Testing Perceptual Disfluency as a Strategy to Diminish Illusory Truth Effects

By

Erin J. Dowling

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

In partial fulfilment of the requirements of the degree of

MASTER OF ARTS

Department of Psychology

University of Manitoba

Winnipeg

Copyright © 2022 by Erin J. Dowling

Abstract

People exhibit a persistent bias to believe information that is not true and to act on these beliefs. Contemporary findings illuminate the destructive costs of misinformation and false claims in the media; belief in misinformation has been flagged as a significant contributor to recent and ongoing querulous, polarizing events including the handling of the global pandemic, election outcomes, and events of racial and political persecution. One persistent obstacle to preventing belief in misinformation is the fluency with which previously encountered information is processed. *Illusory truth effects* refer to the robust finding that memory for information can be distorted (and is more likely to be remembered as true) due to mere exposure of false or misleading details; information that is re-encountered is processed with ease, and this fluency is misattributed to a statements' veracity. The current study is an attempt to replicate earlier work conducted in the domain of illusory truth effects (i.e., Gilbert et al., 1993), and involves two experiments which test a novel *disfluency* manipulation of false statements (e.g., tHe RoBbEr HaD a GuN) in a mock criminal trial scenario. Currently, empirical research testing the potential influence of fluency disruptions on memory for false claims is limited, and findings concerning the efficacy of disfluency as a strategy to deter belief in a statement are inconclusive. Results from both experiments support the hypothesis that people perform poorly when asked to disregard information explicitly identified as false and will use this information in decisions of high consequence, despite receiving specific instruction to disregard it. Although this outcome was tempered slightly when a warning about the presence of false information was provided, participants who encoded false crime report details DiSfLuEnTlY were 9% more likely to misremember these details as true and assigned harsher character assessments and prison sentences to suspects as a result. In addition, multiple regression analysis finds that scoring high in right-wing authoritarianism (RWA) and/or social dominance orientation (SDO) significantly predicted the length of prison sentence that participants assigned to suspects and the proportion of false details remembered as true, particularly when the false details reported exacerbated the suspect's crime.

Keywords: illusory truth effect, misinformation, fluency, disfluency, right-wing authoritarianism, social dominance orientation

Acknowledgements

I would like to extend my utmost gratitude to my advisor, Dr. Launa Leboe-McGowan, for her exceptional guidance and support throughout the course of my degree, through a pandemic. I also wish to extend sincere appreciation and thanks to my committee members, Dr. Jason Leboe-McGowan and Dr. Kelley Main for their valuable feedback and thoughtful insights throughout the entirety of this project.

This research was supported by a Canada Graduate Scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC) in addition to a Tri-Council Top-Up Award and two Program Completion Scholarships awarded by the University of Manitoba.

Dedication

Thank-you to my amazing parents, Bill and Heather Dowling for their steadfast love, support and encouragement. I also am so grateful to my incredible support system of wonderful friends, most notably my dear friends Susan Hunter and Gord Zacharias—thank-you for letting me take over your dining room table, and for cheering me across the finish line.

Table of Contents

Abstract…….…. ii

Acknowledgements……....iii

Dedication……...iv

Table of Contents……..... v

List of Tables……. ix

List of Figures…… x

Testing Perceptual Disfluency as a Strategy to Diminish Illusory Truth Effects…….…. 1

The Illusory Truth Effect……5

Factors Known to Contribute to Illusory Truth Effects ……..…7

Repetition and Time…….……..8

Warnings and Specific Cues …….…….9

Individual Differences ……. 10

Right-Wing Authoritarianism (RWA) ……10

Social Dominance Orientation (SDO) …….12

Effects of Disfluency …….…. 13

The Present Study ……15

Experiment 1 ……... 18

Method……. 18

Participants ……. 19

Design ……. 20

Materials ……. 21

Independent Variables…… .22

Crime Reports……....22

Exacerbating and Extenuating Conditions ……..…. 24

Disfluent Text Condition ……..…… 25

Warning Condition …….…...…….. 26

Interruption Condition ……. 26

Conditional Instructions ……29

Warning Condition Instruction……. 29

Interruption Condition Instruction……30

Dependent Variables……. 31

Character Assessment Measures……31

Sentencing Measure……...32

Recognition Memory Tests ……...32

Personality Constructs as Moderating Variables……...33

Right-Wing Authoritarianism (RWA) Scale…….34

Social Dominance Orientation (SDO) Scale…….35

Procedure……...36

Results……37

Data Exclusion Criteria……. 37

Testing the Significance of Order and Suspect as Variables…….39

Character Assessments and Prison Sentencing Scores …….40

Likeability Ratings ……40

Danger Ratings……..……43

Likelihood to be Helped by Counselling Ratings.....44
Prison Sentencing …….……45
Recognition Memory Test Performance ……...46
Comparing Recognition Memory Outcomes with Gilbert's Results…….……..….48
Effects of False Format on Memory and Sentencing Outcomes …….…….51
Assessing RWA and SDO as Potential Moderating Variables ……..……53
Gender Differences ……...53
Analysis of Right-Wing Authoritarianism Scores…….54
Analysis of Social Dominance Orientation Scores……...…..…58
Discussion ……...….61
Experiment 2……..……64
Method…….……..64
Participants …….……..64
Design…….……...65
Materials …….……..66
Independent Variables …….…….66
Repetition Condition …….……66
Results…….…….…….67
Data and Variable Exclusion …….……. 60
Character Assessments and Prison Sentencing Scores……...……..69
Likeability Ratings …….…….69
Danger Ratings …….…….69
Likelihood to be Helped by Counselling ……...…...70

Prison Sentencing……. 71
Recognition Memory Test Performance ……..…..72
In Comparison to Gilbert's Interruption Findings ……72
Effects of Repetition and False Format on Memory
and Sentencing Outcomes…….73
Assessing RWA and SDO as Potential Moderating Variables……..…75
Gender Differences……75
Analysis of Right-Wing Authoritarianism Scores ……75
Analysis of Social Dominance Orientation Scores ……..78
Discussion……79
General Discussion……81
References……88
Appendices……...105

List of Tables

Table 1: Mean Character Assessment Ratings and Prison Terms Assigned to Suspects (Experiment 1)……..42

Table 2: Comparison of Interruption Effects on Memory Errors (Experiments 1 & 2)…….. 50

Table 3: Correlations Between Tendency to Respond 'True' to False Statements and Character Assessment Ratings (Experiment 1).…….…… 51

List of Figures

Figure 1: Main Effects of False Detail Severity and Interaction Effects Between Severity and Warning on Character Assessments (Experiment 1) ……...……... 43

Figure 2: Main Effect of Disfluency on Tendency to Misremember False Details as 'True' (Experiment 1)...……52

Figure 3: Level of RWA Predicts Length of Prison Sentence for Exacerbated Crimes (Experiment 1)……...55

Figure 4a: Level of RWA Predicts Tendency to Misremember False Details as 'True' (Experiment 1)..……. 57

Figure 4b: Level of SDO Predicts Tendency to Misremember False Details as 'True' (Experiment 1)……60

Figure 5: Interaction Effect Between False Format and Repetition on the Tendency to Misremember False Details as 'True' (Experiment 2)……..74

Figure 6: Level of RWA Influences Length of Prison Sentence for Exacerbated Crimes (Experiment 2)…….77

Figure 7: Level of RWA Influences Tendency to Misremember False Details As 'True' (Experiment 2)……..78

Testing Perceptual Disfluency as a Strategy to Diminish Illusory Truth Effects

Misinformation, recently defined as "any information that turns out to be false" has been a hallmark of mass communication for centuries (Ecker et al., 2022, p. 13); the mythologies and beliefs of large groups of people are as varied as they are ancient—nonetheless surely brushed aside by outsiders as nonsense (Levine, 2022; Posetti & Matthews, 2018). Awareness of the pervasiveness and consequence of misinformation, however, is relatively contemporary and of growing concern as the prevalence and dissemination of mis- and disinformation has increased exponentially in recent decades. The advent of the internet and social media platforms has effectively catapulted the global audience into an era where information—accurate or otherwise—has the capacity to travel more rapidly and vastly than ever before, and anyone can be an 'expert' (Brashier & Marsh, 2020; Larsen, 2014; Levine, 2022, Lewandowsky et al., 2017; Love et al., 2020; Swire-Thompson & Lazer, 2020).

People are generally not very good at assigning credibility to information they encounter in the media, tending to gauge authenticity based on heuristic cues such as the esthetic appeal of a website over substance and evidence supporting or refuting the message (Fogg et al., 2003; Metzger et al., 2010). Perhaps even more troubling is that false information moves faster and reaches more people than accurate information in many online environments (Visoughi et al., 2018); environments such as Facebook 'news' feeds and 5-second TikTok clips, now specifically tailored and targeted to attract (and actively provoke) select individuals (Berger, J., 2011; Gabielkov et al., 2016; Gottfried & Shearer, 2016; Johnson & Kaye, 2016; Tufecki, 2015).

Perhaps the only thing as multi-faceted and far-reaching as misinformation itself are the consequences of it. Belief in misinformation has been flagged as a significant contributor to several recent and ongoing querulous, polarizing events including the handling of the global

Covid-19 pandemic (Love et al., 2020), election outcomes (Hughes & Waismet-Manor, 2020), and events of racial and political persecution (Gover et al., 2020; Lewandowsky et al., 2022; Write & Duong, 2021). Reliance on misinformation in these scenarios has led to drastic consequences including cell phone towers set ablaze as a direct result of 5G and QAnon conspiracy theories (Meese et al., 2020), and communities experienced sharp increases in hate crimes directed toward Asian Canadians following widespread misinformation concerning the origin of the then novel coronavirus (Bresnahan et al., 2022; Mian & Khan, 2020; Tessler et al., 2020). As Love and colleagues point out, scientific mapping of rumor spreading demonstrates striking similarities to the epidemiological spread of communicable viruses (Daley & Kendall, 1964; Love et al., 2020). The mortality and comorbidity rates of the Covid-19 pandemic, and other epidemics, are only compounded by belief in misinformation and unproven claims. For example, in Iran, a rumor claiming that alcohol could cure the Covid-19 virus caused at least 300 deaths due to toxic methanol poisoning (Love et al., 2020). In the U.S., similar outcomes have been the result of bleach consumption following the brazen claims of a particular political figure of the time (Dharawat et al., 2022; Frenkel & Alba, 2020).

Misinformation is a problem that will inevitably persist. Developing a clearer understanding of the mechanisms that influence a person's likelihood of resorting to false information is crucial if we hope to develop strategies to assist cognitive deflection of inaccurate claims. As the literature review that follows highlights, bringing people to successfully override information that was initially encoded as true has proven to be quite a feat (c.f., Vraga et al., 2022) for researchers and policymakers alike (Larsen, 2014; Levine, 2020; Lewandowsky et al., 2022; Marsh & Yang, 2018). This phenomenon describes the *continued influence effect*, the finding that originally encoded information frequently and persistently affects people's later

reasoning and memory judgments (Johnson & Siefert, 1994). In the case of continued influence effects for misinformation (i.e., *illusory truth effects*—the focus of the present research), this robust and well-replicated effect continues to inform cognitions aligned with the original (false) information, despite retractions or warnings indicating that the original information should be disregarded (e.g., Gilbert et al., 1990; 1993; Pennycook et al., 2018).

The *fluency heuristic* refers to the relative ease of processing or recollection of information which elicits strong feelings of familiarity and influences our memory experiences (Jacoby & Dallas, 1981; Whittlesea, 1993; Whittlesea & Leboe, 2000). The *fluency-misattribution paradigm* posits that when we process information that is cohesive and resonates semantically, the resulting feelings of familiarity generate internal signals that are ambiguous and are thus favorably attributed to a variety of sources, (Reber et al., 1998; Trippas et al., 2016). Thompson and Morsanyi (2012) refer to this phenomenon as a "feeling of rightness" (FOR); people are inclined to favor the positive affect that arises from processing ease, which can sometimes lead them astray. A relatable example is the experience of bringing an incorrect name to mind when encountering a specific person. If you have ever had this experience, it can be quite odd—on the one hand you have an intuitive sense that the name you want to use is incorrect, but why did it *feel* so right? Possibly you were incorrectly introduced initially, or perhaps the person reminds you of someone else you know who has this incorrect name. If you have had this experience, you may have also noticed it often takes repeated exposure to the *correct* name to prevent this memory error from occurring.

Earlier studies have found a range of outcomes to be attributable to fluent processing. Earlier work conducted by Whittlesea (1993) found that people processed sentences that were completed by a semantically predictive (i.e., *fluent*) word (e.g., "the dentist cured the tooth…

ACHE") faster and judged these sentences as more pleasant than sentences that were completed by semantically non-predictive words (e.g., "the soft pillow under her head was … PINK"; Whittlesea, 1993, p. 1252). In another study, Trippas and colleagues (2016) demonstrated that manipulating the logical structure of arguments (i.e., logically valid versus logically invalid) influenced participant's perception of the presented physical brightness of the arguments themselves; the greater the perceptual fluency of the argument (i.e., the more logically sound the argument was), the brighter and more 'likeable' participants judged it to be.

