT should lead to less pronounced inhibition of stereotype-inconsistent information, and less stereotypical judgments, relative to outcomes that result from negligible thinking (i.e., MD). Alternatively, the research of Van Knippenberg and Dijksterhuis's (1996)—and other research that demonstrates stereotype bias resulting from unconscious stereotype activation—highlights the possibility that stereotypes could bias UT (i.e., the permeable hypothesis) when activated during an unconscious-thinking period, and thus inhibit memory for stereotype-inconsistent information and lead to more stereotypical impressions. Further, given UT is an active process that integrates and organizes information into memory (Bos et al., 2008), the influence of a stereotype on

this process should lead to more pronounced inhibition of stereotype-inconsistent information, and more stereotypical judgments, relative to outcomes that result from negligible thinking (i.e., MD). Thus, although stereotypes activated just prior to information encoding do not bias UT (Bos & Dijksterhuis, 2011), there are two opposing and untested predictions about whether stereotypes activated after encoding will bias outputs from UT. Of note, UTT does not specifically offer predictions about stereotype-biased outcomes within MD conditions. However, the general literature on stereotype activation following the presentation of information suggests that participants in a context such as a MD condition would manifest a stereotype bias (less memory for stereotype-incongruent information, and greater stereotypical impressions among people in whom a relevant stereotype is activated, versus not). Thus, both of the above viewpoints suggest stereotypes should bias outcomes in MD conditions. However, relative to biased outcomes in MD conditions, the impermeable hypothesis (UTT) predicts less biased outcomes from UT, whereas the permeable hypothesis predicts more biased outcomes from UT. My proposed experiment is designed to test these opposing predictions.

Experimental Overview

The current experiment was a 2 (thought: UT vs. MD) x 2 (word search puzzles: primed vs. control) x 3 (stereotype-relevancy: consistent vs. inconsistent vs. neutral) mixed-design with the last factor varied within-subjects. A sample of undergraduate participants first learned that their task was to form an overall impression of a person, and then viewed a series of descriptions about ‘Person 1.’ Some of the descriptions implied

traits that are consistent with the stereotype of African-American men, others implied traits that are inconsistent with the stereotype, and still others were neutral to the stereotype of African-American men. After reading the descriptions, participants were randomly assigned to either a period of UT or MD for 3min, during which time they solved word search puzzles (WSPs) as a distraction task. Further, participants were randomly assigned to complete a WSP task that included names that are stereotypical of African-American men (primed-WSPs), or a WSP task with names that are neutral to the stereotype of African-American men (control-WSP). In this way, some participants should have had the African-American stereotype activated by the WSP task, whereas other participants should not. Following the WSP task, all participants completed an LDT. Some of the target words on the LDT were the stereotype-consistent and inconsistent traits implied by the earlier descriptions and others were unrelated to the behavioral descriptions of 'Person 1.' Lastly, participants were asked to rate the extent to which they felt 'Person 1' possessed the specific traits implied by the descriptions.

Hypotheses

The permeable hypothesis (UTT) suggests that individuals who think unconsciously, versus think negligibly, about information will come to judgments and responses that are relatively unaffected by stereotypes. Although this is likely true in some cases (e.g., when stereotypes are activated before individuals receive information, see Bos & Dijksterhuis, 2011), I outlined rationale to suggest that UT may at times be biased by stereotypes, such as when stereotypes are activated subtly during a period of

UT. In this case, the active, and arguably biased, processing of information among people who think unconsciously, versus think negligibly, about information will result in judgments and responses that are particularly affected by stereotypes. Thus, UTT predicts UT is impermeable to stereotypes and will not produce outcomes that are biased by stereotypes within the current experiment, whereas the permeable hypothesis (that still assumes the operation of unconscious thinking process) suggests that UT will produce outcomes that are biased by stereotypes. Hypotheses 1 and 2 (specified below) are consistent with both the impermeable and the permeable hypotheses presented in this thesis. However, the impermeable hypothesis (UTT) and the permeable hypothesis make divergent predictions as well, and these are elaborated in Hypotheses 3a and 4a (impermeable predictions), and Hypotheses 3b and 4b (permeable predictions).

Overall, both the impermeable hypothesis (UTT) and the permeable hypothesis predict an interaction between thought, prime, and measurements for stereotype-consistent and inconsistent information (a within-participants factor). However, the specific pattern of this interaction differs in critical ways according to the two alternative perspectives. Of note, greater differences between measurements for stereotype-consistent and inconsistent information suggests greater bias from a stereotype in the below hypotheses.

Shared Predictions

Hypothesis 1: Participants in the MD conditions who received the primed-WSPs should respond significantly faster to the stereotype-consistent targets than the stereotype-inconsistent targets, relative to those who received the control-WSPs. Such a result would

simply demonstrate the effects of stereotype activation in the control conditions, and is consistent with previous research showing a consistency bias following stereotype activation (Heider et al., 2007; Van Knippenberg & Dijksterhuis, 1996). Similarly, participants in the MD conditions who received the primed-WSPs should rate 'Person 1' higher on the stereotype-consistent trait dimensions relative to the stereotype-inconsistent trait dimensions, compared to participants who received the control-WSPs.

Hypothesis 2: Among participants who received the control-WSPs, those in the UT condition should respond significantly faster to both the stereotype-consistent and inconsistent targets on the LDT, relative to participants in the MD condition. According to UTT, participants in the UT conditions should process and organize the information contained in the descriptions, whereas participants in the MD conditions should not, resulting in greater activation of all traits implied by those descriptions. Similarly, among participants who received the control-WSPs, those in the UT condition should rate 'Person 1' higher on both the stereotype-consistent and inconsistent trait dimensions, relative to participants in the MD condition.

Divergent Predictions

Hypothesis 3a: According to the impermeable hypothesis (UTT), the difference between participants' response times to stereotype-consistent and inconsistent traits on the LDT should be negligible if they engaged in UT. Furthermore, these differences should be comparable for participants receiving a stereotype prime and no prime during a period of UT. This prediction accords with the results of Bos and Dijksterhuis (2011) who found

that among participants in the UT condition, stereotype activation prior to information acquisition did not bias the accessibility of stereotype-consistent and inconsistent traits. Similarly, the impermeable hypothesis (UTT) would predict that the difference between participants' ratings on the stereotype-consistent and inconsistent trait dimensions should be negligible if they engaged in UT. Furthermore, these differences should be comparable for participants receiving the primed-WSPs and control-WSPs.

Hypothesis 3b: In contrast to UTT, if stereotypes do bias outcomes among participants in UT conditions, such participants would respond significantly faster to the stereotype-consistent trait targets relative to the stereotype-inconsistent trait targets on the LDT if they received the primed-WSPs compared to the control-WSPs. Similarly, among participants in the UT conditions, participants who received the primed-WSPs should rate 'Person 1' higher on the stereotype-consistent trait dimensions and lower on the stereotype-inconsistent trait dimensions, relative to participants who received the control-WSPs. This permeable hypothesis opposes Hypothesis 3a.

Hypothesis 4a: According to the impermeable hypothesis (UTT), among participants who received the primed-WSPs, participants in the MD condition should respond significantly faster to the stereotype-consistent targets relative to the inconsistent targets on the LDT, compared to participants in the UT condition. This prediction is consistent with Bos and Dijksterhuis (2011) who found that among participants in the UT condition, stereotype activation did not bias the accessibility of stereotype-consistent and inconsistent information. Similarly, the impermeable hypothesis (UTT) would predict that among participants who received the primed-WSPs, participants in the MD condition should rate

‘Person 1’ higher on the stereotype-consistent and lower on the stereotype-inconsistent trait dimensions, relative to participants in the UT condition.

Hypothesis 4b: Alternatively, if stereotypes can permeate UT, then participants in UT conditions will be more biased by stereotype activation compared to participants in MD conditions, for whom negligible thought should occur. Specifically, among participants who received the primed-WSPs, participants in the UT condition will respond significantly faster to the stereotype-consistent targets relative to the inconsistent targets, compared to participants in the MD condition. This is expected because UT is an active process, and stereotype activation should bias the accessibility of information in this active process to a greater extent than it would in a passive process like MD. Similarly, among participants who received the primed-WSPs, participants in the UT condition should rate ‘Person 1’ higher on the stereotype-consistent trait dimensions relative to inconsistent trait dimensions, compared to participants in the MD condition. This hypothesis opposes Hypothesis 4a.