Fluency is a powerful vector in which people intuitively place strong reliance when forming their judgments, and it influences aspects of daily life all the time. People judge easy-to pronounce rides and food additives as safer (Song & Schwarz, 2009), stocks as more profitable (Alter & Oppenheimer, 2006; Green & Jame, 2013), and people with easier to pronounce names are judged as more likeable, are preferred in hiring decisions and election campaigns (Bertrand & Mullainathan, 2004; Laham et al., 2012), and their claims are judged as more truthful (Newman et al., 2014). This inflated sense of confidence is proposed to arise from the experience of arriving at a solution with relative ease (e.g., reading and pronouncing a name with minimal cognitive effort), and not necessarily from the content of the response itself (Thompson & Morsanyi, 2012). As a result of this heightened fluency, people also routinely express high confidence in false memories or inaccurate information. Thus, fluency misattribution can lead people to misremember previously encountered (and fluently processed) false information as more truthful or accurate than new information, a phenomenon referred to as the *illusory truth effect* (De Keersmaecker et al., 2020; Hasher et al., 1977; Henderson et al., 2021).

For a phenomenon with such a far-reaching array of implications, studies investigating bottom-up approaches such as fluency manipulations and their potential to alleviate certain

memory errors within the domain of misinformation are surprisingly limited. However, the idea of exposing people to cues indicating the veracity of an item of information while it is encoded has received recent attention and brought about attempts to develop fact-checking technologies capable of debunking false claims in real time (Bode & Vagra, 2015; Guo et al., 2022; Lanius et al., 2021). Importantly, the scientific literature at present is lacking empirical cognitive data demonstrating the effects of disfluency on whether people remember information as true or false. With the possibility that real-time fact checking may become a viable computational approach to dealing with misinformation in the future, it is critical that we have a thorough understanding of the effects that manipulations such as fluency exert on memory for the information we encounter. In response, the present study proposes that a *disruption* to the fluency with which people initially encode information might reduce reliance on explicitly identified false details in subsequent decision-making and memory tasks.

The Illusory Truth Effect

The imperfect nature of memory makes people especially vulnerable to inaccurate memory experiences. Researcher Elizabeth Loftus dedicated much of her career to uncovering evidence demonstrating the susceptibility of memory to the influence of misinformation (Loftus, 2005; Loftus et al., 1978; Loftus & Palmer, 1974). Since her debut work demonstrating this extremely common occurrence, herself and many others have well-established the robust *misinformation effect*, the finding that people will insert new information (newly encountered details suggested in a question, for example) into previous autobiographical memory accounts (Loftus, 2005: Frenda et al., 2011). The present research however is concerned with the closely related *illusory truth effect*—human memory can be distorted significantly as a direct result of repeated exposure to false or misleading details. That is, the pronounced fluency that is

experienced when previously encountered information is reintroduced is mistakenly ascribed to its accuracy (Henderson et al., 2021). This effect has been demonstrated for people's recollection of words (Deese, 1959; Roediger & McDermott, 1995), faces, (Shimae et al., 2020), news headlines (Pennycook et al., 2018), and autobiographical events (Conway, 1995; Ost et al., 2002; Pezdek, 2003; Read, 1996; Schmidt, 2004).

The traditional experimental paradigm used to test for illusory truth effects involves participants encoding information (e.g., from written scripts, photographs, video clips, etc.) and subsequently subjecting them to deliberate exposure of information that differs from the original material on one or several descriptive points. Finally, in a test phase, participants are required to report or recognize details pertaining to the original information (e.g., via questionnaires, memory tests). As a result, people inadvertently include elements from the interjected misinformation into the memory from the original source (Frenda et al., 2011; Gilbert et al., 1993; Loftus et al, 2005). Similarly, in other experiments, participants are first exposed to various ambiguous claims and then later asked to assign truth ratings to previously seen statements versus new statements. In these experiments, participants consistently judge the previously encountered details as more truthful than new details, even when provided with explicit instruction to disregard these details as false (Gilbert et al., 1993; Jalbert et al., 2020; Pennycook et al., 2018).

A major component to assigning judgments of truth—and a central argument of this manuscript—is that the experience one has processing information and whether a statement *feels* familiar are associated with how accurately it will be judged. Having had previous exposure to a piece of information can create the illusion that the information is valid when it is next encountered, functioning to increase a person's belief in it. For instance, after first hearing the

statement, *"The temperature of a hen's body is 104º F"*, you may not immediately believe it to be true. However, upon hearing or reading this information again, it begins to gain convergent validity—separate sources spouting the same claim lends credence to the possibility that there is truth to the message (Fazio et al., 2015; Lewandowsky et al., 2012, Pennycook et al., 2018).

As the following discussion illustrates, numerous cognitive factors compound the issue of illusory truth effects and render the beliefs stemming from these falsehoods as notoriously "sticky" and difficult to correct (Scholl et al., 2014; Swire & Ecker, 2018; Lewandowsky et al., 2012, *p.*107). There are various cognitive mechanisms found to contribute to the likelihood that previously encountered information may be misremembered as true, including time (Knowlton & Squire, 1995), repeated exposure (Pillai & Fazio, 2021, Swire et al., 2017), specific cues (Pennycook et al., 2018), cognitive resources available (Loftus, 2005), emotion (Forgas et al., 2005; Koch & Forgas, 2012), stress (Morgan et al., 2013), and various individual differences (Pennycook et al., 2016; 2020; Zhu et al., 2010). Importantly, as this study attempts to illustrate, these factors cannot account for the significant and lasting effects that fluency exerts on people's assessments of the truth.

Factors Known to Contribute to Illusory Truth Effects

Recent decades have produced a range of findings pertaining to the contexts and outcomes of illusory truth effects (see Henderson et al., 2022 for a recent review). Considering the scope of misinformed content, the possible influences and factors relevant to this issue are practically limitless. Nevertheless, there are several themes which have emerged as consistent contributors of this effect: time, repetition, specific cues, and individual differences. These are precisely the factors involved in the present study, though my intention is to highlight the overarching influence that fluency exerts in this effect by testing whether a *disruption* to the

fluency of false claims (e.g., ThE rObBeR hAd A gUn") might alleviate some reliance on these details in later tests of judgment and memory. Furthermore, I also investigated whether two separate personality constructs that are relatively well-known for predicting prejudice as well as a greater affinity for negatively-valenced stimuli (Carraro et al., 2011; Dodd et al., 2012), and for having specific views on equality and punishment (Altemeyer, 2004; Capps, 2002) might also have some bearing on the information that is ultimately remembered as true.

Repetition and Time

In alliance with fluency, repetition is considered the quintessential driver of illusory truth effects, primarily since— as the literature currently stands—repetition is inherently necessary to measure illusory truth effects (Hintzman, 2010; Jalbert et al., 2020). If one is to 'remember' whether a detail is false or true, this is generally based on the assumption that the person has encountered some version of the tested content on a previous occasion. If a statement provides resonance for the reader, sparking a feeling that they have read or seen something similar at an earlier time, the reader will be more likely to judge it as true (Fazio et al., 2019; Henderson et al., 2021). These findings coupled with the intrinsic nature of the internet and social media platforms (where many tend to find and consume their 'news'; Gottfried & Shearer, 2016), make it the perfect breeding ground for misinformation and its repetition— false claims can be repeated thousands of times with only a few 'shares' (Bakshy et al., 2012). Additionally, people who seek their 'news' from social media tend to engage with headlines rather than actual content (Ecker et al., 2014)—articles and stories which have been funnelled through the lenses of social media contacts. This combination of factors presents a near perfect opportunity for *confirmation bias* to ensue – people demonstrate a strong preference for sources whose message confirms their own

worldview because of the natural inclination to attend to information that supports one's beliefs and reject claims that are incongruent to their position (Evans et al., 2005; Zhao et al., 2020).

Warnings and Specific Cues

Warnings, retractions, and other cues alerting that a message might be or *is* false are only modestly effective in diminishing the challenge of illusory truth effects. Pennycook et al., (2018) presented participants with fake and real news headlines in a format resembling an article one might typically see in a Facebook 'news' feed. For half of the sample, the false headlines were accompanied by a strong warning label: "⚠ ***Disputed by 3rd Party Fact-Checkers!".*** Despite these distinct warnings, having previously seen the headlines dramatically increased the truth ratings of false headlines. In fact participants rated previously seen false headlines as *more* accurate than new headlines which hadn't been accompanied by any 'fake news' warning.

In relation to illusory truth effects, a recently published study demonstrates the effects of the above mentioned factors nicely and is of a similar nature to the present study. Jalbert and colleagues (2020) were interested in understanding the extent that providing warnings to participants regarding the presence of false information might reduce illusory truth effects. These researchers included delays of three to six days between the initial exposure to false information and a test phase, to simulate the approximate delay one might experience naturally before re-encountering some morsel of information. These researchers utilized rather ambiguous trivia claims as stimuli such as, "*Walruses use their tusks primarily for mating"* (true), or *"The mouth of a sea urchin is on top"* (false) (Jalbert et al., 2020; p. 605) and had participants read the various trivia claims as they were presented individually on a computer screen. Prior to viewing the statements, half of participants were warned that half of the statements about to be presented were false. The test phase involved presenting another set of trivia statements and all

participants were told that half of those statements were items they had seen previously, while half of them were new, though no participants were provided any information regarding the veracity of the trivia statements at test. All participants were asked to indicate the veracity of each claim, using a 6-pt Likert scale (i.e., 1 = *definitely true,* 6 = *definitely false)*.

Results of this study indicated that participants who received the warning judged statements as less true than unwarned participants, though the veracity judgments for new statements did not differ between warned and unwarned groups. Participants judged the new statements as true approximately 50% of the time, though truth assessments for repeated claims rose to 70% for unwarned participants, while the assessments of repeated statements of warned participants was 60%, representing a 50% reduction in truth ratings for repeated statements when a pre-exposure warning was provided. A follow-up experiment found that receiving a pre-test (as opposed to a pre-exposure) warning was as ineffective at reducing the illusory truth effect as receiving no warning at all (Jalbert et al., 2020)

Individual Differences

The range of individual differences that have been studied with respect to affinity for misinformation is extensive and includes age (Brashier & Marsh, 2020), emotion (Forgas et al., 2005), cognitive style (Pennycook et al., 2020), religious beliefs (Pennycook et al., 2016), political affiliation (Carraro et al., 2011), and various personality constructs (De keersmaecker et al., 2020). As noted earlier, two constructs that may have some bearing on illusory truth effects but have received very little empirical attention in this regard are right-wing authoritarianism (RWA; Altemeyer 1981; 1998; 2004; Zakrisson, 2005) and social dominance orientation (SDO; Pratto et al., 1994), both of which are shown to uniquely predict prejudice and attitudes toward social justice.

Right-Wing Authoritarianism. Right-wing authoritarianism is described as the personality construct most predictive of prejudice (Altemeyer, 2004). Generally, those who score high on the RWA scale (Altemeyer, 1981; 2004; Zakrisson, 2005) tend to be i) relatively submissive to authority (or those whom they deem to be an 'authority'), ii) aggressive, particularly when aggression is condoned by said 'authority', and iii) highly conventional, generally set in their 'traditional' values and are often overtly religious (Altemeyer, 2004). As Yelderman & Miller (2017) illustrate, religious fundamentalism is one of the most well-documented characteristics related to legal decisions. For example, those high in religious fundamentalism demonstrate strong support for the death penalty and endorse tougher sentencing than more liberal-leaning thinkers (Capps, 2002). Thus, given that the experiments described next involve participants ascribing character judgements and prison terms to suspected criminals, it follows that RWA should have some bearing on the information one clings to when facing decisions of punishment.

In addition, recent studies have found that individuals high in RWA were much less able to update their original belief in popular myths (e.g., *"deoxygenated blood in your veins is blue"* (p. 1350) following correction as compared to those with a more open-minded thinking style and those scoring low in RWA, and that this was particularly true if the belief to be updated was a myth that the participant had held strong belief in originally (Sinclair et al., 2020; Voss et al., 1991). Individuals with high RWA scores are also found to have a higher tolerance for stereotypical beliefs and hate-speech (Bilewicz, et al., 2017; Strube & Rahimi, 2005), and have a stronger affinity to aversive stimuli compared to those who identify with more liberal or democratic views (Carraro, et al., 2011; Dodd et al., 2012). With these considerations in mind, my expectation is that those scoring higher on the RWA scale (Zariksson, 2005) will also have

the highest tendency to misremember false information as truthful, and to assign longer prison sentences as compared to their lower-scoring RWA counterparts.

Social Dominance Orientation. Comparatively speaking, social dominance orientation (SDO) is described as a strong preference for unfair social systems, and a belief that equality is not a worthy aspiration for society, generally because people high in SDO have a lot to gain by a hierarchical social structure in which they are placed near the top. Those high in SDO tend to promote "*legitimizing myths*" (Pratto et al., 1994, p. 741) that function to stabilize inequality and oppression, engaging strong assertions toward how things of societal value (e.g, jobs, real estate, wealth, positions of power, legal persecution) should be allocated (Pratto et al., 1994). Those high in SDO encompass several personality traits that are generally viewed quite negatively in society including the tendency to act with minimal moral restraint, an extreme drive for personal power, and meanness (Altemeyer, 2004; Friedman, 2018). These individuals tend to disagree with statements such as, "*It is better to be loved than to be feared"*, while agreeing with statements like, "*It is a mistake to interfere with the law of the jungle; some people are meant to dominate others"* (Altemeyer, 2004, p. 425).