METHOD

Participants and Design

Two hundred nineteen undergraduates from Montana State University were recruited to participate in this study. Participants received partial course credit as compensation for their participation. Due to technical issues with the software running the experiment, 28 participants were not able to finish the experiment, leaving a final sample of 191 individuals. Participants were randomly assigned to the conditions of a 2 (Thought: UT vs. MD) x 2 (WSPs: primed vs. neutral) x 3 (stereotype-relevancy: consistent vs. inconsistent vs. neutral) mixed-design with the last factor varied within-participants. The 28 lost participants where comparably distributed across the four conditions (final number of participants in each condition; UT-primed = 44, MD-primed = 52, UT-neutral = 47, MD-neutral = 48).

Procedure

Participants received consent forms, and then completed the experiment on individual computers. Participants were first told that the experiment was investigating how people form impressions of others and that their goal was to form an overall impression of a person based on various descriptions. Then, participants were presented with 32 descriptions of ‘Person 1.’ Twelve of these descriptions were consistent with the stereotype of African-American men, 12 were inconsistent with the stereotype, and eight

were irrelevant to the stereotype of African-American men. These descriptions were presented one at a time in random order.

After the presentation of all 32 descriptions, participants were randomly assigned to either hold an impression-formation goal (UT), or no goal (MD), then engaged in a distraction task in which they solved WSPs for 3min. Further, participants were randomly assigned to complete WSPs in which some hidden words were stereotypical names of African-American men (primed-WSPs) or WSPs in which all of the hidden names were neutral to the stereotype of African-American men (control-WSPs). Following the WSPs task, all participants completed an LDT in which some of the target words were the stereotype-consistent and inconsistent traits implied by the earlier descriptions, whereas others were unrelated to the descriptions of 'Person 1.' Then, participants were asked to rate the extent to which they felt 'Person 1' possessed the specific traits implied by the initial descriptions (i.e., traits consistent and inconsistent with the stereotype). Lastly, participants were probed for suspicion of the priming procedure before being debriefed and excused.

Materials

Description Stimuli

'Person 1' was portrayed by 32 descriptions randomly presented one at a time for 4s each. Twelve of the descriptions implied traits that are consistent with the stereotype of African-American men (i.e., athletic, musical, criminal) with each trait represented by 4 descriptions (e.g., "Person 1 plays the bass in a reggae band"). Twelve of the descriptions implied traits that are inconsistent with the stereotype of African-American

men (i.e., intelligent, hardworking, shy) with each trait represented by 4 descriptions (e.g., "Person 1 enjoys solving math proofs"). Lastly, eight of the descriptions conveyed information that was neutral to the stereotype of African-American men (e.g., "Person 1 has 2 siblings").

These stimuli were pilot tested with a sample of 120 native English-speaking Americans on Amazon's Mechanical Turk (MTurk). All participants were presented the descriptions one at a time in random order. After viewing each description, and depending on random assignment, participants were asked to either: "Indicate how consistent this description is with the following traits, (i.e., athletic, musical, criminal, intelligent, hardworking, shy)" (-3 = *very inconsistent* to 3 = *very consistent*), "indicate the extent to which the description is consistent with the cultural stereotype of African-American men" (-3 = *very inconsistent* to 3 = *very consistent*), or "list four traits that came to mind after reading the description."

Based on the pilot study, the stereotype-consistent (stereotype-inconsistent) descriptions were chosen based on receiving a rating that is significantly more (less) stereotype-consistent than the stereotype neutral descriptions. A description was considered stereotype neutral if it's rating did not differ significantly from zero (the midpoint of the scale). Additionally, each description-trait combination was tested against zero (the mid point of the scale). Descriptions were considered to imply a trait if they were rated as significantly greater than zero on that trait. Some descriptions received ratings that were significantly greater than zero on more than one trait. Those descriptions were included if ratings on the intended trait were significantly higher than the secondary

trait and the traits were congruent with each other. For example, the description "Person 1 has outrun the cops before" could imply both criminal and athletic, but would be considered a criminal description as long as the rating on criminal was significantly higher than the rating on athletic. Finally, for a description to imply a given trait, the majority of participants must have listed it (or a synonym) as an implied trait after reading the description.

Independent Variables

Thought Condition

Following presentation of the descriptions for 'Person 1,' participants were asked to rate the extent to which they found the descriptions useful in forming an impression before being randomly assigned to a UT or MD condition. Next, participants in the MD condition were told that the impression-formation portion of the experiment is now over, and then were directed to complete a series of WSPs for 3min. For these participants, the combination of answering a question about the descriptions and being told that they are done with the impression-formation stage should have been sufficient to disabuse them of the goal to process the information further (see Bos et al., 2008). Therefore, these participants should not have engaged in UT during the WSP task. Alternatively, participants in the UT condition were informed that they would have to answer further questions about 'Person 1,' but that they would first complete a series of WSPs for 3min. Therefore, these participants should engage in UT during the WSP task because they should have retained the goal to form an overall impression of 'Person 1.' In both the UT

and MD conditions, the WSPs served as a distractor task to prevent participants from consciously thinking about the descriptions of 'Person 1.'

WSPs

From the outset of the distractor task, participants were randomly assigned to either a primed-WSPs condition or a control-WSPs condition. All of the hidden words were male names. In the primed-WSPs conditions, some of the names were stereotypical names of African-American men (e.g., Jamal), thus non-consciously priming the stereotype of African-American men. Other researchers had validated these names (Bertrand & Mullainathan, 2004; Greenwald, McGhee, & Schwartz, 1998; Rudman, Ashmore, & Gary, 2001). In contrast, in the control-WSPs condition all the targets on the WSPs were names that are neutral to the stereotype of African-American men (e.g., Brian). Some of these stereotype-neutral names were taken from the existing literature (Kasof & Steinberg, 1993), whereas others were gathered from 120 MTurk workers each listing 10 names that they thought were equally likely to belong to European and African-American men.

The WSPs appeared one at a time in the center of the computer screen. The participants' task was to identify and type the name into a box locate below the WSPs. To avoid participants getting stuck on an individual WSP, the computer automatically advanced the participants to the next WSP after 15s. Because successful identification of the words is crucial for the priming manipulation, all target words were vertically or horizontally displayed in the WSP to facilitate identification. After 15s had passed or the participant had entered in a word, the hidden name was highlighted on the screen for 1s,

before the next WSPs was presented. To ensure that participants in the primed-WSP conditions receive sufficient exposure to the primes, the primed-WSPs were front-loaded. Specifically, 12 of the first 20 WSPs were primes, however the order of these 20 WSPs was randomized. After the initial 20 WSPs, the ratio of primed-to-neutral WSPs became 1 to 1. This prime to neutral ratio is an adaptation of the ratio used by Welsh and Ordonez (2014).

Dependent Variables

LDT

Participants completed an LDT following the WSP task. Participants were told they are going to perform a word recognition task and their job is to respond as quickly as possible indicating whether a letter string on their computer screen is a word or a non-word by pressing one of two keyboard buttons. Participants pressed the space bar to start each trial, then an orienting "+" appeared in the center of the screen for 500ms, followed by the target letter string which disappeared once the response was made or 2000ms had elapsed.

The LDT consisted of 82 trials across two blocks. The first block included 10 practice trials of 5 non-words and 5 words that are unrelated to the stereotype of African-American men. The second block consisted of 72 trials, 36 of which were non-words and 36 actual English words. Of the English words, 12 were stereotype-consistent, with three being the actual trait implied by the earlier descriptions (e.g., criminal) and nine being related words (e.g., felon) as identified by a large online database described by De Deyne, Navarro, and Storms (2013). Similarly, 12 of the words were stereotype-

inconsistent, with three being the actual trait implied by the earlier descriptions (e.g., intelligent) and nine being related words (e.g., smart). It should be noted that Bos and Dijksterhuis (2011) only had two targets on the LDT for each trait; I decided to use four targets per trait here in order to increases the sensitivity of the measure. Finally, 12 of the English words were stereotype neutral. The decision to have an equal number of stereotype-consistent, inconsistent, and neutral target words is based on the LDT procedure done by Dijksterhuis and Van Knippenberg (1996).