The relationship between RWA and SDO has been studied at length, and correlations between the constructs are found to be strongest among citizens of countries marked by strong ideological divides (Roccato & Ricolfi, 2005). Altemeyer (2004) makes an important distinction between the two personality constructs in terms of their prejudiced ideologies; those high in RWA demonstrate a primary motivation of fear and self-righteousness, whereas those high in SDO are driven by their vendetta for personal power and to dominate others. In addition, those high in SDO tend to have awareness of their prejudiced beliefs towards others and will justify their message in a variety of ways, while those high in RWA are more typically lacking the

insight necessary to recognize their own prejudiced attitudes (Altemeyer, 2004). Interestingly, and perhaps unfortunately, those high in RWA and those high in SDO often complement one another; one group is eager to lead and dominate with aggression, while the other is quite comfortable following the charge and being 'led' by the those they believe to be powerful. Perhaps unsurprisingly, Son Hing et al. (2007) found that pairing people high in both constructs resulted in more unethical decisions as compared to dyads with low scores for these constructs. Therefore, my expectation for those high in SDO is similar to my prediction for those scoring high in RWA—I expect to find a positive relationship between SDO score and proportion of false information misremembered as true, and potentially longer prison term assignments from these participants as a result.

Effects of Disfluency

Fluency is a clear and complex obstacle to preventing belief in misinformation. Therefore, this study proposes that ruining the fluency of false information when it is first experienced, or during a repetition, might be one strategy to curtail people from encoding explicitly identified false information as true. Tests using disfluency to counteract an effect that largely arises from fluency are limited, and generally involve adjusting the physical brightness, font size, or color of the information marked as 'disfluent' (e.g., Meyer et al. 2015; Motyka et al., 2016). The disfluency manipulation implemented in these experiments (i.e., AlTeRnAtInG cAsEs), has been investigated to a much lesser extent. While saturation, contrast and font adjustments may very well present less fluently than the typical sans serif font, having participants encode false information in alternating cases may do a better job at disrupting fluency at its core—whether you read false information in an *unfamiliar text* or read letters and words that are more s p a c e d o u t than you might typically expect to encounter, presenting

false information in a cOmPlEtElY uNuSuAl WaY challenges fluency differently; in a manner that directly affects the extent of processing at the time of encoding.

The literature concerning the advantages and disadvantages of disfluency are mixed, and research concerning any *positive* effects of disfluency is a fertile ground of unanswered questions (Dunlosky & Mueller, 2016). Practically speaking, disfluency can certainly present a disadvantage—a blurry advertisement or menu with words that are barely readable will be off-putting to most. In terms of the present study, on one hand people demonstrate a clear preference for fluent stimuli, and tend to intuitively respond to disfluently presented stimuli with a negative bias (Lee & Labroo, 2004; Motyka et al., 2016). Alternatively, previous research in the decision-making literature reports that presenting test items in a difficult-to-read (i.e., disfluent) font provides a cue to process the information more deeply, resulting in higher scores for university student participants prompted to answer disfluently presented test questions as compared to test-takers who answered fluently presented questions (i.e., "desirable difficulties"*;* Bjork & Bjork, 2011). In recent years, several of these findings have come under scrutiny, prompting accusations of attempting to find specific materials (i.e., font variations, degree of blurriness, or presenting materials on computer screens versus on paper; Sidi, et al., 2016) to demonstrate the effect, rather than focusing on understanding how disfluency as a specific cognitive mechanism influences the manner that we process and retain information. Though replication studies have found differences in response times between fluent and disfluent encoders, they find little if any effect of improved analytical reasoning due to reading information with less-than-ideal clarity (Dunlosky & Mueller, 2016; Meyer et al., 2015). Importantly, any recent critiques of the beneficial cognitive effects of disfluency acknowledge that disfluency likely holds more influence in tests of memory than of reasoning or decision-making tasks (Meyer et al., 2015).

Various outcomes of perceptual disfluency in memory tasks have been identified, but unfortunately do not offer much more clarity on the issue of how disfluency might impact the preservation of false details in memory. For instance, Sanchez & Naylor (2018) found that disfluent presentation of words significantly boosted recall for unpresented words but made no difference in recall for presented words (Deese, 1959; Roediger-McDermott, 1995), and suggest that disfluency could in fact *aid* false memory creation. Others agree and suggest that disfluency can provide considerable benefits to memory performance and learning (Bjork & Bjork, 2011), particularly if participants are aware of an upcoming test (Geller & Peterson, 2021). On the contrary, Yue et al., (2013) found that manipulating fluency by blurring the stimuli resulted in worse memory performance and a clear memory deficit compared to those who encoded the stimuli fluently.

Given that the type of disfluency manipulation used in the present study is yet to be tested in this type of experiment, there is evidence to support the hypothesis that such a manipulation could make the false information that is presented disfluently more distinct and memorable than regular looking black text, and thus more likely to be used in decisions and incorrectly reported as true in tests of memory. Alternatively, this manipulation may make this information more memorable in such a way that it reminds participants to disregard it—remembering it was presented in an unusual format may be enough to *reduce* participants' reliance on it, and thus their likelihood of reporting that a false detail was presented as true. Finally, participants may disregard the disfluent text altogether as it is more difficult to read than fluently presented information, particularly if the participant was assigned a concurrent interruption task. Whatever the outcome, the aim of this study is to provide initial evidence of support or non-support for this

mode of disfluency as a mechanism to attenuate belief in false claims, and hopefully toward some future consensus on the effects of disfluency on recall for information that is not true.

The Present Study

This research is an attempt to replicate, and expand upon, an earlier study conducted in 1993 by Daniel Gilbert and colleagues who gathered some of the first evidence to suggest that experiencing an interruption at the time that declared false information is encoded not only leaves people in their initial state of belief when this information is re-encountered, but people will rely on this information as criteria in future judgments, including judgments of high consequence (i.e., the sentencing of a defendant). As these authors note, remembering that a novel phrase was labelled as 'false' does not harness the essence of that which they believe to be true; remembering that certain information had been corrected appears to be a separate endeavor from believing and basing decisions or actions upon this information in the future (Gilbert et al., 1993; Spinoza, 1982).

To summarize, Gilbert et al. (1993) had participants read aloud two separate crime reports—one report accusing Tom, and one report accusing Kevin—each of which contained false details. Included in the general instructions for all participants was the warning that any details presented in red text (as compared to standard, black text) were false and were to be struck from the record in their final assessments of the purported criminals, which included assigning likeability, danger, and potential for rehabilitation ratings, along with a final prison sentence (0 – 20 years) for each crime. There were two versions of each crime report for each of the suspects—in one, the false details exacerbated the reported crime, while in the other the false details extenuated the crime. Each participant read one exacerbating report, and one extenuating report for each of Tom and Kevin's crimes. In addition, half of the participants were assigned an

'interruption task'. For this group of participants, the false text was accompanied by a string of blue numeral digits scrawling beneath the text, and participants were tasked with indicating whenever they identified a '5' cross the screen. In the uninterrupted condition, there was no concurrent task assigned. Following the presentation of both crime reports, all participants completed two recognition memory tests where they were asked to respond TRUE, FALSE, or NEVER to various gist statements about the reports they had just read.

Gilbert and colleagues (1993) report that interrupted participants recommended prison terms that were significantly longer than the prison terms assigned by the uninterrupted participants—when the false details exacerbated the crime. In addition, the recognition memory test results indicated that interrupted participants were more likely than their uninterrupted counterparts to mis-identify false sentences and foils (details that were never presented) as originally true statements, though these groups were equally likely to correctly remember true statements as true. Despite explicit warnings provided to all participants, the criminal for whom the details exacerbated the reported crime was consistently judged as less likeable, more dangerous, and less likely to be helped by counselling, while also receiving the longer prison sentence. This response pattern was reversed for the suspect whose associated crime details extenuated their crime. Gilbert et al. argue that the distraction experienced made it more difficult for participants to remember which details were presented in red text, leaving these details free to influence character assessments and prison sentencing for each case.

Though Gilbert et al. (1993) do not mention the role of fluency in their work, it is clear that the fluency with which the content of the false statements is experienced is superseding any memory participants may have about the statement that was presented in red text. After all, we read information printed in red all the time, and likely just as effortlessly as reading text

presented in black text. Therefore, there are several aims of the present study. Across two experiments, I will attempt to replicate the findings of Gilbert, but with four additional goals in mind. First, since Gilbert warned each of his participants about the presence of false details in the reports, it is difficult to gauge a true baseline for how memorable (or not) the false details are when participants are provided no other warning instruction or cues, except for the different-looking text scattered throughout the report. Additionally, this allows an opportunity to replicate the findings of Jalbert et al., (2020) discussed earlier who found that providing warnings prior to exposure of false information can greatly reduce later truth ratings for that information. According to these results, I expect to find that providing a warning will influence participants' decisions to some extent, likely reducing the harshness of assessments towards one or both of the perpetrators on one or all of the character assessment variables. My second goal is to test an untested disfluency manipulation for false content with the prediction that the negative attribution generated when disfluent details are presented might encourage people to later discount these ideas in their judgments of the crimes. Third, as mentioned, the context of the study (i.e., reading and judging crime report details) presents a unique opportunity to investigate RWA and SDO as personality constructs with potential moderating roles for illusory truth effects. This experimental method also avoids presentation of stimuli with political connotations and side-steps potential confounding issues like previous knowledge that arise in memory tests for ambiguous statements. Lastly, in Experiment 2 I test whether *repeating* false statements either in the same regular black text (*fluent is false*) or in alternating cases (*disfluent is false*) might curtail reliance on these details, or, as previous research has consistently shown, whether a repetition of this nature only makes such details significantly *more* memorable and relied upon compared to instances where the false detail was only presented once.

Experiment 1

Method

Experiment 1 builds upon Gilbert et al.'s (1993) earlier work by adding a novel disfluency condition to test whether embedding the construct of disfluency within a message itself might interfere with its encoding, rendering the falsely described details as less likely to later be remembered as true than when these same false statements were presented in a fluent, red text. As in Gilbert's study, each participant was asked to assume the role of a judge and told they will later be tasked with assigning a prison sentence to each of two independent suspects based on the details they read for each suspect's crime. In Gilbert's original study, all participants were informed of the presence of false information in the reports, whereas in the present study half of the participants were assigned to a *no warning* condition and were never told that some of the information presented in each of the reports was false and was to be excluded from consideration when evaluating appropriate punishments and attitudes toward the suspects. In this condition, participants were left to assume that all information included in the reports was accurate. In addition, for half of the participants, false details were presented in red text (as in Gilbert et al., 1993) representing the *fluent* condition while the remaining participants were presented with false details in a *disfluent* alternating-case format (e.g., hIs aFfAiRs WiTh OtHeR wOmeN"). Similar to Gilbert's work, half of the participants experienced a concurrent distraction task that involved monitoring whether the number '50' repeated as changing double-digit numbers appeared on the screen below the text statements of the crime report as it was read. Finally, for each participant, the false information *exacerbated* the crime in one report (if, of course, the details were remembered and considered to be true), while the false information *extenuated* the circumstances of the crime in the alternate report.

Participants

Five-hundred and five participants (299 females, 199 males, 7 'other', M_{age} = 19.58) were recruited through the Department of Psychology Online Research Participation System at the University of Manitoba (Winnipeg, Canada) and received partial credit in their Introductory Psychology course (PSYC-1200) in exchange for their participation. The recruitment message specified that participants must be fluent in English, that they must have normal or corrected-to-normal vision including color vision, and that they must complete the study in one sitting at a desktop or laptop computer. Participants were never informed that they were participating in a study of misinformation, but instead were told that this was a study of memory for details of crime reports. Prior to their participation, every participant signed an online consent form as required for approval by the University of Manitoba Fort Garry Research Ethics Board (see Appendix A). Following completion of the study, 23 participants (11 female, 12 male, M_{age} = 19.17) were excluded for incomplete data— only the data of participants who completed the experiment in its entirety were included in the data analysis. Upon completion of data collection, the data exclusion criteria outlined in the Results section removed an additional 21 participants (12 females, 9 males, M_{age} = 19.86) from analysis. The final sample for Experiment 1 included 461 participants (276 female, 178 male, 7 other, M_{age} = 19.99, SD = 3.82, *range* = 17 – 49 years). This sample size was comparable to that of Gilbert et al., (1993) and allowed for approximately 15 responses in each cell of the design, though notably Gilbert tested only females.