Using the English Lexicon Project (Balota et al., 2007), the stereotype-consistent, inconsistent, and neutral targets were matched, as best as possible, on word length, contextual diversity (a measure of for the number of contexts in which a word appears, specifically LgSUBTLCD), mean reaction times, and orthographic neighborhood (i.e., the number of words that could be created by changing one letter in the word; Veraga-Martines & Swaab, 2012). Separate one-way ANOVAs were run comparing the consistent, inconsistent, and neutral targets on each of the four above parameters, none of which showed significant group differences (all Fs < 1.25).

Reaction times to the 12 stereotype-consistent targets were averaged to create a consistent-target score. This provided a measurement for the accessibility of the stereotype-consistent traits, and there was good reliability among these traits (Cronbach's alpha = .708). Similarly, reaction times to the 12 stereotype-inconsistent targets were averaged to create an inconsistent-target score. Again, this provided a measurement for the accessibility of the stereotype-inconsistent traits, and reliability was good (Cronbach's alpha = .788).

Trait ratings

Lastly, participants were asked to report their impression of 'Person 1' by rating the extent to which they felt he possessed the three stereotype-consistent traits (athletic, musical, criminal; e.g., to what extent do you feel Person 1 is athletic?) and the three stereotype-inconsistent traits (intelligent, hardworking, shy; e.g., to what extent do you feel Person 1 is intelligent?) implied by the earlier descriptions of him. Participants provided responses to each trait on 7- point scales ranging from 1 *(not at all*) to 7 (*very much so*). Independently, the ratings for the 3 consistent traits, and the 3 inconsistent traits, were averaged to create a measure of stereotype application. The reliability among the 3 consistent and inconsistent traits was quite poor ($\alpha = .292$; $\alpha = .513$, respectively). The primary analyses still utilize these two composite scores; however, their poor reliability was one justification for conducting additional exploratory analyses examining the traits individually.

RESULTS

Awareness of Priming

A funnel debrief was included at the end of the experiment to probe for awareness of the priming procedure and its link to impressions of 'Person 1.' Fourteen participants indicated awareness of the theme of stereotypical names in the WSPs and hypothesized that the purpose of this theme was to bias their impressions (bias-aware group)[1]. An additional 9 participants indicated awareness of the theme and came close to guessing its purpose (marginally bias-aware group). Lastly, 9 participants indicated awareness of the theme but did not indicate that it was meant to bias their impressions of 'Person 1'(theme-only). All analyses were conducted four times: with all participants included, with the bias-aware group excluded, with the bias-aware and marginally bias-aware groups excluded, and with the bias-aware, marginally bias-aware, and theme-only groups excluded. Because awareness of the prime could easily impact an explicit measure like trait ratings, the most stringent exclusionary criteria were utilized when analyzing the trait ratings data.

However, when analyzing the LDT data, it is justifiable to consider the results of all participants for several reasons. First, none of the participants who noticed the stereotypical theme in the WSPs noted that it might have affected their responding on the LDT; thus, based on their own self-reports, it seems unlikely that participants'

---

[1] Participants in the UT and MD conditions were more or less equally represented in the bias-aware (UT = 5, MD = 9) marginally bias-aware (UT = 3, MD = 6), and theme only groups (UT = 6, MD = 3).

knowledge of this theme led them to strategically alter their responses on the LDT. Additionally, even if participants were aware of the prime and its intended effects on the LDT, it is unlikely that they would have accurately guessed the specific traits tested in this experiment. Third, and perhaps most compelling, Banaji and Hardin (1996) found comparable stereotype-priming effects on a LDT for participants who were aware and unaware of the prime's link to the task. Thus, this evidence suggests that differences in response times between consistent and inconsistent trials for the LDT reflect stereotype activation regardless of whether participants within the current experiment might have realized the stereotype could influence responses within the LDT. Further, removing participants did not change the pattern of results, and changes in statistical significance may well have occurred due to decreased statistical power. Lastly, the effect size observed in the critical three-way interaction barely changed from excluding no one (η^2 = .038) to removing all 32 participants (η^2 = .037) who indicated awareness of the theme. Given the above points, it seems appropriate to consider the results from the analyses including all participants for which I obtained sufficiently valid data. However, when the significance of results changed based on exclusions, I report results from multiple analyses in the footnotes.

Reaction Times

Data Cleaning

As noted earlier, each stereotype-consistent and inconsistent trait was tested with four targets on the LDT. Incorrect responses to a target or reaction times that were 3 or

more SDs from the mean for a given target were excluded from analysis of that target. Participants who responded incorrectly to 2 or more targets of a given trait were excluded from analysis of that *specific* trait (e.g., athletic; see Exploratory Analyses). Further, participant's data were removed from the *overall* LDT analysis if they had an overall incorrect response rate greater than 40% or had 4 or more incorrect responses within the 12 consistent or inconsistent trials.

LDT

Average target reaction-times were subjected to a 2 (Thought: UT vs. MD) x 2 (WSPs: primed vs. control) x 2 (target type: consistent vs. inconsistent) mixed-model analysis of variance (ANOVA), with target type serving as a within-subjects factor. Analysis on the within-subjects effects demonstrated a significant main effect of target type, $F(1, 136) = 14.69$, $p < .001$, $\eta^2 = .097$, observed power = .962, reflecting that reaction times were overall faster for the stereotype-consistent targets ($M = 713.87$, $SD = 110.48$) relative to inconsistent targets ($M = 739.38$, $SD = 124.07$). The analysis revealed no significant main effects of Thought or WSPs, nor any significant 2-way interactions (all Fs < 0.832, ps $> .05$). Importantly, however, the predicted 3-way interaction among Thought, WSP, and target type was significant, $F(1, 136) = 5.33$, $p = .022$, $\eta^2 = .038$, observed power = .630.

To understand the nature of this three-way interaction, and to more specifically test Hypotheses 1 and 3, I conducted 2 separate analyses to look at the 2-way interaction between WSP and target-type at each level of Thought. Among participants in the MD conditions, the 2 (WSPs) x 2 (target-type) mixed-model ANOVA revealed only a

significant main effect of target-type $F(1,74) = 6.06$, $p = .016$, $\eta^2 = .076$, observed power = .68 (all other Fs < .9), showing that on average participants responded faster to the

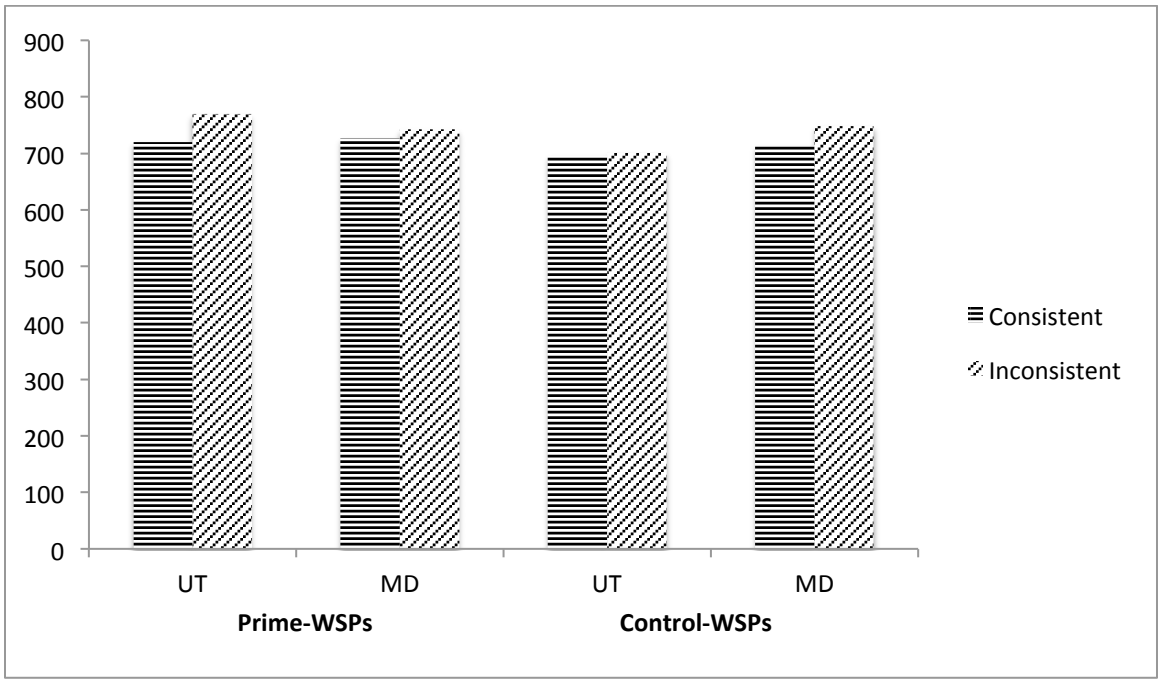

Figure 1. Reaction times (in ms.) to stereotype consistent and inconsistent targets as a function of Thought and WSP conditions.

consistent targets than inconsistent targets. Further, simple effects tests revealed that participants who received the control-WSPs responded significant slower to the inconsistent targets relative to the consistent targets $F(1,74) = 5.96$, $p = .017$, $\eta^2 = .074$, observed power = .67, whereas this difference was not significant among participants who received the primed-WSPs $F(1,74) = 1.12$, $p = .293$, $\eta^2 = .015$, observed power = .181. This is contrary to Hypothesis 1, and suggests that among participants in the MD conditions, those who received the control-WSPs exhibited biased accessibility of the stereotype-relevant information, and participants who received the primed-WSPs did not. This finding among participants in the control-WSP condition is curious because it

suggests the stereotype was more activated in them relative to participants who actually received the stereotype prime. However, because the WSPs x target-type interaction was not significant, the extent of bias observed among participants who received the control-WSPs was not significantly larger than among participants who received the primed-WSPs. This suggests that the priming tasks did not differentially activate the stereotype among participants in the MD conditions. Possible explanations for this counter intuitive finding are discussed later.

Among participants in the UT conditions, the 2 (WSPs) x 2 (target-type) mixed-model ANOVA revealed a significant main effect of target-type $F(1,62) = 9.79$, $p = .003$, $\eta^2 = .136$, observed power = .869, again showing that across conditions, participants responded faster to the consistent targets relative to the inconsistent targets (see table 1). The main effect of WSPs was not significant ($F[1,62] = 2.58$, $p > .05$), but the WSPs x target-type interaction was significant $F(1,62) = 6.27$, $p = .015$, $\eta^2 = .092$, observed power = .693[2]. Simple effects analysis revealed that participants who received the primed-WSPs responded significantly slower to the inconsistent targets relative to the consistent targets $F(1,62) = 14.5$, $p < .001$, $\eta^2 = .19$, observed power = .963, whereas this difference was not significant among participants who received the control-WSPs $F(1,62) = .215$, $p = .644$, $\eta^2 = .003$, observed power = .074. This suggests that stereotype activation during UT biased the accessibility of stereotype-relevant information.

---

[2]Note, this interaction remained significant after removing the bias-aware participants $F(1,58) = 7.04$, $p = .010$, $\eta^2 = .108$, observed power = .742, and additionally, the marginally bias-aware participants Interaction. $F(1,55) = 4.828$, $p = .032$, $\eta^2 = .032$, observed power = .579. However, this interaction was not significant when the bias-aware, marginally bias-aware, and theme only groups were excluded, $F(1,51) = 3.51$, $p = .067$, $\eta^2 = .064$, observed power = .452.

Furthermore, the significant WSPs x target-type interaction implies that the extent of this bias was significantly larger among participants who received the primed relative to the control-WSPs. This finding supports Hypothesis 3B suggesting that UT can be biased by stereotypes, and refutes Hypothesis 3A that suggests UT is not biased by stereotypes.

To compare the impact of priming across thought conditions, a 2 (Thought) x 2 (target-type) mixed-model ANOVA was conducted on participants who received the primed-WSPs. This analysis revealed a main effect of target-type, indicating that across conditions, participants responded faster to the consistent targets relative to the inconsistent targets $F(1,64) = 9.46$, $p = .003$, $\eta^2 = .129$, observed power = .857. Neither the main effect of Thought, nor the Thought x target-type interaction were significant (all Fs < 2.64) Simple effects analysis revealed that participants in the UT condition responded significantly slower to the inconsistent targets relative to the consistent targets $F(1,64) = 9.84$, $p = .003$, $\eta^2 = .113$, observed power = .871, whereas this difference was not significant among participants in the MD conditions, $F(1,64) = 1.2$, $p = .277$, $\eta^2 = .018$, observed power = .191. Overall, these results suggest that stereotype activation did bias the accessibility of information among participants engaged in UT, supporting Hypothesis 3b. However, because the Thought x target-type interaction was not significant, and priming failed to activate the stereotype among participants in the MD conditions, there is not clear evidence that participants in the UT condition were more biased by stereotype activation compared to participants in the MD condition (supporting neither Hypotheses 4a or 4b) .

To test Hypothesis 2, average reaction times to the stereotype-consistent and inconsistent targets were separately analyzed using 2 (Thought) x 2 (WSPs) between-subjects ANOVAs. For the stereotype-consistent targets, the analysis revealed no significant main effects nor an interaction. Further, planned simple effects tests found no difference in reaction times between participants in the UT and MD conditions who received the control-WSPs, $F(1,149) = 1.21$, $p = .274$, observed power = .194. Similarly, for the stereotype-inconsistent targets, the analysis revealed no significant main effects nor an interaction. Further, planned simple effects testing found no difference in the reaction times between participants in the UT and MD conditions who received the control-WSPs, $F(1,156) = .585$, $p = .445$, observed power = .118. Both of these null results contradict Hypothesis 2, which stated that UT should lead to faster reaction times to both consistent and inconsistent targets relative to MD in the absence of priming. Rather, it appears that the presumed active-integration process of UT did not lead to greater accessibility of stereotype-relevant information compared to the passive processes within the MD conditions.

Trait Ratings

As mentioned above, awareness of the prime and its purpose is very problematic for an explicit measure like trait ratings. Thus, I employed stringent exclusion criteria. Specifically, the primary analysis reported excluded all 32 participants who correctly identified the theme of the WSPs, regardless of whether they indicated it was meant to bias them. However, in terms of significance, none of the results changed based on exclusionary criteria. The average ratings on the consistent and inconsistent traits were

subjected to a 2 (Thought: UT vs. MD) x 2 (WSPs: primed vs. control) x 2 (Trait: consistent vs. inconsistent) mixed-model ANOVA, with Trait serving as the within-subjects factor. The analysis demonstrated a significant main effect of thought, $F(1,155) = 6.292$, $p = .013$, $\eta^2 = .039$, observed power = .703, such that participants in the MD conditions rated 'Person 1' higher on the stereotype-consistent ($M = 5.654$, $SD = .904$) and inconsistent ($M = 5.757$, $SD = .869$) trait dimensions relative to participants in the UT conditions ($M = 5.393$, $SD = .809$; $M = 5.483$, $SD = .924$). This finding contradicts UTT, and suggest that overall, UT led to less polarized impressions of 'Person 1' than the passive processing assumed to occur within participants in the MD conditions. However, the analysis revealed no significant main effects of WSPs or Trait, nor any significant 2 or 3-way interactions (all Fs < 1.126, ps > .05).

To test the specific predictions made in Hypotheses 1, 3, and 4, the average ratings of the inconsistent traits were subtracted from the average ratings of the consistent traits to create a difference score (higher values indicating relatively more bias). Then, the Thought and WSPs conditions were combined into one factor containing four levels to allow contrast testing using a one-way ANOVA. First, a planned contrasts revealed no impact of priming on difference scores within the MD conditions, $t(155) = .037$, $p = .971$, observed power = .056. Thus, inconsistent with Hypothesis 1, participants in MD conditions did not form more stereotypical impressions of 'Person 1' after receiving the primed relative to the control-WSPs. This suggests that priming did not impact the impressions formed by participants in the MD conditions (similar to the results for the LDT measure). Second, an additional planned contrast revealed no impact of priming on

the difference score among participants within UT conditions, $t(155) = .773$, $p = .440$, observed power = .183. This null result is consistent with hypothesis 3a and contradicts hypothesis 3b, suggesting that stereotype activation did not impact the evaluations of participants in UT conditions. Lastly, there was no effect of thought condition on difference scores among participants who received the primed-WSPs, $t(155) = .505$, $p = .614$, observed power = .098. This result contradicts both Hypotheses 4a and 4b, and suggests that participants in UT and MD conditions formed similar impressions of 'Person 1' when they were primed with a stereotype during distraction. However, it should be noted that there were no differences in any of these contrasts, including no evidence of priming at all on these measures. Therefore, it is difficult to draw conclusions about the relative amount of bias in the various conditions. Implications and possible explanations for these null findings are discussed later.