Design

Two-hundred and thirty-three participants were randomly assigned to one of sixteen *fluent* conditions in which all false details within each crime report were presented in bright red text, as Gilbert and colleagues (1993) also did. The remaining two-hundred and twenty-eight

participants were randomly assigned to one of sixteen *disfluent* conditions, wherein all false statements for each report were presented in the alternating-case format (e.g., "…where he had BeEn DrInKiNg BeEr AnD sCoTcH most of the evening"). The order that each suspect's crime was presented (*Tom then Kevin* versus *Kevin then Tom*), the severity of the false statements embedded within each crime report *(better* versus *worse,* for simplicity), the presence of the instruction to disregard any red or alternating-case text format as false (*no warning* versus *warning*) and whether the participant experienced the interruption task while reading the details of the crimes (*no interrupt* versus *interrupt*) was counterbalanced across all conditions. This resulted in a complex mixed design, treating the order of the reports, the format of the false statements, whether participants were explicitly made aware of false information and whether participants experienced a simultaneous interruption task as between-subjects variables, while the severity of the false information embedded in the reports and the suspect represented by exacerbating versus extenuating details were included as a within-subjects variables. That is, if a participant first viewed Tom's exacerbating report, the second report they viewed was Kevin's extenuating report. This is an improvement to the design implemented by Gilbert et al. (1993) as in those studies all participants were privy to the presence of misinformation, making it difficult to determine the extent to which this differently encoded information would be remembered as true without any explicit indication that any details were to be disregarded as false. In addition, while there is no theoretical reason to expect a difference in length of prison sentence assigned based on the order that the reports were presented, Gilbert and colleagues always presented *Tom*'s crime report first and therefore these changes make it possible to formally test Gilbert's original assumptions.

Materials

All materials related to the crime reports, character assessments and sentencing measures, as well as the recognition memory tests were obtained through Daniel Gilbert's professional website (Gilbert, n.d.). Due to the ongoing COVID-19 pandemic at the time this study was conducted, all participants participated in this study online at a personal desktop or laptop computer (the experiment was not mobile friendly). The experiment was designed in such a way that all responses were provided using a mouse or trackpad cursor to either click a button icon, or to move a slider bar. When the participant was required to enter their age or assign a prison sentence, they entered these responses using the number pad on their keyboard. The experiment was conducted through Qualtrics (Qualtrics, Provo, UT, 2005) online platform.

Independent Variables

Crime Reports

The two crime reports described separate crimes— one crime was said to be committed by the perpetrator Tom, who robbed a stranger, and the other crime committed by the suspect Kevin, who robbed a convenience store. The reports used in this study were borrowed from Gilbert's original materials and kept virtually identical but for a few, minor adjustments. First, the location and date of the crimes were changed from taking place in 1988 in Austin, TX to taking place in in 2019 in Winnipeg, Manitoba, Canada. The locations mentioned in the reports (e.g., highway I-35) were also adjusted to reflect locations relevant to Winnipeg (e.g., Portage Avenue). Since the original experiments were conducted in the early 1990's on a sample from the geographical area of Austin, TX, this change was consistent with the relevant setting described in Gilbert's reports.

In addition, the number of sentences between the original reports for Tom and Kevin differed; the report concerning Kevin's crime included 11 short-to medium length sentences,

while Tom's report consisted of 14 sentences which included two very short sentences (e.g., "Larry complied."). Given that Gilbert had participants read the reports aloud as the words 'crawled' across the screen (at approximately 16 characters per second; p. 224) and that, in terms of word count, the reports were roughly the same length, these minor differences between reports are insignificant to Gilbert's study. However, since participants in the current study were to read the reports on the screen one sentence at a time, it was necessary to ensure that the number of sentences and average sentence length were equal across both crime reports. This was achieved by making a few minor adjustments to the sentences in Tom's report while leaving the overall integrity of the story intact. All original crime report details for Tom and Kevin's crimes are found in Appendices B and C, respectively.

First, the fourth and fifth sentences of the original report were altered so that the fourth sentence read, "Travelling down Main Street, Tom indicated that he wanted to get off in the North end.", and the fifth read, "Larry recognized that Tom was nervous, preoccupied, and strangely silent". Next, the eighth sentence, "Larry complied" was eliminated and the sentence that followed was changed to "Larry did as he was told, and with the valuables in his possession Tom began muttering that he thought *drug dealers were really despicable/crippled people were really disgusting*." Third, the original 10th and 11th sentences were combined to produce the following sentence: "Tom then ordered Larry to take an early exit and directed him through a residential area, forcing him to stop at a deserted corner." Finally, the original 13th sentence was combined with the 12th to read, "As Tom turned to get out of the car, Larry began pleading with him to return his wedding ring— but Tom continued to exit the vehicle." These changes ensured that all crime reports read by participants included 11 sentences each and brought the average sentence length of Tom's report up from 17.5 to 21.5 words, equal to the average sentence length

in Kevin's report. The only statement in Kevin's report that was altered was sentence 11, where "as he drove away" was added at the very end to help the sentence sound more complete, and to equalize word counts between reports. None of the false details were altered from the original materials. The final word count for each of the four versions of crime reports varied only slightly; Tom false-extenuating, 249 words; Tom false-exacerbating, 241 words; Kevin false-extenuating, 241 words; and Kevin false-exacerbating, 244 words. The revised versions of the reports for Tom and Kevin's alleged crimes used in this study are found in Appendices D and E, respectively.

Exacerbating and Extenuating Conditions

For each suspect, in one crime report the false information exacerbated the reported crime, and in the other report the false information extenuated the crime. The only difference between the two reports for each suspect, for each version, was the content of the false information. That is, the details that were to be considered *true* were identical between extenuating and exacerbating versions of the report for each suspect. In addition, each version reported six sentences which included false information, and five sentences which did not contain any false information. When the false information is removed from the reports, the crimes are described as Tom having robbed a man at knifepoint, and Kevin having robbed a convenience store at gunpoint.

When the extenuating false information is believed for Tom's crime, the details include Tom robbing a well-known drug dealer who was off to join his friends at a strip club, and that Tom tossed the wedding ring he had stolen back in the car before leaving the scene. When the exacerbating false information was read for Tom's report, the details included that Tom had been drinking all night, had hitchhiked a ride with a disabled veteran before robbing him with a knife

to his throat, and uttered a discriminating slur before running away. The extenuating false information for Kevin's crime involved Kevin having had loud arguments with his wife over her overspending, and after decompressing with friends he allegedly robbed a convenience store at gunpoint "to cover his wife's debts". When the exacerbating false information was included for Kevin's crime, the participants read that Kevin had been arguing with his wife about his affairs with other women, assaulted his brother-in law, and then purchased and consume a large amount of cocaine with friends before robbing a convenience store. During the convenience store robbery, Kevin reportedly held the gun to the clerk's head and threatened to sexually assault her.

Disfluent Text Condition

The main objective of Experiment 1 was to determine whether a novel disfluency manipulation had any effect on the information that participants retained, and ultimately used, in their decisions concerning the suspects as evidenced by their responses to sentencing and character assessment measures described above, and by their performance on each of the two recognition memory tests. Similar to Gilbert's study, half of the participants were randomly assigned to the *fluent* condition (n = 233) in which all of the false information was presented in bright red, default 12-pt font provided within the rich content editor of the Qualtrics (Qualtrics, Provo, UT, 2005) platform. In the *disfluent* condition (n = 228), all information in the crime reports was presented in black, default 12-pt font, except that each word of each of the false details was presented in alternating cases. For example, the second sentence of Kevin's crime report appeared as: "During an argument that had erupted OvEr HiS wIfE's ExTrAvAgAnT sPeNdInG, Kevin had yelled at his wife several times aNd HaD tHrEaTeNeD tO dIvOrCe HeR." Following this pattern, the next word of the next false detail of the report would begin with a lowercase letter. This allowed all possibilities of cases of letters to be viewed by participants.

That is, half of the participants in the *disfluent* condition saw the opposite presentation of the alternating cases for false details in the crime reports (i.e., "oVeR hIs WiFe'S eXtRaVaGaNt SpEnDiNg").

Warning Condition

In Experiment 1 half of all participants were randomly assigned to a *no warning* condition (n = 236) in which participants were not provided with any prior indication that some of the crime report details would be false. This manipulation made it possible to better understand the extent to which different-looking content within a statement (i.e., either *fluent* or *disfluent*) might influence judgements when participants were not explicitly made aware that this information should be treated differently from the details appearing in the predominant black-text without alternating cases. The remaining participants were randomly assigned to the *warning* condition (n = 225) in which all participants received the same initial warning instruction as Gilbert's original participants. The specifics of the instructions provided based on each experimental condition are found below in the "Conditional Instructions" section (p. 29).

Interruption Condition

In the *interrupt* condition (n = 231), participants were tasked with attending to a series of two-digit integers which were presented centered, directly beneath each detail of the crime and changed once each second for the entirety of the 12 s presentation of each sentence. The participant's task was to monitor whether the number '50' was repeated while simultaneously reading the sentence detail of the crime report presented directly above the changing numbers. All two-digit integers were presented in bold, in light blue font color using 12-point default font. Since each crime report consisted of eleven sentences, 22 unique sets of 2-digit integers ranging from 10 to 99 were generated using a random integer generator (www. random.org). These

integer sequences were then manipulated to ensure that the digit '50' was repeated in eleven of the sequences and was not repeated in the remaining eleven sequences. Six of these 'no repetition' sequences included the presentation of the digit '50' one time while the remaining five sequences did not present the digit '50' at any point. Across all sequences, no digit other than '50' was ever repeated. In any instance where the random integer generator produced a sequence with a repeating integer other than '50', the second instance of that digit was changed to that digit plus one. In addition, the sequences wherein the digit '50' repeated were created while minding several constraints. First, the integer '50' was only ever repeated once in the sequence and was never presented as the first or last of the twelve integers in any sequence, which ensured that the switch to the next statement of the report did not interfere with participants' ability to perform the interruption task. In addition, the first presentation of the '50' integer was always within the third to eighth position, and there was always at least one other digit separating the first and second presentation of the integer. In four of the repetition sequences, the second presentation of '50' took the second last position (the 11th presented integer), in another four sequences the repeated '50' was presented in the third last position (the 10th integer), and the remaining three repeating sequences placed the second '50' in the fourth to last position (the 9th integer) of each respective 12-integer sequence. These constraints were included primarily to simplify the task and ensure that participants' attention was also directed to the sentence displayed.

Once these criteria were met for all 22 semi-randomized integer sequences, the sequences themselves were randomized into sixteen unique organizations and then randomly assigned to each of the sixteen counterbalances that included the repeating-integer interruption task. In any condition that included an interruption task, there were at least four sentences for which the

correct response to the question, “Was the digit ‘50’ repeated?”, was ‘YES’, and never fewer than four sentences for which the correct response was ‘NO’. In cases where the participant experienced the interruption, each of the twelve 2-digit integers were presented centered and immediately beneath the text for 1 second before changing to the next integer in the sequence while the report text remained static. Immediately following the presentation of the twelfth integer in the sequence for each sentence, both the statement and the integer disappeared from the screen and were replaced by the question, “Was the digit ’50’ repeated?”. Using their mouse cursor, participants had to select a bubble icon beside either a ‘Yes’ or ‘No’ response before proceeding to the next sentence (and a new sequence of interrupting integers) of the crime report.

This interruption task was conducted slightly differently from Gilbert’s original study. First, for those in Gilbert’s interruption condition, participants were asked to indicate whenever they witnessed the number 5 appear in a string of numbers scrolling beneath the text of the report. In this study, the experimenters had participants seated at private cubicles with computer screens and devices described as ‘counters’ which were seemingly connected to another computer present. All participants were instructed to press the button on the counter when they saw the digit ‘5’ appear beneath the text they were reading and were told that they would be recording their accuracy throughout the experiment. After all participants had completed a practice phase of the interruption task, half of the participants were told that they were part of the control condition, their counters were removed, and they were instructed not to pay any attention to the digits scrolling beneath the (also scrolling) details of the reports, throughout the experiment. This characteristic of Gilbert’s study makes it difficult to assert that those in Gilbert’s *no interrupt* condition were truly undisturbed by the scrolling digits. It seems likely that digits appearing concurrently with information that is meant to be retained would provide some

degree of distraction, whether the participant was instructed to pay attention to it or not. In addition, Gilbert only presented digits beneath the false information, and never beneath any details that should have correctly been recalled as true. This presents the problem of differences in degree of distraction between true and false items, perhaps providing an unfair memory advantage for the true items. To address these concerns, participants in this study who did not experience an interruption task were not shown any digits concurrently with the sentences of the report. In addition, for those in the *interrupt* condition, the 2-digit integers changed consistently each second of the 12 s that the sentence was presented, distracting the participant while they read *both* the true and false information within the report.

Conditional Instructions

In every condition, **e**ach participant was presented with the following identical initial instructions, centered on their computer screen: "Today's task requires you to assume the role of a judge. You will be reading police reports of two separate crimes committed recently in the Winnipeg, Manitoba area. You will read through the reports one at a time."