To test Hypothesis 2, average ratings on the stereotype-consistent and inconsistent trait dimensions were separately analyzed using 2 (Thought) x 2 (WSPs) between-subjects ANOVAs. For the stereotype-consistent trait dimensions, there was a non-significant main effect of thought, $F(1,155) = 3.83$, $p = .052$, $\eta^2 = .024$, observed power = .494 such that participants in UT conditions rated 'Person 1' lower on the stereotype-consistent trait dimensions ($M = 5.393$, $SD = .809$) compared to participants in the MD conditions ($M = 5.654$, $SD = .904$). Further, the analysis revealed no significant main effect of WSPs nor a significant WSPs x Thought interaction (all Fs < 1). Additionally, planned simple effects testing found no difference in ratings between participants in the UT and MD conditions who received the control WSPs, $F(1,155) = 1.25$, $p = .266$,

observed power = .199. This is inconsistent with Hypothesis 2, which predicted that in the absence of priming, UT should lead to higher ratings on the stereotype-consistent trait dimensions.

The analysis of the stereotype-inconsistent trait dimensions revealed only a significant main effect of Thought, $F(1,155) = 4.43$, $p = .037$, $\eta^2 = .028$, observed power = .553, such that participants in the UT conditions rated 'Person 1' significantly lower on the stereotype-inconsistent trait dimensions ($M = 5.483$, $SD = .924$) compared to participants in the MD conditions ($M = 5.757$, $SD = .869$). Further analysis revealed no significant main effect of WSPs, nor a significant WSPs x Thought interaction (all Fs < 1.39). Finally, planned simple effects testing found no difference in ratings between participants in the UT and MD conditions who received the control WSPs, $F(1,155) = .529$, $p = .468$, observed power = .112. Overall then, the results from both analyses contradict Hypothesis 2, which stated that UT should result in higher ratings on *both* the stereotype-consistent and inconsistent trait dimensions. This is consistent with the LDT data, which found no difference in the accessibility of stereotype-relevant information between participants in the UT and MD conditions in the absence of priming. Rather, it appears that the presumed active-integration process of UT did not result in an impression that was more polarized around the underlying traits compared to the passive process within MD conditions.

Exploratory Analysis

One assumption underlying the procedures in this experiment was that priming would have equivalent facilitative effects on the 3 stereotype-consistent traits and inhibitory effects on the 3 stereotype-inconsistent traits. However, it is possible that the processing and accessibility of these traits were differentially affected by the prime (and perhaps even a potential interaction between the prime and thought manipulations). For example, perhaps the stereotypical names primed the trait athleticism more than musical, thus weakening the results on the composite measures of stereotype-consistent items on the LDT and trait ratings. To explore this possibility, additional analyses were separately conducted on each trait measure for both the LDT and trait ratings data, the results of which are presented in Table 1 and Table 2.

LDT

Reaction times for the 4 trials representing each trait were combined to create a measure of that trait's accessibility. Then, the Thought and WSPs conditions were again combined into one factor containing four levels. Next, the average reactions times to each trait were subjected to separate one-way ANOVAs involving this factor, with planned contrasts to test for differences predicted in my 4 hypotheses (see Table 1). Looking first at Hypothesis 1, there was no impact of priming on reaction times within the MD conditions on any of the traits (all ts < 1.17 ps > .05). This is consistent with the earlier pattern of results from the primary analysis, and suggests that priming did not impact the accessibility of the stereotype-relevant traits among participants in MD conditions.

Table 1. Mean Reaction Times (in ms.) to Trait Items by Thought Condition and WSP Condition

	UT				MD			
	Primed		Control		Prime		Control	
Trait	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Athletic	691.3	(123.4)	663.9	(110.6)	699.1	(108.3)	694.4	(119.2)
Musical	795.9	(179.9)	735.1	(195.3)	796.7	(194.7)	786.5	(208.1)
Criminal	752.1	(123.3)	731.6	(180.0)	736.5	(148.5)	745.0	(151.9)
Intel.	769.3	(178.4)	723.6	(150.5)	741.5	(155.4)	751.8	(156.8)
Shy	749.6	(178.8)*	718.9	(174.3)	690.7	(123.8)*	729.9	(160.3)
HW	821.0	(172.0)	794.2	(176.8)	843.7	(172.4)	798.0	(155.6)

Note: * Signifies significant differences between the means at $p < .05$.

Further, there was no difference in the reaction times on any traits between the UT and MD conditions among participants who received the control-WSPs (all *t*s < 1.36, *p*s > .05). This contradicts Hypothesis 2, and supports the findings from the primary analysis, that in the absence of stereotype activation, the accessibility of the stereotype-relevant traits was similar among participants within the UT and MD conditions. Third, there was no impact of priming on reaction times to any of the traits among participants within the UT conditions (all *t*s < 1.31, *p*s > .05). This is inconsistent with Hypothesis 3b and the logic undergirding the primary analysis in which priming led to significantly larger difference scores among participants in UT conditions. Lastly, among participants who received the primed-WSPs, reaction times differed significantly between thought conditions only for the shy targets, $t(173) = 1.67$, $p = .048$ (one-tailed) $d = .25$, observed power = .45, such that participants in the UT condition responded slower on the shy trails compared to participants in the MD condition (all *t*s < .81, *p*s > .05 for this comparison on the other trait dimensions)[2].

Overall, the exploratory analysis on the LDT data does not suggest that priming disproportionately impacted certain traits. This conclusions was also supported by the results of a 2 (Thought) x 2 (WSPs) x 3 (Consistent Traits) mixed-model ANOVA, and a 2 (Thought) x 2 (WSPs) x 3 (Inconsistent Traits) mixed-model ANOVA, neither of which found the Trait x WSPs, or 3-way interactions that would indicate disproportionate priming of certain traits. Further, given Hypothesis 3b was supported in the primary analysis using the composite measures, but not on the individual items, suggests that the individual items might not be as reliable of measures.

Trait Ratings

As indicated in the Method section, the reliability among the three stereotype-consistent and three inconsistent traits was extremely poor ($\alpha = .292$; $\alpha = .513$) respectively. This low reliability brings into question whether the three consistent (and inconsistent) traits should actually be viewed as measuring the same construct and represents a strong reason to analyze the traits individually. The trait ratings data was subjected to the same analyses above, and the results are displayed in Table 2. Only two of the contrasts resulted in significant differences (all other ts < 1.52, ps > .05). First, among participants within the MD conditions, those who received the primed-WSPs rated Person 1' significantly higher on the trait of criminal than participants who received the control-WSPs, $t(155)=1.961$, $p = .026$ (one-tailed), $d = .32$, observed power = .59. This is

---

2 This was not significant when both the bias-aware and marginally bias-aware groups were excluded from analysis $t(153)= 1.608$, $p = .055$ (one-tailed), or when the theme-only group was additionally excluded $t(146)= 1.136$, $p = .129$ (one-tailed).

consistent with Hypothesis 1, and suggests that among participants in the MD conditions, priming lead them to view 'Person 1' as more criminal (stereotypical). Second, among participants who received the primed-WSPs, those in the UT condition rated 'Person 1' significantly lower on intelligence than participants in the MD condition $t(155) = 2.245$, $p = .013$ (one-tailed), $d = .36$, observed power =.62, consistent with Hypothesis 4b. Thus, the conditions did not differ much on their ratings of the individual traits, although 2 significant comparisons among all of the trait ratings supported Hypotheses 1 and 4b.

Table 2. Mean Trait Ratings by Thought Condition and WSP Condition

	UT				MD			
	Primed		Control		Prime		Control	
Trait	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Athletic	5.97	(1.354)	6.09	(0.830)	6.26	(1.024)	6.26	(0.871)
Musical	4.48	(1.655)	4.87	(1.498)	5.26	(1.88)	5.34	(1.536)
Criminal	5.42	(1.119)	5.19	(1.345)	5.74*	(1.355)	5.15*	(1.414)
Intel.	5.52*	(1.208)	5.66	(1.273)	6.15*	(1.019)	6.00	(1.00)
Shy	4.74	(1.653)	5.17	(1.34)	5.21	(1.647)	5.00	(1.518)
HW.	5.90	(1.193)	5.81	(1.076)	6.24	(0.923)	6.04	(1.179)

Note: * Signifies significant differences between the means at $p < .05$

Overall, the results of the exploratory analysis do not support the notion that priming differentially impacted certain traits. Similar to the LDT data, this conclusion was also supported by the results of a 2 (Thought) x 2 (WSPs) x 3 (Consistent Traits) mixed-model ANOVA, and a 2 (Thought) x 2 (WSPs) x 3 (Inconsistent Traits) mixed-model ANOVA, neither of which found the Trait x WSPs, or 3-way interactions that would indicate disproportionate priming of certain traits. Potential reasons for the trait ratings lacking sensitivity, including the possible existence of a ceiling effect, are discussed later.

DISCUSSION

According to UTT, UT has a vast processing capacity that allows it to process information in a bottom-up, aschematic manner (Dijksterhuis & Nordgren, 2006). Based on these principles, UTT predicts that UT should utilize stereotypes less than schema-driven top-down process like CT. Consistent with this prediction, Bos and Dijksterhuis (2011) found evidence for greatly reduced stereotype activation and biased impressions following a period of UT relative to CT—at least when a stereotype was explicitly activated prior to information acquisition. From their results, Bos and Dijksterhuis concluded that UT is not biased by stereotypes when forming impressions of a person from a stereotyped group. Nonetheless, there is evidence to suggest that participants may form biased impressions in experimental contexts arguably suitable to the operation of UT if a stereotype is non-consciously activated during distraction periods following information acquisition (van Knippenberg & Dijksterhuis, 1996). Thus, the current experiment sought to expand the literature on UT and investigate whether the theoretically active process of UT and the passive processes within MD conditions are affected differently by non-conscious stereotype activation during a period of distraction.

This experiment tested these competing predictions using a person-impression paradigm. All participants read various descriptions of 'Person 1,' and received the goal to form an overall impression of him. Some of the descriptions implied traits that were consistent with the stereotype of African-American men, others were inconsistent, and still others were stereotype-irrelevant. After viewing the descriptions, participants were randomly assigned to periods of distraction in which they held an impression-formation

goal (i.e., thought unconsciously) or not while completing a 3min distraction task (word-search puzzles). Further, participants were randomly assigned to complete a distraction task designed to either unconsciously prime, or not prime, the stereotype of African-American men. After the distraction period, the accessibility of the stereotype-consistent and inconsistent traits implied by the earlier descriptions was measured using an LDT. Lastly, participants were asked to rate the extent to which they felt 'Person 1' possessed the specific stereotype-related traits implied by the earlier descriptions.

Focusing first on participants in the MD conditions, previous theorizing and research suggests that stereotype activation (versus no activation) within passive contexts should facilitate the accessibility of stereotype-consistent information and inhibit the accessibility of inconsistent information (e.g., Macrea, et al., 1994), as well as result in more stereotypical impressions. Inconsistent with Hypothesis 1, the results from the current experiment did not align with this common pattern of results. Specifically, the difference in reaction times to the stereotype-consistent and inconsistent trials on the LDT was comparable for participants who received the primed and control-WSPs within the MD conditions. Similarly, these participants reported comparable differences in ratings of 'Person 1' on the stereotype-consistent and inconsistent trait dimensions. Overall then, these results suggest a basic lack of priming in the MD conditions; possible explanations for this finding are discussed later.

General Predictions from UTT

In the absence of priming, participants in the UT condition should process all of the information to a greater extent than participants in the MD condition. Therefore, one would expect to see greater accessibility of stereotype-consistent *and* inconsistent information among participants in UT conditions compared to MD conditions if they received the control-WSPs (Hypothesis 2). Similarly, UT should have resulted in more polarized impressions of 'Person 1' (i.e., higher ratings on all traits; Bos et al., 2008). Contrary to these predictions, participants in the UT conditions did not respond faster (or form more polarized impression) than those in the MD conditions on the stereotype-consistent or inconsistent targets within the LDT (or trait ratings) when they were not primed with the stereotype (i.e., control-WSPs). This is contradictory to the principles of UTT, and previous research. For instance, Zhong, Dijksterhuis, and Galinsky (2008) found that participants in a UT condition reacted more quickly to goal-relevant information on a LDT compared to participants in a MD condition. Both of these null results in the present experiment are inconsistent with Hypothesis 2, and suggest that in the contexts for which a stereotype was not activated, participants in the UT condition did not actively process the information to a greater extent than participants in the MD condition. It is important to note that one of the indicators of this active-integration process is a more organized impression. Although this organization should result in participants forming more polarized impressions (i.e., higher ratings on all of the traits) in the UT condition, clustering of recall may serve as a more accurate marker. As an exploratory measure, at the end of the experiment all participants were allowed 2min to

recall as much information about 'Person 1' as they could. This data has not been coded and analyzed yet, but it might offer more insight into the impressions formed by participants in the UT and MD conditions when no stereotype was activated.

Can Stereotype Activation Bias the Outcomes of Unconscious Thinking?

According to the impermeable hypothesis proposed by UTT, stereotype-consistent and inconsistent information should be comparably accessible among participants in the UT conditions, regardless of whether or not they have an activated stereotype. Similarly, the impermeable hypothesis would predict that participants should form comparable impressions regardless of priming conditions if they engage in UT (see Hypothesis 3a). Alternatively, Hypothesis 3b (the permeable hypothesis) predicts that UT will (under certain circumstances) inhibit the accessibility of stereotype-inconsistent concepts but facilitate the accessibility of stereotype-consistent concepts, and lead to more stereotypical impressions if stereotypes are active (i.e., primed), but not otherwise.

The results from the current experiment offer mixed support for Hypotheses 3a and 3b. Participants in UT conditions responded significantly slower to the stereotype-inconsistent targets than the consistent targets on the LDT when they received the primed compared to the control-WSPs. This finding is consistent with the permeable hypothesis (Hypothesis 3b) and opposes Hypothesis 3a (UTT), and suggests that stereotype activation can bias the accessibility of stereotype-relevant information while individuals engage in UT. Interestingly, this difference in stereotype accessibility observed among participants who were, and were not, primed in the UT conditions did not result in more

biased impressions of 'Person 1.' Rather, participants in the UT conditions who received the primed and control-WSPs rated 'Person 1' similarly on the stereotype-consistent and inconsistent trait dimensions. However, the exceptionally poor reliability among ratings for the consistent and inconsistent traits suggests that the three consistent (and inconsistent) ratings were not measuring the individual construct of a stereotypical (or non-stereotypical) person, casting doubt on their utility as a measure of stereotype application. Further, it is likely that a ceiling effect was present, as will be discussed later. Lastly, neither the significant effect in the LDT data, nor the non-significant effect in the trait ratings were driven by a few participants[4]. Overall then, these results suggest that stereotype activation during periods of UT can bias the *accessibility* of stereotype-relevant information (supports Hypothesis 3b for the LDT). However, it remains unclear how this biased accessibility would impact the impressions formed by UT.

Do Activated Stereotypes Bias Outcomes from UT more than Negligible Thought?

According to the impermeable hypothesis, UT processes are less biased by stereotype activation relative to contexts in which negligible thinking occurs, a finding which would be consistent with UTT. On the other hand, the permeable hypothesis

---

4 This was tested by subtracting the number of participants not showing the effect from the number showing the effect and dividing it by the square root of n, to create a z-score. For instance, to show that the biased responding among participants in the UT-prime condition was not due to a few subjects, the number of subjects with a negative difference score was subtracted from the number with a positive difference score, and then divided by the square root of the total subjects in the UT-prime conditions (23-8/ $\sqrt{31}$ = 2.69). Scores greater then 1.96 indicate that a significant portion of participants in the condition showing the effect drives these results. Alternatively, the *null effect* among the trait ratings data for participants in the UT-prime condition has a score of -.179, suggesting that a significant portion of participants in the UT-prime condition also showed the null effect.

predicts that stereotype activation should bias UT processes to a *greater* extent than negligible thinking process. Inconsistent with both of these Hypotheses, there was no significant difference in responding to the consistent and inconsistent targets among participants in the UT and MD conditions who received the primed-WSPs. Similarly, participants in the MD and UT conditions formed comparable impressions of 'Person1,' a pattern of results that supports neither Hypotheses 4a nor 4b. As previously noted, the poor reliability of the composite trait-ratings DVs casts doubt on their utility as a measure of stereotype application. Further, the lack of significant differences among the contrasts for all of the trait ratings data, including no observable effect of priming, makes it difficult to draw conclusions from those results.