Warning Condition Instruction

For those assigned to the *warning* condition an additional instructional paragraph was included here: "You will notice that some of the statements appear in red text (an odd-looking format). Any statement that appears in this red text (alternating case format - e.g., ThE rObBeR hAd A gUn) should be considered FALSE, as this detail is from an unrelated crime report and was accidentally included, as sometimes can occur in a court of law. <u>You should not include these false statements in your assessment of the suspect, or the sentencing of the crime.</u>" This information was not included for those participants assigned to the *no warning* condition. Participants in both the *warning* and *no warning* conditions continued to read the following

instructions: “Each of the two reports will appear only once, so you will have only one opportunity to learn what happened. **It is important that you read each statement presented, as there will be a memory test at the end of this experiment**. After you have read both reports, you will be asked to make judgments about the suspect accused of each crime, to assign a prison sentence (from 0-20 years) to each suspect, complete two personality measures, and your memory for the details of each crime will be tested.” The participant was then required to use their mouse cursor to select a button labelled ‘Click to Continue’ to proceed.

Interruption Condition Instruction

For those participants assigned to the *interrupt* condition, a final instruction appeared on the screen that read: “As you read through the crime reports, there is one additional task to keep in mind. As you read, you will notice that numeral digits in **BLUE FONT** [typed boldface and in the same color of the digits that were presented in the interruption condition] appear below the details of the crime. Your task is to read the details while paying attention to the digits below. After each detail, you will be asked whether the number ‘50’ was repeated. Once you indicate whether the integer was repeated, you will move on to the next detail of the crime report”. This instruction was omitted and there were no digits presented concurrently with the crime report details for those participants assigned to the *no interrupt* condition.

After participants were presented with their conditionally appropriate instructions, all participants were presented with the same review instruction: “Let’s review. You will soon be reading a series of two crime reports, trying to learn and remember as much as you can about the two criminals and their reported actions. **Remember that you must read each statement in order to perform well on the recognition memory test at the end of the experiment.** Finally, at the end of the experiment, you will complete two personality measures related to the concept

of equality. Once you have read and understood these instructions, click 'I Understand' to begin reading the crime reports". The participant then used their mouse cursor to click a button icon labelled, "I Understand, Let's Begin!" This action prompted the display of the first sentence of the first crime report to appear on the screen (along with the concurrent first 2-digit number in the series of interrupting digits, in those conditions).

Dependant Variables

Character Assessment Measures

As in Gilbert's original study, all participants responded to three questions about each of the suspected perpetrators. First, participants were reminded of the suspect they should refer to with the instruction: "Please indicate your perception of TOM (KEVIN), the suspect described in the FIRST crime report. For each question, drag the slider to the rating you feel is appropriate." Directly beneath this instruction was the first question; "What are your general feelings of likeability toward TOM (KEVIN)?" Beneath the question was a 7-pt slider anchored by neutral (0) and extreme dislike (7), with the slider always set to 3.5 (somewhat dislike). Once participants made their selection they were presented with the next question; "How dangerous do you perceive TOM (KEVIN) to be?" Another 7-pt slider was presented beneath the question and was anchored by slightly dangerous (0) to extremely dangerous (7), and the slider was always set to 3.5 (moderately dangerous). Finally, participants were asked, "Do you think TOM (KEVIN) would be helped if he received counselling?" The final 7-pt slider was anchored by helped a great deal (0) to not helped at all (7) and the slider was always set to 3.5 (helped a little).

Sentencing Measure

After completing the final attitudes measure, participants received the instruction; "A jury has found TOM (KEVIN), the suspect described in the FIRST crime report, guilty of his crime.

As the judge serving in this case, you must sentence TOM (KEVIN) to an appropriate prison sentence (0 – 20 years).” Below the instruction appeared the question; “How many years should TOM (KEVIN) serve? Please indicate your sentence by entering a number between 0 – 20 years.” Once the participant had entered their sentence in the text box provided for the first crime report, the same instruction and question was repeated for the second suspect. It should be noted that Gilbert had participants assign prison sentences for both suspects simultaneously and *before* responding to the likeability, danger, and counselling rating measures. The decision to change the order of questions was made to ensure that participants had considered these variables prior to assigning their sentence to each suspect, as would be the case in an actual judicial decision in a court of law. Due to the online nature in which this study was conducted, it was imperative to ensure that the instructions made it clear to participants which suspect and crime they should be referring to when responding to the questions described above. In addition to providing clear instructions, keeping the order of suspect presentation consistent for participants throughout the experiment was intended to help avoid any confusion and minimize errors.

Recognition Memory Tests

Following completion of the RWA and SDO measures (described below), participants entered the final phase of the experiment—the recognition memory tests phase. The recognition memory tests used in this study were identical to those used in Gilbert’s original experiment (see Appendix F). Again, each participant’s memory for the details of the reports was tested in the same order that the participants read the reports, and this was again made clear in the instructions they received. Each test consisted of 30 sentences and the participant’s task was to indicate whether they remembered the statement as a detail that; a) WAS REPORTED and is TRUE, b) WAS REPORTED and is FALSE or c) WAS NEVER REPORTED. Each sentence was

presented on the screen individually, and participants responded by clicking a bubble icon beside their choice and selecting 'CONTINUE' which prompted the next question of the test. As Gilbert notes, no item in either test was taken verbatim from either of the reports, but rather each statement conveyed the gist of a detail that had been reported (either as true or false) or a detail that was never mentioned in the associated report (e.g., "Kevin's wife was named Julie"). For participants that were warned of the presence of false information, there were 19 statements with details that had never been mentioned in the report *(Foils*), seven details that had been reported as false *(False*), and 4 details that had been reported as true *(True*). This was true for the test related to Tom's report and the test related to Kevin's report. Since participants in the *no warning* condition were given no reason to assume that any of the details provided in either report were untrue, 11 of the statements would correctly be identified as 'true', and 19 statements would correctly be identified as having 'never' been reported. Finally, since these participants were not warned, none of the details should have been identified and reported as 'false'. That is, any 'false' response in the *no warning* condition is considered an error response.

Personality Constructs as Moderating Variables

In this study, each participant completed two additional measures that were not included in Gilbert's research: the Right-Wing Authoritarian (RWA) scale (Altemeyer, 1998; Zakrisson, 2005, see Appendix G) and the Social Inequity (SI) scale (Altemeyer, 2004, p. 423-424; see Appendix H) which taps the construct of social dominance orientation (SDO; Pratto et al., 1994). These measures were administered in this order for all participants and were completed immediately following the participants' assignment of the prison sentence for the second suspect, and prior to the recognition memory test phase. The completion of these scales not only provided a convenient distractor task between the time of encoding the crime reports and participant's

exposure to the memory test statements, but more importantly their inclusion will provide an opportunity to explore whether scores for these personality constructs might serve as moderating variables that help to explain any relationship that is found between the proportion of false information that is ultimately remembered as true, and the length of prison sentences assigned to the suspects. These constructs were chosen because, taken together, they are reported to be rather robust predictors of prejudice, political orientation, attention, and behaviour (Altemeyer, 1998; 2004; Dodd, 2012; Pratto et al., 1994; Zakrisson, 2005).

Right-Wing Authoritarianism (RWA) Measure

Altemeyer (2004) describes people who score high on the RWA scale as willing to submit to those they deem established authorities, as prone to aggressive behaviour when they feel that resolve is condoned by said authorities, as conventional, and as highly prejudiced toward minorities. People who score in the upper quartiles on this trait also tend to be religious, conservative, and have a strong preference for conservative values and political parties.

The RWA scale used in this study was a shorter version of Altemeyer's (1998) original 30-item measure constructed by Zakrisson and colleagues (2005, see Appendix H). This 15-item measure was used because it excluded some potentially problematic items that were excessively long or that contained extreme wording or cultural connotations (e.g., "rotten apples"; Zakrisson, 2005). All individual statements tested in this measure were presented simultaneously on the screen, and participants were required to indicate their level of agreement to each statement on a 7-pt Likert scale[1] by selecting a bubble icon immediately to the right of the statement; *strongly*

[1] It should be noted that due to an unfortunate oversight in the Qualtrics program, those in the *disfluent* condition in Experiment 1 were only given a 5-pt scale, anchored by strongly disagree (1) and strongly agree (5). Prior to data analysis, all RWA scores in this condition were converted to scores on a 7-pt scale.

disagree (1), *disagree* (2), *somewhat disagree* (3), *neutral* (4), *somewhat agree* (5), *agree* (6), or *strongly agree* (7).

Social Dominance Orientation (SDO) Measure

In this study SDO was measured using Altemeyer's Social Inequality (SI) scale (2004, pp. 423; see Appendix I). Those scoring high on SDO are described as people who have a strong preference for social systems that benefit them at the expense of others and, who believe that equality is unrealistic and 'unnatural', asserting instead that some people should have to earn their way in society (even if they themselves have not). Those scoring high in SDO also tend to score high on scales of meanness, manipulation, and dishonesty. As Altemeyer (2004) alludes, including both measures in the current study is likely to be beneficial when exploring possible moderating effects as the RWA and SDO scales capture prejudices in people for different reasons with each construct uniquely explaining variance in prejudice scores while achieving a minimal correlation of approximately r = .20 -.45 between the two scales (Altemeyer, 2005; Strube & Rahimi, 2006).

The SI scale always followed the RWA scale and consisted of 22 statements presented in exactly the same format as the RWA scale, with each statement appearing on the screen simultaneously, and participants were required to select the bubble icon directly to the right of each statement that indicated their level of agreement; *strongly disagree* (1), *disagree* (2), *somewhat disagree* (3), *neither agree nor disagree* (4), *somewhat agree* (5), *agree* (6), or *strongly agree* (7).

Procedure

Prior to beginning the experiment, each participant read the letter of informed consent and selected the bubble icon "I consent" at the bottom of the page before proceeding (see

Appendix A). Participants then provided their age and preferred gender (female, male, or 'other'). Participants were then presented with instructions specific to their randomly assigned experimental condition, as discussed above, followed by the general review instruction. Once participants had finished reading the instructions, they indicated their understanding by selecting the bubble icon "I understand. Let's begin!". This prompted a new screen to appear with the first sentence of the first crime report presented centered on the screen. After twelve seconds, the sentence was replaced with the next sentence of the report, and this process continued through the eleventh and final sentence of the first crime report. For those in the *interrupt* condition, the 12 s presentation of each sentence was replaced with the question: "Was the digit '50' repeated?", and the participant was required to select the bubble icon associated with either a 'Yes' or 'No' response before proceeding to the next sentence of the report. Once the last sentence of the first report was displayed (or the participant had responded to the repetition question for the final sentence), all participants received instruction that they were about to begin reading the details of the second crime report: "This concludes the details of the crime report involving TOM (KEVIN). When you are ready, select 'Click to Continue' to begin reading the details of the crime involving KEVIN (TOM)". This action prompted the second report to be presented in the exact same manner as the first, one sentence at a time, each for a total of twelve seconds. Following the presentation of the last sentence of the second report (or, the participant's response to the repetition question for the final sentence of the report), all participants received the following instruction: "This concludes the details of the crime report involving KEVIN (TOM). Next you will be asked to make judgments about the two suspects and decide on an appropriate prison sentence. When ready, select 'Click to Continue".

Following the presentation of the second crime report, participants responded to each of the three character assessment measures and assigned a prison sentence for each suspect. Each measure (likeability rating, danger rating, and counselling rating) and the sentencing measure was presented in the same order for each suspect for each participant, and participants assigned their ratings to suspects in the same order that the crime reports had been presented. For example, if a participant had first been presented with Tom's crime report, they first made a judgement about Tom's likeability, then how dangerous they perceived Tom to be, and finally how likely they felt Tom would be to benefit from counselling. After finally assigning a prison sentence to Tom, the participant moved on to assign likeability, danger, and counselling ratings for Kevin, and lastly to assign a prison sentence for Kevin. All participants were required to indicate a score for each measure before moving to the next prompted question. Following the character assessment and sentencing measures, each participant was required to complete the RWA and the SDO measure, in that order. Finally, participants completed the recognition memory test for each suspect, in the same sequence that the reports were presented initially. Following the completion of both memory tests, all participants were thanked and debriefed on the true purpose of the experiment. Each experiment presented took approximately 25 minutes for participants to complete.

Results

Data Exclusion Criteria

Prior to data analysis, it was necessary to consider outlier exclusion criteria carefully. Due to the subjective nature of the character assessment measures used in this study (e.g., "*How likeable do you judge Tom to be?"* or "*To what extent do you feel the suspect might be helped by counselling?*") and considering the nature of the experimental manipulation (i.e., to *not* include

information identified as false in any assessments), I correctly anticipated that box-plot and extreme score output analysis would reveal a considerable number of outliers. For instance, it was expected that a number of participants might reasonably judge the extenuating misinformation as *extremely dangerous* or, that any participant might assign a much lesser prison sentence (or no prison time at all) to the suspect described by the exacerbating false information, particularly if they were adhering to the instructions provided (i.e., remembering and *discrediting* the exacerbating details), or responding based on their perception of a realistic punishment for the described crime in Manitoba, Canada in 2022.