Interim Summary

Overall, the pattern of results from this experiment offer mixed support for the conflicting hypotheses, meaning the extent to which UT is susceptible to stereotype activation and use remains unclear. Perhaps the most puzzling finding is the lack of a priming effect in the LDT data among participants in the MD conditions. There are at least four possible explanations for this. First, the priming procedure might not have effectively activated the stereotype. However, this interpretation fails to explain the priming effect observed among participants within the UT conditions. The second possibility is that participants in the MD conditions somehow actively suppressed the stereotype. Although this prospect cannot be completely ruled out, there is nothing in the

literature to suggest that the passive processes assumed to be operating within MD conditions actively suppress stereotypes. Third, and alternatively, perhaps the stereotype was activated, but decayed before it could be measured on the LDT (e.g., Kunda & Spencer, 2003). The stereotype not decaying, or decaying at a slower rate among participants in the UT condition, could result from the active processes within UT maintaining the stereotype. Finally, it is possible that the difficulty of the distractor task hampered participants' performance on the LDT, and this was sufficient to mask the effects of priming on participants within the MD condition for whom the effect was predicted to be less pronounced. To elaborate, participants in the priming conditions solved significantly less WSPs (M = 8.66, SD = 2.16) than participants in the control conditions (M = 9.98, SD = 3.32) $t(188) = 3.240$, $p = .001$, $d = .47$. Assuming African-American names are not commonly encountered among members of the MSU subject pool (the vast majority of whom are White), the African-American names in the primed-WSPs may have been harder to identify than the names in the control-WSP. It is possible that the increased difficulty of finding uncommon names was more exhausting than finding fairly common names, and this exhaustion hindered participants' performance on the LDT. Indeed, previous research has shown that the difficulty of a prior task can impact LDT performance (Kneer, Elson & Knapp, 2016). If this were the case, one would expect to see slower reaction times to the 12 neutral targets on the LDT among participants who received the primed compared to control-WSPs. Indeed participants who received the control-WSPs did respond numerically faster to the neutral targets (M = 716.739, SD =111.124) compared to participants who received the primed-WSPs (M =

746.774, *SD* = 116.737), however the difference was not statistically significant $t(150) = 1.625$ $p = .106$. Although this trend was not significant, perhaps it was sufficient to mask the effects of priming among participants in the MD condition for whom the impact of priming was hypothesized to be less pronounced. Regardless of the correct explanation, the most important findings occurred among participants in the UT conditions. Contrary to the predictions of the impermeable hypothesis (UTT), these results suggest that stereotype activation during UT can bias the accessibility of stereotype-relevant information.

However, it remains unclear why the biased accessibility of information among participants in the UT condition did not lead to stereotype use. It could be that the accessibility of information in UT can be biased by stereotype activation, but the products of UT (e.g., impressions, judgments, choices) are not impacted by the stereotype. Indeed, previous research has shown that stereotype activation does not always translate into application (e.g., Gilbert & Hixon, 1991). Still, and perhaps most obviously, it is likely that the trait ratings were simply not a sufficiently sensitive, or reliable, measure of stereotype application given the current sample size. As previously stated, the poor reliability among the consistent and inconsistent trait ratings casts doubt on their utility as a measure of stereotype application. Further, the descriptions may have so unambiguously characterized the traits that there was little room for thought or priming conditions to impact participants' impressions (i.e., a ceiling effect occurred such that 'Person 1' was clearly high on all traits). Indeed, looking at the skewness of the trait-rating data (see Table 3), there does appear to be a clustering of scores around the top end

of the distribution. This leaves open the possibility that the lack of an observable impact of stereotype activation on judgments resulted from too little variability in the measures of participants' impressions. Overall, the results from the present experiment suggest that stereotype activation can bias the accessibility of stereotype-relevant information within UT conditions, but it remains unclear how this biased accessibility compares to non-UT conditions and whether it impacts impressions.

Table 3. Descriptive statistics for the Trait-Ratings data.

Trait	Mean (SD)	z-skew	z-kurtosis
Athletic	6.15 (1.00)	-8.391*	9.313*
Criminal	5.34 (1.34)	-2.932*	-.900
Musical	5.09 (1.63)	-3.531*	-.807
Hard working	5.99 (1.10)	-4.953*	.548
Intelligent	5.84 (1.15)	-5.479*	4.097*
Shy	5.04 (1.52)	-2.276*	-1.68
Consistent	5.53 (.866)	-1.875	-.493
Inconsistent	5.63 (.904)	-2.589*	-.287

Note:* signifies a significant Z-score at $p < .05$.

Limitations and Future Directions

This experiment utilized a modified version of the person-impression task from Bos and Dijksterhuis (2011), with a novel priming manipulation and materials; thus, the experiment has several potential limitations. First, although the trait descriptions were validated by a sample of native English-speaking American citizens sampled from MTurk, it is possible that the participants from the subject pool of introductory psychology students at MSU perceived the descriptions as implying different traits, or as being less stereotypical than did the MTurk sample. Indeed, workers on MTurk are more

ethnically diverse (e.g., 7.1% African-American, 5.6% Hispanic, 71.3% White; Levay, Freese & Druckman, 2016) then the MSU population in 2017 (e.g., 0.45% African-American, 3.9% Hispanic, White = 83.4%; Ethnic Distribution of Majors by College, n.d.). Future researchers could pre-test materials and conduct experiments drawing from the same subject pools to eliminate this possibility. Second, it is highly plausible that a ceiling effect was present in the trait ratings data. This might have occurred because 24 of the 32 descriptions implied stereotype-related traits, whereas only 8 of the descriptions were neutral. This high ratio of stereotype-related to neutral traits could have left little room for priming or thought conditions to impact participants' impressions. Future research could attempt to eliminate this problem by using descriptions that are slightly more ambiguous, and include more neutral descriptions. Third, the effects of stereotype activation on accessibility of information were measured using a LDT. Future research could employ different implicit measures (e.g., IAT, word-stem completion tasks) to replicate and generalize these findings.

This experiment utilized WSPs as a distractor task and to prime the African-American stereotype in a manner that is novel within UTT research. When designing these manipulations, I made many choices that I felt were justifiable; yet they still present some possible limitations. First, I employed a relatively high ratio of priming to neutral words in the primed-WSPs (i.e., overall ratio of 16/12, with the first 20 having an even higher ratio of 12/8). This high ratio was selected in order to maximize the chance of actually priming participants with the WSPs, and is an adaptation of a 12/8 ratio utilized by Welsh and Ordonez (2014). However, the heavy handedness of such a priming

procedure has several limitations. First, Welsh and Ordonez utilized a sentence unscrambling task to prime ethical norms; perhaps the theme of morality was harder for participants to detect than the theme of African-American men's names in the current experiment, justifying the high ratio Welsh and Ordonez used. Alternatively, it is unclear if a smaller ratio, perhaps where only 25% of the WSPs involve stereotype-priming names, would be sufficient to influence stereotype accessibility and bias impressions. Future research could investigate this possibility. Similarly, the funnel debrief revealed that 32 participants noticed the theme of African-American men's names, and between 14-23 of them hypothesized that the purpose of the theme was to impact their impressions of 'Person 1.' Such a high rate of awareness leaves open the possibility that many other participants also were aware of the primes and either forgot by the time they reached the funnel debrief or did not disclose this awareness. Alternatively, the number of truly aware participants could be much lower, and the step-wise nature of the funnel debrief made them "connect the dots" later, as one participant openly claimed. Either way, future research could employ a more covert method of priming, perhaps even using subliminal primes, to remove any possible confounding effects from awareness of the prime. Similarly, as previously mentioned, it is possible that solving the primed-WSPs was more difficult then the control-WSPs, and this might have influenced LDT performance; thus, possibly masking the effects of priming among participants within the MD conditions, where the impact of priming was hypothesized to be less pronounced. Therefore, employing a subliminal priming procedure could also eliminate the confounding effects of difficulty. Lastly, the number of primed-WSPs a participant saw depended on their rate

of completion. Thus, in order to maximize exposure to the primed-WSPs, I made the decision to front-load the distraction period with a higher ratio of primed to control-WSPs. There are two possible implications of this decision. First, some participants in the priming conditions received more primed-WSPs then others. Consequently, some participants may have been more heavily primed than others, adding variability to the study's measures. Second, the stereotype was repeatedly activated early on during the distraction task. Because information is actively integrated and organized during UT, it is possible that early and repeated activation led to greater bias than would occur from latter activation, when the information is likely already organized to a greater extent. Future research could manipulate when the stereotype is first activated during the distraction task to determine if this differentially impacts impressions and the organization of information in memory among participants in UT and MD conditions. Looking at the clustering of recall after manipulating timing of activation could be especially revealing when trying to characterize how activation impacts integration.