With these considerations in mind, I opted to include all univariate outliers—with the exception of any identified problematic responders—and instead focused on removing only extreme multivariate outliers. Boxplots and extreme values were inspected for all variables and three participants were flagged for suspicious responding; two of these three participants responded with extreme values to all character assessment and sentencing measures (i.e., responding with only 7's and assigning both suspects 20 years in prison) while the third also answered extremely in this manner, and only provided 'neutral' or 'neither' responses to all RWA and SDO questions, and 'never' to all recognition memory test items. These three participants were removed from the analysis. To identify the extreme multivariate outliers in the dataset, I conducted a mahalanobis distance test (Ghorbani, 2019) which identified 11 extreme multivariate outliers ($p < .001$), each of which was inspected and removed. Finally, for the recognition memory test variables to be meaningful, it was important to ensure that the participants included in the analysis demonstrated some recollection of the details in reports as reflected by their recognition memory test accuracy. However, due to the nature of the experiment, it was also expected that many well-intentioned test-takers would perform below

chance on one or both memory tests. Therefore, only those participants who scored lower than 30% (i.e., answered fewer than 9 test questions correctly) on *both* recognition memory tests were excluded from data analysis. This exclusion criteria identified an additional 7 problematic responders, and each of these were removed from the data set bringing the total number of outliers removed from the data set prior to analysis to 21. All results were computed using IBM SPSS Statistics for Windows, Version 28.0 (2021).

Testing the Significance of Order and Suspect as Variables

Recall that one difference between this study and Gilbert's was the counterbalancing of the order of the presentation of Tom and Kevin's crime— in Gilbert's original experiment, all participants read Tom's crime report prior to Kevin's crime report, though the severity of the false information was counterbalanced across participants. Although there is no theoretical reason to expect that order of presentation would have any bearing on the results, a series of repeated-measures analysis-of variance (ANOVA) tests were conducted to confirm that the order of crime report presentation nor any differences in the reports themselves (i.e., specifics of *Tom*'s report versus *Kevin*'s report) had any significant effect on any of the dependent variables included in the experiment. A 2 x 2 mixed ANOVA treating order of suspects (*Tom then Kevin* versus *Kevin then Tom*) as the between-subjects factor and likeability ratings as the within-subjects factor revealed no significant effect of order or suspect, $F(1, 459) = 0.83$, $MSE = 2.18$, $p = .36$, $\eta_p^2 = 0.002$. Three identical individual mixed ANOVAs were performed, each assigning a different dependent variable as the within-subjects factor. These additional tests revealed similar insignificant effects of order or suspect on danger ratings, $F(1, 459) = 0.003$, $MSE = 0.01$, $p = .96$, $\eta_p^2 = .000$, perceptions of suspects' potential to be helped by counselling, $F(1, 459) = 0.05$, $MSE = 0.17$, $p = .82$, $\eta_p^2 = .000$, or the length of prison sentence assigned, $F(1, 459) = .56$, *MSE*

= 18.11, p = .45, η_p^2 = .001. Three additional repeated-measures ANOVA tests further confirmed no significant effect of order on memory accuracy, F(1,459) = 2.68, MSE = .17, p = 0.10, η_p^2 = .006, the tendency to judge false statements as true, F(1,459) = 1.48, MSE = 0.33, p= .23, η_p^2 = .007, or to judge foil statements as true, F(1,459) = 1.21, MSE = 0.03, p = .27= η_p^2= .003. These results confirm Gilbert's original assumption— the specifics of Tom or Kevin's report nor the order of their presentation had any effect on the results of the experiment. These results allowed for collapsing across the between-subjects factors of order and suspect, leaving the severity of the false information as the sole within-subjects variable for the analyses that follow. Collapsing across these variables resulted in approximately 55 to 60 participants in each cell of the design. In all cases wherein a significant main effect or interaction is reported—in both Experiment 1 and Experiment 2—post-hoc pairwise comparison analyses always included Bonferroni adjustments for multiple comparisons.

Character Assessments and Prison Sentencing Scores

Likeability Ratings

All character assessment rating and prison sentencing means, along with those reported by Gilbert et al. are presented below in **Table 1**. Bearing in mind that for each character assessment measure (i.e., 7-pt Likert scales) higher scores reflect a more negative assessment of the suspect than lower scores, I conducted a series of mixed ANOVAs, the first of which revealed a main effect of the severity of false information on likeability ratings, F(1, 453) = 278.61, MSE = 549.06, p < 0.001, η_p^2 = .381. Overall, participants indicated greater dislike for the suspect when the false information presented exacerbated the crime than when the false information extenuated the crime (M_{worse} = 5.72, M_{better} = 4.17, SE = .09, p < .001). There were

no significant effects found for likeability ratings due to the format of the false information, $F(1, 453) = .075$, $MSE = 0.20$, $p = .79$, $\eta_p^2 = .000$, whether participants were warned, $F(1, 453) = 0.14$, $MSE = 0.37$, $p = .71$, $\eta_p^2 = .000$, or whether participants experienced the interruption task, $F(1, 453) = 0.19$, $MSE = 0.49$, $p = .67$, $\eta_p^2 = .000$.

The analysis also revealed a significant interaction between the severity of the false information and whether a warning of its presence was provided, $F(1, 453) = 39.43$, $MSE = 77.70$, $p < .001$, $\eta_p^2 = .080$ (see **Figure 1**)**.** Post-hoc pairwise comparisons revealed that when the false information extenuated the crime, those in the *no warning* condition rated the suspects as more likeable than those who were warned ($M_{\text{nowarning}} = 3.90$, $M_{\text{warning}} = 4.44$, $SE = 0.15$, $p < .001$) however when the false information exacerbated the crime the pattern reversed—those who were not warned of the presence of false information rated the suspect as less likeable than those in the warning condition ($M_{\text{nowarning}} = 6.02$, $M_{\text{warning}} = 5.40$, $SE = 0.13$, $p < .001$).

Table 1

Mean Character Assessment Ratings and Prison Terms Assigned to Suspects

Uninterrupted Groups

	FLUENT FALSE						DISFLUENT FALSE					
	Better	*(SD)*	**Worse**	*(SD)*	**Diff**	*p-value*	**Better**	*(SD)*	**Worse**	*(SD)*	**Diff**	*p-value*
Prison Term												
Gilbert et al.	**6.03**	*n/a*	**7.03**	*n/a*	**-1.00**	*<.07*	**6.03**	*n/a*	**7.03**	*n/a*	**-1.00**	*<.07*
Exp 1 Sentence (yrs)	**6.51**	*4.05*	**8.19**	*4.54*	**-1.68**	*<.001*	**4.82**	*3.05*	**8.00**	*4.28*	**-3.18**	*<.001*
Likeable Rating												
Gilbert et al.	**5.83**	*n/a*	**6.62**	*n/a*	***-0.79***	***<.05***	**5.83**	*n/a*	**6.62**	*n/a*	***-0.79***	***<.05***
Exp 1 Converted Score	**5.59**		**6.90**		**-1.31**		**5.49**		**6.90**		**-1.41**	
Exp 1 Raw Score	**4.35**	*1.61*	**5.37**	*1.44*	**-1.02**	*<.001*	**4.27**	*1.57*	**5.37**	*1.76*	**-1.10**	<.001
Danger Rating												
Gilbert et al.	**5.36**	*n/a*	**6.41**	*n/a*	**-1.05**	***<.05***	**5.36**	*n/a*	**6.41**	*n/a*	**-1.05**	***<.05***
Exp 1 Converted Score	**5.62**		**6.69**		**-1.07**		**4.96**		**6.92**		**-1.95**	
Exp 1 Raw Score	**4.37**	*1.68*	**5.20**	*1.31*	**-0.83**	*<.001*	**3.86**	*1.43*	**5.38**	*1.32*	**-1.52**	<.001
Benefit from Counselling												
Gilbert et al.	**6.45**	*n/a*	**5.85**	*n/a*	**0.60**	***0.22***	**6.45**	*n/a*	**5.85**	*n/a*	**0.60**	**0.22**
Exp 1 Converted Score	**5.00**		**3.95**		**1.05**		**4.22**		**4.05**		**0.17**	
Exp 1 raw score	**2.33**	*1.66*	**3.15**	*1.67*	**-0.82**	<.001	**2.94**	*1.89*	**3.07**	*1.57*	**-0.13**	<.001

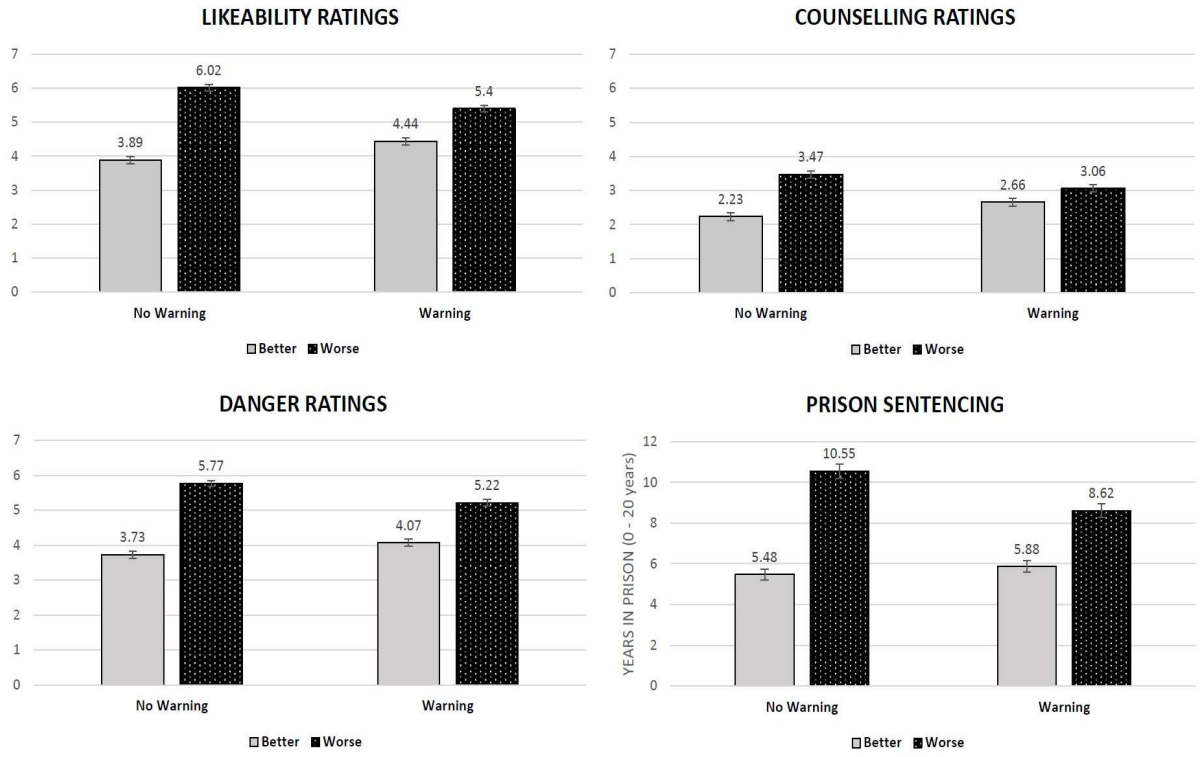
Interrupted Groups

	FLUENT FALSE						DISFLUENT FALSE					
	Better	*(SD)*	**Worse**	*(SD)*	**Diff**		**Better**	*(SD)*	**Worse**	*(SD)*	**Diff**	
Prison Term												
Gilbert et al.	**5.83**	*n/a*	**11.15**	*n/a*	**-5.32**	***<.001***	**5.83**	*n/a*	**11.15**	*n/a*	**-5.32**	***<.001***
Exp 1 Sentence (Yrs)	**6.31**	*4.48*	**8.93**	*5.38*	**-2.62**	<.001	**5.88**	*3.69*	**9.36**	*5.64*	**-3.48**	<.001
Likeable Rating												
Gilbert et al.	**5.18**	*n/a*	**7.00**	*n/a*	**-1.82**	***<.05***	**5.18**	*n/a*	**7.00**	*n/a*	**-1.82**	***<.05***
Exp 1 Converted Score	**5.75**		**6.74**		**-0.99**		**5.99**		**7.23**		**-1.23**	
Exp 1 Raw Score	**4.47**	*1.66*	**5.24**	*1.70*	**-0.77**	<.001	**4.66**	*1.48*	**5.62**	*1.76*	**-0.96**	<.001
Danger Rating												
Gilbert et al.	**4.82**	*n/a*	**7.06**	*n/a*	**-2.24**	***<.05***	**4.82**	*n/a*	**7.06**	*n/a*	**-2.24**	***<.05***
Exp 1 Converted Score	**5.18**		**6.53**		**-1.35**		**5.17**		**6.72**		**-1.56**	
Exp 1 Raw Score	**4.03**	*1.51*	**5.08**	*1.43*	**-1.05**	<.001	**4.02**	*1.47*	**5.23**	*1.46*	**-1.21**	<.001
Benefit from Counselling												
Gilbert et al.	**6.45**	*n/a*	**5.03**	*n/a*	**1.42**	***<.05***	**6.45**	*n/a*	**5.03**	*n/a*	**1.42**	***<.05***
Exp 1 Converted Score	**3.92**		**4.43**		**-0.50**		**5.17**		**3.82**		**1.35**	
Exp 1 raw score	**3.17**	*1.90*	**2.78**	*1.53*	**0.39**	<.001	**2.2**	*1.42*	**3.25**	*1.51*	**-1.05**	<.001

Note. There was a difference between the Likert scales used in Gilbert's study and those used in this research. Gilbert used 9-pt Likert scales for Liking, Danger, and Counselling ratings while this study used 7-pt Likert scales. Raw scores based on the 7-point Likert scale are reported along with original standard deviations. All character assessment means were also converted to a score on a 9-pt Likert scale and are reported above the original means. Scores for counselling ratings were first converted to a 9-pt Likert score before being subtracted from 8 (maximum scale point plus one) to obtain the reverse coded score. All means reported are restricted to warned participants only (n = 225)

Figure 1

Interaction Effects Between Receiving a Warning and the Severity of False Details on Character Assessments (Experiment 1)



Note. For all ratings, a higher score represents a more negative assessment of the suspect. Likeability: *0 = neutral, 7 = extreme dislike;* Danger: *0 = slightly dangerous, 7 = extremely dangerous;* Counselling: *0 = helped a great deal, 7 = not helped at all;* Prison Sentence: *option to sentence suspects from 0 to 20 years for the crimes.* Error bars represent standard error. (*n = 461)*

Danger Ratings

The same mixed ANOVA analysis described above was performed for the danger ratings assigned to suspects and once again revealed a significant main effect of the false information severity, $F(1,453)=327.46$, $MSE= 586.70$, $p < .001$, $\eta_p^2 = .420$. Participants perceived the suspect as significantly more dangerous when the false information exacerbated the crime than when the false information extenuated the crime ($M_{worse} = 5.50$, $M_{better} = 3.90$, $SE = .09$, $p < .001$). The analysis revealed no significant effects of the format of the false information, $F(1, 453) = 0.64$, $MSE = 1.48$, $p = .42$, $\eta_p^2 = .001$, whether the participant was warned, $F(1, 453) =$

1.03, $MSE = 2.37$, $p = .31$, $\eta_p^2 = .002$ nor whether the participant was interrupted, $F(1, 453) = 2.54$, $MSE = 5.82$, $p = .11$, $\eta_p^2 = .112$, on danger assessments of the suspects.

This analysis revealed another significant interaction between the severity of the false information and whether the participant was warned of its presence on danger assessments, $F(1, 453) = 25.46$, $MSE = 45.62$, $p < .001$, $\eta_p^2 = .053$. Pairwise comparison analysis revealed that when the false information extenuated the crime, those in the *no warning* condition rated the suspects as less dangerous than those who were warned *(*$M_{\text{nowarning}} = 3.73$, $M_{\text{warning}} = 4.07$, $SE = 0.15$, $p = .02$), however when the false information exacerbated the crime, those in the *no warning* condition rated the suspect as *more* dangerous than those who were warned ($M_{\text{nowarning}} = 5.77$, $M_{\text{warning}} = 5.22$, $SE = 0.12$, $p < .001$). No other interactions were significant.

Likelihood to Benefit from Counselling Ratings

An identical mixed ANOVA revealed a significant main effect of severity on participants' assessments of how likely they believed counselling might benefit each suspect, $F(1, 453) = 76.03$, $MSE = 155.80$, $p < .001$, $\eta_p^2 = .144$. Overall, participants judged suspects to be more likely to benefit from counselling when the false information extenuated than when the false information exacerbated the crime ($M_{\text{better}} = 2.44$, $M_{\text{worse}} = 3.27$, $SE = .09$, $p < .001$). The analysis revealed no significant effects on counselling ratings as a result of the fluency of the false information, $F(1, 453) = 0.70$, $MSE = 2.30$, $p = .41$, $\eta_p^2 = .002$, whether a warning was provided, $F(1, 453) = 0.01$, $MSE = 0.05$, $p = .91$, $\eta_p^2 = .000$, or whether the participant experienced the interruption task, $F(1, 453) = 1.13$, $MSE = 3.73$, $p = .29$, $\eta_p^2 = .002$.

This analysis also revealed a significant interaction between severity of the false information and whether a warning was provided for counselling ratings, $F(1, 453) = 20.12$, *MSE*

= 41.24, p < .001, η_p^2 = .043. Pairwise comparisons revealed that when the false details exacerbated the crime, unwarned participants judged the suspect as less likely to be helped by counselling than warned participants ($M_{nowarning}$ = 3.47, $M_{warning}$ = 3.06, SE = 0.15, p = .007) and this pattern was reversed when the false details extenuated the crime ($M_{nowarning}$ = 2.23, $M_{warning}$ = 2.66, SE = 0.15, p = .005). Finally, a significant three-way interaction was identified between the severity of the false information, the fluency of the false information, and whether the participant experienced an interruption, $F(1,453)$ = 16.74, MSE = 34.31, p < .001, η_p^2 = .036. Pairwise comparisons further indicated that in all conditions, participants rated suspects as less likely to be helped by counselling when the false details exacerbated the crime. When the false information was presented fluently, there was a significant difference in counselling ratings for suspects with extenuating crime details depending whether or not the participant experienced an interruption; those who were interrupted indicated that the suspect would be less likely to benefit than was indicated by those who were not interrupted, despite being presented with identical false details in the crime report ($M_{interrupt}$ = 2.96, $M_{nointerrupt}$ = 2.11, SE = .22, p < .001). There were no other significant differences of note within this interaction.

Prison Sentencing

The final ANOVA analysis of the series revealed a significant main effect of the severity of the false details on the prison sentences assigned, $F(1, 453)$ = 290.83, MSE = 3510.30, p < .001, η_p^2 = .391. Participants assigned significantly longer prison sentences to suspects when the false information embedded in the report exacerbated as opposed to when it extenuated the suspect's reported crime (M_{worse} = 9.61, M_{better} = 5.68, SE = .23, p < .001). There was also a significant main effect of warning on length of prison sentence assigned, $F(1, 453)$ = 4.17, MSE

= 133.50, p = .042, η_p^2 = .009. Participants assigned significantly longer prison sentences to suspects when they were not warned of any misinformation as compared to when they were warned and instructed to disregard information presented as false ($M_{nowarning}$ = 8.01, $M_{warning}$ = 7.25, SE = .37, p =.04). There was no significant effect of the format of the false information, $F(1, 453) = 1.36$, MSE = 43.49, p = .25, η_p^2 = .003, nor of the interruption task, $F(1,453) = 3.64$, MSE = 116.52, p = .057, η_p^2 = .008 on the length of the prison sentences assigned.

The mixed ANOVA revealed a significant two-way interaction between the severity of the false report details and the presence of a warning on the length of prison sentence assigned, $F(1,453) = 25.78$, MSE = 311.14, $p < .001$, η_p^2 = .054. Exploring this interaction, post-hoc comparisons revealed that when the false details exacerbated the crime, those in the *no warning* condition assigned significantly longer prison terms compared to those who had been warned ($M_{nowarning}$ = 10.55, $M_{warning}$ = 8.62, SE = .49, $p < .001$), and this pattern was reversed when the false details extenuated the crime, though this difference was non-significant ($M_{nowarning}$ = 5.48, $M_{warning}$ = 5.88, SE = .38, p = .29). There was no difference in prison sentences assigned between warned and unwarned groups when the false details extenuated the crime, and there were no other significant interactions. Main effects of severity and interaction effects between receiving a warning and the severity of the false details for character assessment ratings are presented in **Figure 1**.

Recognition Memory Test Performance

In their analysis of the memory test results, Gilbert and colleagues' main objective was to determine the effects of the interruption task on participants' memory for the information presented in the reports. The overall accuracy on the interruption task in Experiment 1 was quite

high (M = 84.2%, SD = 0.10), an indication that participants were adhering to the task and were adequately distracted in that condition. It should be noted that due to a programming error, accuracy data was only recorded for 186 of the 231 participants assigned to the *interrupt* condition, though it was confirmed that indeed all 231 participants experienced the same interruption task. Given the large sample size and overall high accuracy scores recorded for this task, this minor error was corrected and is not considered further. All participants in Gilbert's study were warned about the presence of misinformation, and so the following analyses of recognition memory scores is restricted to only those participants who were randomly assigned to the *warning* condition (n = 225).

As stated above, previous tests demonstrate that neither order nor suspect had any effect on participants' memory accuracy, nor on their likelihood of responding 'true' to either *False* or *Foil* statements. To simplify the analysis and results further, three additional repeated-measures ANOVAs were conducted to test for any effects of severity (*better* versus *worse*) on the three primary memory outcomes of interest in this and Gilbert's study; memory accuracy, the likelihood of responding 'true' to *False* statements, and the likelihood of responding 'true' to *Foil* statements. All tests indicated there was no significant effect of the severity of the false details on memory accuracy, $F(1, 224) = 2.63$, $MSE = .02$, $p = .11$, $\eta_p^2 = .012$, on participants' likelihood of responding 'true' to *False* statements, $F(1, 224) = 0.11$, $MSE = .003$, $p = .76$, $\eta_p^2 = .000$, or the likelihood of responding 'true' to *Foil* statements, $F(1, 224) = 0.21$, $MSE = .001$, $p = .69$, $\eta_p^2 = .001$. As a result, the severity of the false statements was removed as a within-subjects factor for the recognition memory test analyses, unless otherwise stated. The memory test dependent variables were then calculated by computing the average of the two scores for

each variable for each participant. The ten new dependent variable scores were: i) overall memory accuracy, ii) proportion of correct 'true' responses, iii) proportion of correct 'never' responses (foils), iv) proportion of correct 'false', v) proportion of *Foils* remembered as 'true', vi) proportion of *False* remembered as 'true', vii) proportion of *True* remembered as 'never', viii) proportion of *False* remembered as 'never', ix) proportion of *True* remembered as 'false', and x) proportion *of Foils* remembered as 'false'. Exactly as the case in Gilbert's study, for any participant included in this analysis (i.e., a warned participant), a perfect memory score would consist of 19 correct 'never' responses, 7 correct 'false' responses, and 4 correct 'true' responses on each of the two memory tests.

Comparing Recognition Memory Outcomes with Gilbert's Results

Gilbert and colleagues' analyses of memory test results were conducted and reported differently than that reported here since in their study, only false statements were interrupted, and only correlational results were reported. With these considerations in mind, Gilbert et al. report four key findings pertaining to the recognition memory test: 1) interrupted and uninterrupted participants were equally likely to misremember *True* statements (this result is rather ambiguous, but was interpreted as either responding 'never' or 'false' to *True* statements), 2) interrupted participants were more likely to respond 'true' *to False* test items, 2) interrupted participants were more likely to misremember *Foil* items as 'true', 3) interrupted participants were more likely to respond 'true' to *False* test items, and 4) the proportion *of False* statements remembered as 'true' was positively correlated with the length of prison term assigned when the participant experienced the interruption. Specifically, Gilbert found that those who experienced the interruption assigned nearly twice the length of prison term when the false details in the report

exacerbated the crime compared to when they extenuated the crime, but this difference did not occur for those who did not experience the interruption task.

To compare Experiment 1 results with these findings I first conducted a one-way multivariate analysis of variance (MANOVA) assigning interruption as my between-subjects factor and the memory outcomes listed above as dependent variables (i.e., i) proportion of 'never' responses to *True* items, ii) proportion of 'false' responses *to True* items, iii) proportion of 'true' responses to *False* items, and iv) proportion of 'true' responses to *Foil* items). Gilbert's first finding was confirmed: there were no significant differences between interrupted and uninterrupted participants in terms of their likelihood for misidentifying *True* statements as 'never', $F(1, 223) = 1.21$, $MSE = .017$, $p = .27$, $\eta_p^2 = .005$, or misremembering *True* statements as 'false', $F(1, 224) = 1.69$, $MSE = 0.015$, $p = .20$, $\eta_p^2 = .008$. However, interrupted participants were neither more likely to misremember *Foil* statements as 'true', $F(1, 224) = 0.96$, $MSE = .012$, $p = .33$, $\eta_p^2 = .004$, nor more likely to respond 'true' to *False* statements, $F(1 ,224) = 2.41$, $MSE = 0.265$, $p = .12$, $\eta_p^2 = .011$, than their uninterrupted counterparts. Finally, a one-way ANOVA revealed no difference between interrupted and uninterrupted participants in terms of the length of prison sentence assigned, $F(1, 223) = 2.14$, $MSE = 29.53$, $p = .15$, $\eta^2 = .009$. A second repeated-measures ANOVA confirmed that there was no main effect of interruption, $F(1, 223) = 2.16$, $MSE = 59.78$, $p = .14$, $\eta_p^2 = .010$, nor any interaction effect between interruption and the severity of the false information on assigned prison sentences, $F(1, 223) = 0.92$, $MSE = 11.27$, $p = .34$, $\eta_p^2 = .004$. These and Gilbert's comparative findings for interrupted and uninterrupted part are presented in **Table 2.**

Table 2

Comparison of Interruption Effects on Memory Errors

	No Interruption Condition		
	Gilbert et al.	**Experiment 1**	**Experiment 2**
Correct "False"	0.69	0.41	0.38
"True" to FALSE	0.23	0.42	0.46
"Never" to FALSE	0.09	0.16	0.14

	Interruption Condition		
	Gilbert et al.	**Experiment 1**	**Experiment 2**
Correct "False"	0.34	0.45	0.37
"True" to FALSE	0.44	0.36	0.43
"Never" to FALSE	0.23	0.19	0.19

Note. Cell means represent average proportion of each type of response for each experiment. Warned participants only (n = 225).