The purpose of this study was to explore the conditions under which UT could be biased by stereotypes. Thus, using MD as a control condition was desirable because the experience of participants in the two conditions is almost identical. However, perhaps the passive processes assumed to be operating within MD conditions are not sufficient to maintain an activated stereotype. This might explain the null effects of priming on LDT performance in the MD conditions. It may be more appropriate to compare the effects of stereotype activation during UT to another condition that includes active processes (i.e., CT). Such a study might actually support the *general* conclusions of Bos and Dijksterhuis

(2011) if it finds that UT is less biased by stereotype activation than CT. Although the current method of priming excluded the possibility of a CT condition, future research could attempt to employ a method that would allow for priming during CT.

Conclusion

In conclusion, contrary to the predictions of UTT, this experiment found evidence that stereotype activation during UT can bias the accessibility of stereotype-relevant information (supporting the permeable Hypothesis). Importantly, stereotype activation during MD did not bias the accessibility of information, suggesting that some characteristic of UT is responsible. However, due to possible issues with the measurement of stereotype application used in the experiment, it is unknown if stereotype activation did or did not impact the impressions of UT, and future work could attempt to employ more sensitive and reliable measures to capture possible undetected effects.

REFERENCES CITED

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The english lexicon project. *Behavior research methods*, *39*(3), 445-459.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological science*, *7*(3), 136-141.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American economic review*, *94*(4), 991-1013.

Bos, M. W., & Dijksterhuis, A. (2011). Unconscious thought works bottom-up and conscious thought works top-down when forming an impression. *Social Cognition, 29*(6), 727-737.

Bos, M. W., Dijksterhuis, A., & van Baaren, R. B. (2008). On the goal-dependency of unconscious thought. *Journal of Experimental Social Psychology, 44*(4), 1114-1120.

De Deyne, S., Navarro, D., Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single word associations. *Behavior Research Methods*, 45 (2), 480-498.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5-18.

Dijksterhuis, A. (2004). Think different: The merits of unconscious thought in preference development and decision making. *Journal of Personality and Social Psychology, 87*, 586–98.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F. & van Baaren, R. B. (2006). On making the right choice: The deliberation-without-attention effect. *Science, 311*(5763).

Dijksterhuis, A., Bos, M. W., Van der Leij, A., & Van Baaren, R. B. (2009). Predicting soccer matches after unconscious and conscious thought as a function of expertise. *Psychological Science*, *20*(11), 1381-1387.

Dijksterhuis, A. & Nordgren, L. F. (2006). A theory of unconscious thought. *Perspectives in Psychological Science, 1,* 95–109.

Dijksterhuis, A., & Van Knippenberg, A. (1995). Memory for stereotype-consistent and stereotype-inconsistent information as a function of processing pace. *European Journal of Social Psychology, 25*(6), 689-693.

Dijksterhuis, A., & van Knippenberg, A. (1996). The knife that cuts both ways: Facilitated and inhibited access to traits as a result of stereotype activation. *Journal of Experimental Social Psychology, 32*(3), 271-288.

Dijksterhuis, A., van Knippenberg, A., Kruglanski, A. W., & Schaper, C. (1996). Motivated social cognition: Need for closure effects on memory and judgment. *Journal of Experimental Social Psychology, 32*(3), 254-270.

Ethnic Distribution of Majors by College. (n.d). Retrieved from http://www.montana.edu/opa/facts/ethnicm.html#TOT

Fyock, J., & Stangor, C. (1994). The role of memory biases in stereotype maintenance. *British Journal of Social Psychology, 33*, 331–343.

Garrison, K., & Handley, I. (2017). Not Merely Experiential: Unconscious Thought Can Be Rational. *Frontiers in Psychology, 8*, 1096.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60*(4), 509-517.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology, 74*(6), 1464.

Heider, J. D., Scherer, C. R., Skowronski, J. J., Wood, S. E., Edlund, J. E., & Hartnett, J. L. (2007). Trait expectancies and stereotype expectancies have the same effect on person memory. *Journal of Experimental Social Psychology, 43*(2), 265-272.

Kasof, J., & Steinberg, Robert J. (1993). Sex Bias in the Naming of Stimulus Persons. *Psychological Bulletin, 113*(1), 140-163.

Kneer, J., Elson, M., & Knapp, F. (2016). Fight fire with rainbows: The effects of displayed violence, difficulty, and performance in digital games on affect, aggression, and physiological arousal. *Computers in Human Behavior*, *54*, 142-148.

Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological bulletin*, *129*(4), 522.

Levay, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of Mechanical Turk samples. *Sage Open*, *6*(1), 2158244016636433.

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.

Macrae, C. N., Stangor, C., & Milne, A. B. (1994). Activating social stereotypes: A functional analysis. *Journal of Experimental Social Psychology, 30*(4), 370-389.

Madon, S., Guyll, M., Aboufadel, K., Montiel, E., Smith, A., Palumbo, P., & Jussim, L. (2001). Ethnic and National Stereotypes: The Princeton Trilogy Revisited and Revised. *Personality and Social Psychology Bulletin, 27*(8), 996-1010.

Nieuwenstein, M. R., Wierenga, T., Morey, R. D., Wicherts, J. M., Blom, T. N., Wagenmakers, E., & van Rijn, H. (2015). On making the right choice: A meta-analysis and large-scale replication attempt of the unconscious thought advantage. *Judgment and Decision Making, 10*(1), 1-17.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological review*, *111*(1), 159.

Rojahn, K., & Pettigrew, T. F. (1992). Memory for schema-relevant information: A meta-analytic resolution. *British Journal of Social Psychology*, *31*(2), 81–109.

Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). " Unlearning" automatic biases: the malleability of implicit prejudice and stereotypes. *Journal of personality and social psychology*, *81*(5), 856.

Scarabis, M., & Florack, A. (2008). How the Motivation to Make Fair Judgments Influences Memory for In- and Out-Group Behavior. *Swiss Journal of Psychology, 67*(4), 241-248.

Sechrist, G. B., & Stangor, C. (2001). Perceived consensus influences intergroup behavior and stereotype accessibility. *Journal of Personality and Social Psychology, 80*(4), 645-654.

Sherman, J. W., Stroessner, S. J., Conrey, F. R., & Azam, O. A. (2005). Prejudice and stereotype maintenance processes: attention, attribution, and individuation. *Journal of personality and social psychology*, *89*(4), 607.

Sinclair, L. M. (1998). *Justifying desired impressions of evaluators: Motivated activation, application, and inhibition of stereotypes* (Order No. NQ32857). Available from ProQuest Dissertations & Theses Global.

Snyder, M., & Uranowitz, S. W. (1978). Reconstructing the past: Some cognitive consequences of person perception. *Journal of Personality and Social Psychology, 36*(9), 941-950.

Srull, T. K., Lichtenstein, M., & Rothbart, M. (1985). Associative storage and retrieval processes in person memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(2), 316-345

Strick, M., Dijksterhuis, A., Bos, M. W., Sjoerdsma, A., van Baaren, R. B., & Nordgren, L. F. (2011). A meta-analysis on unconscious thought effects. *Social Cognition, 29*(6), 738-762.

Van Knippenberg, A., & Dijksterhuis, A. (1996). A posteriori stereotype activation: The preservation of stereotypes through memory distortion. *Social Cognition, 14*(1), 21-53.

Vergara-Martínez, M., & Swaab, T. Y. (2012). Orthographic neighborhood effects as a function of word frequency: An event-related potential study. *Psychophysiology*, *49*(9), 1277–1289.

Welsh, D. T., & Ordóñez, L. D. (2014). Conscience without cognition: The effects of subconscious priming on ethical behavior. *Academy of Management Journal*, *57*(3), 723-742.

Zhong, C. B., Dijksterhuis, A., & Galinsky, A. D. (2008). The merits of unconscious thought in creativity. *Psychological science*, *19*(9), 912-918.