The results reported here do not reflect the effects of interruption on memory performance presented in Gilbert's study, however correlational analysis did reveal a significant positive correlation between the likelihood of misremembering *False* statements as 'true' and the average prison sentence assigned when the false information exacerbated the crime, $r(225) = .27$, $p < .001$, and a significant negative correlation between these variables when the false information extenuated the crime, $r\ (225) = -.19$, $p = .005$. Finally, Gilbert identifies similar correlational patterns between the character assessments and the tendency to misremember *False* items as 'true', and this pattern was also identified in Experiment 1. Pearson correlations revealed that while the overall tendency to respond 'true' to *False* statements was uncorrelated with the average liking, danger, and counselling ratings, there was a significant positive

correlation between this memory error and character assessments when the details exacerbated the story, and a significant negative correlation between this error and assessments when the details extenuated the crime (with the exception of counselling rating, in this case). Correlational data between this response tendency and character assessments are presented in **Table 3.**

Table 3

Correlations Between Tendency to Respond 'True' to False Statements and Character Assessment Ratings

	Liking	*p-value*	Danger	*p-value*	Counselling	*p-value*	Prison	*p-value*
Better False Details	-0.275	*<.001*	-.281	*<.001*	-0.055	*.42*	-.156	*.019*
Worse False Details	0.324	*<.001*	.326	*<.001*	0.179	*.007*	.235	*<.001*

Note. Pearson's *r* correlations reported. Warned participants only (*n* = 225)

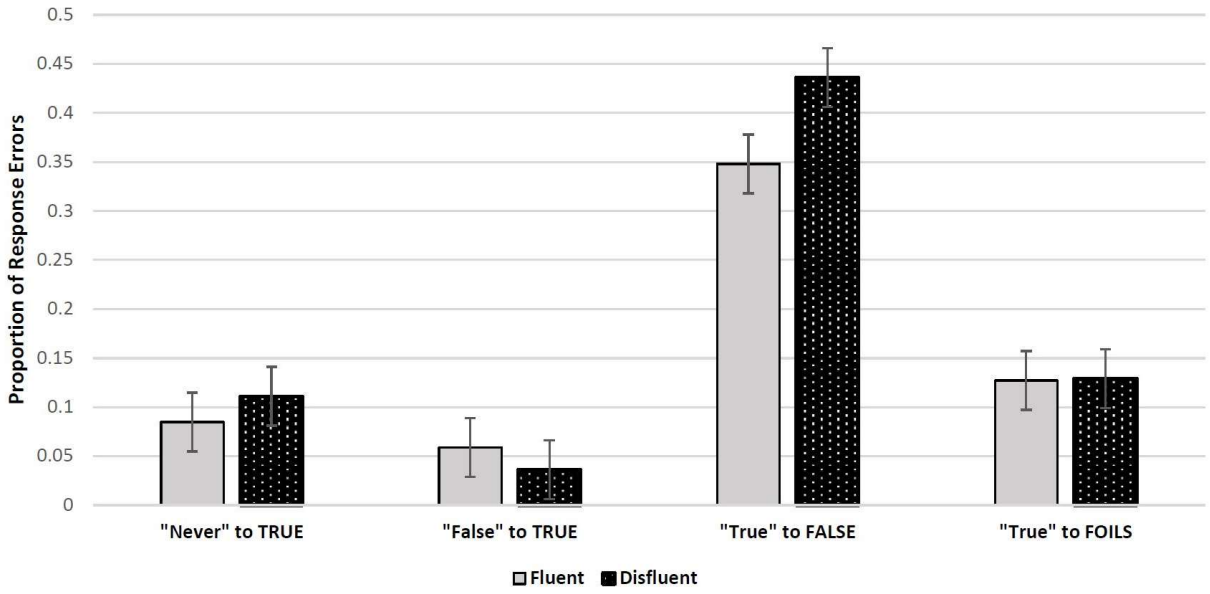
Effects of False Format on Memory and Sentencing Outcomes

Following the analysis comparing Experiment 1 results to Gilbert's, I was interested in identifying whether the disfluency manipulation had any effect on participants' memory accuracy, or on participants' tendency to misremember *False* (or *Foil*) statements as 'true'. To test this, I conducted a similar MANOVA to that performed above however assigning false format as the between-subjects variable and adding *overall accuracy* to the above listed dependent variables. This analysis revealed a significant effect of disfluency on participants' likelihood of responding 'true' to *False* statements, $F(1, 223) = 4.00$, $MSE = .44$, $p = .047$, $\eta_p^2 = .018$ (see **Figure 2**). Exploring this relationship further, pairwise comparisons revealed the (small) main effect of false format; participants were approximately 9% more likely to misremember *False* statements as 'true' when the false details were presented disfluently as compared to when the details were fluently presented ($M_{\text{fluent}} = .35$, $M_{\text{disfluent}} = .44$, $SE = .04$, $p =$

.047). In addition, a repeated-measures ANOVA including severity of the false statements as a within-subjects factor found no interaction between misremembering *False* for 'true' and the severity of the false details $F(1, 223) = 0.08$, MSE - .002, $p = .78$, $\eta_p^2 = .000$. Finally, a one-way ANOVA indicated no main effect of false format on the average length of prison sentence assigned, $F(1, 223) = .90$, $MSE = 12.48$, $p = .34$, $\eta_p^2 = .004$.

Figure 2

Main Effect of Disfluency on the Tendency to Misremember False Details as True



Note. Proportions represent average proportion of responses.
Perfect memory score: 19 correct 'never' responses, 4 correct 'true' responses, 7 correct 'false' responses.
Error bars represent standard errors.
Includes warned participants only.

Assessing RWA and SDO as Potential Moderating Variables

The main hypotheses concerning the inclusion of RWA and SDO scores was to test whether these personality constructs might predict the length of prison sentence assigned to suspects, and therefore possibly one's tendency to misremember false statements as true. Consistent with previous research, the RWA scale (15 items, $\boldsymbol{\alpha}$ = .85) and the SDO scale (22 items, $\boldsymbol{\alpha}$ = .93) demonstrate high internal reliability, and the two constructs were significantly correlated, r = .51, p < .001. This finding is also consistent with previous research (Altemeyer, 2004; Duckitt, 2001; Roccato & Ricolfi, 2005; Strube & Rahimi, 2006; Zakrisson, 2005). Previous studies investigating outcomes related to the two constructs utilize various methods for analyzing scores, though designs chosen for studies involving large datasets tend to implement quartile (Altemeyer, 2004) or tercile (Son Hing, 2007) splits to the datasets. Prior to data analysis, the latter method was implemented, and the data was split into three equal groups *(Low, Med*, and *High*) for each set of scores. Ranges for RWA scores were as follows: *Low*: 0 – 3.20 (*n* = 156), *Med*: 3.21 – 3.92 (*n* = 161), *High*: 3.93 – 5.33 (*n* = 144). Ranges for SDO scores were as follows: *Low*: 0 – 1.91 (*n* = 156), *Med*: 1.92 – 2.73 (*n* =152), *High*: 2.74 – 4.27 (*n* =153). These cut-offs are within range of tercile cut-offs reported in previous studies (e.g., Son Hing, 2007).

Gender Differences

A consistent finding within the RWA and SDO literature is that a robust sex difference exists for social dominance orientation with men tending to have higher SDO scores than women (Dyrendal et al., 2021; Zakrisson, 2005) but that lesser or insignificant differences exist in favor of men in terms of right-wing authoritarianism scores (Heaven & Bucci, 2001; Zakrisson, 2005; c.f., Oliver & Wood, 2014). To determine whether the data of Experiment 1 matched this pattern, I conducted a one-way MANOVA which indicated a significant main effect of gender on SDO

scores, $F(2, 458) = 31.36$, $MSE = 20.14$, $p < .001$, $\eta_p^2 = .120$. Pairwise comparisons revealed that males had significantly higher SDO scores than females ($M_{males} = 2.79$, $M_{females} = 2.19$, $SE = .07$, $p < .001$) and significantly higher SDO scores than those who did not identify as male or female ($M_{other} = 1.94$, $SE = .31$, $p = .006$). The MANOVA analysis also revealed a small main effect of gender on RWA scores, $F(2, 458) = 3.15$, $MSE = 1.60$, $p = .04$, $\eta_p^2 = .014$, however pairwise comparisons revealed no significant difference between males and females ($M_{male} = 3.63$, $M_{female} = 3.48$, $SE = .07$, $p = .09$) or the seven participants who did not identify as male or female ($M_{other} = 3.20$, $SE = .27$, $p = .36$).

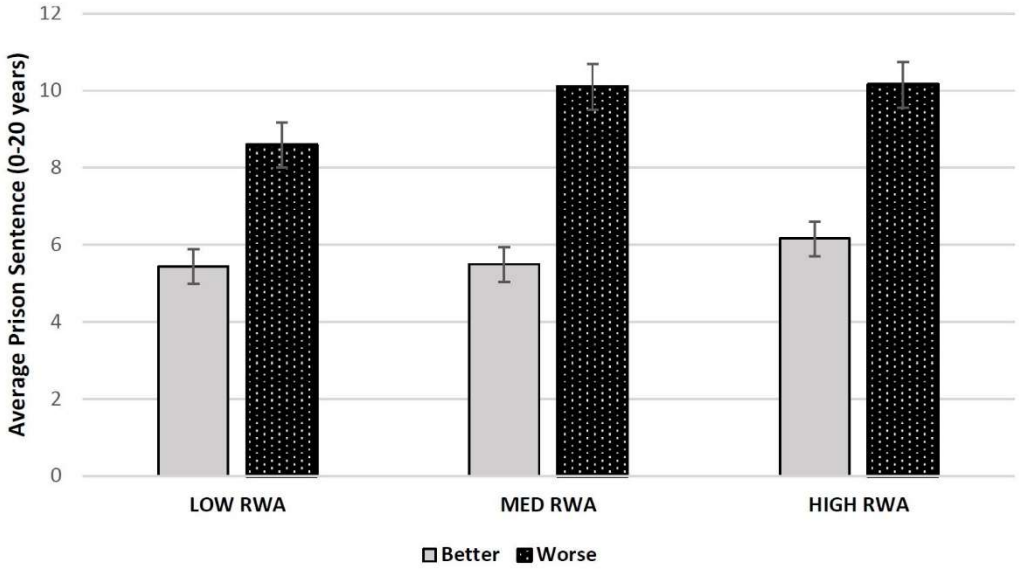
Analysis of Right-Wing Authoritarianism Scores

To determine whether RWA scores contributed to the length of prison sentence assigned, or an individual's tendency to misremember false statements as true, I first conducted a one-way MANOVA setting RWA as my independent variable with three categorical levels (*Low, Med* and *High*), and assigning average prison sentences, prison sentences for suspects when the false details exacerbated the crime, and sentences for suspects when the false information extenuated the crime as my dependent variables. As mentioned previously in this manuscript, these variables were included to determine whether the severity of the false details might affect the decisions of *High* RWA individuals differently from *Low* RWAs, given numerous previous studies reporting that individuals with right-of-center ideologies have a tendency to spend more time attending to—and have stronger responses toward— aversive over appetitive stimuli compared to those with left-leaning ideological beliefs (Carraro et al., 2011; Dodd et al., 2012). Again, all analysis of RWA and SDO measures that follow were restricted to only those participants who had been warned of the presence of the inclusion of false information in the reports. Eliminating the unwarned from the analysis left a sample size of 225 participants with approximately equal

numbers of participants in each of the *Low, Med,* and *High* RWA level groups (n_{low} = 78, n_{med} = 79, n_{high} = 68). Results of this MANOVA reveal a significant main effect of RWA scores on prison sentences, though only when the false information exacerbated the crime, $F(2, 222) = 3.52$, $MSE = 85.75$, $p = .03$, $\eta^2 = .031$. There were no main effects of RWA level on the average length of prison sentence assigned, $F(2, 222) = 1.25$, $MSE = 17.25$, $p = .29$, $\eta^2 = .011$, or the length of prison sentence assigned when the false information extenuated the crime, $F(2, 222) = 0.22$, $MSE = 3.32$, $p = .80$, $\eta^2 = .002$. Post hoc pairwise comparisons revealed that *High* RWA individuals assigned significantly longer prison sentences to suspects than *Low* RWA individuals when the false information exacerbated the crime ($M_{high} = 9.57$, $M_{low} = 7.47$, $SE = .82$, $p = .03$).

Figure 3

Level of RWA Predicts Length of Prison Sentence for Exacerbated Crimes



Note. RWA = Right-Wing Authoritarianism.
Low RWA: 0-3.27 (n = 98); *Med RWA:* 3.28 -3.87 (n = 88); *High RWA:* 3.88 -5.46 (n =84).
Error bars represent standard errors.
Includes warned participants only.

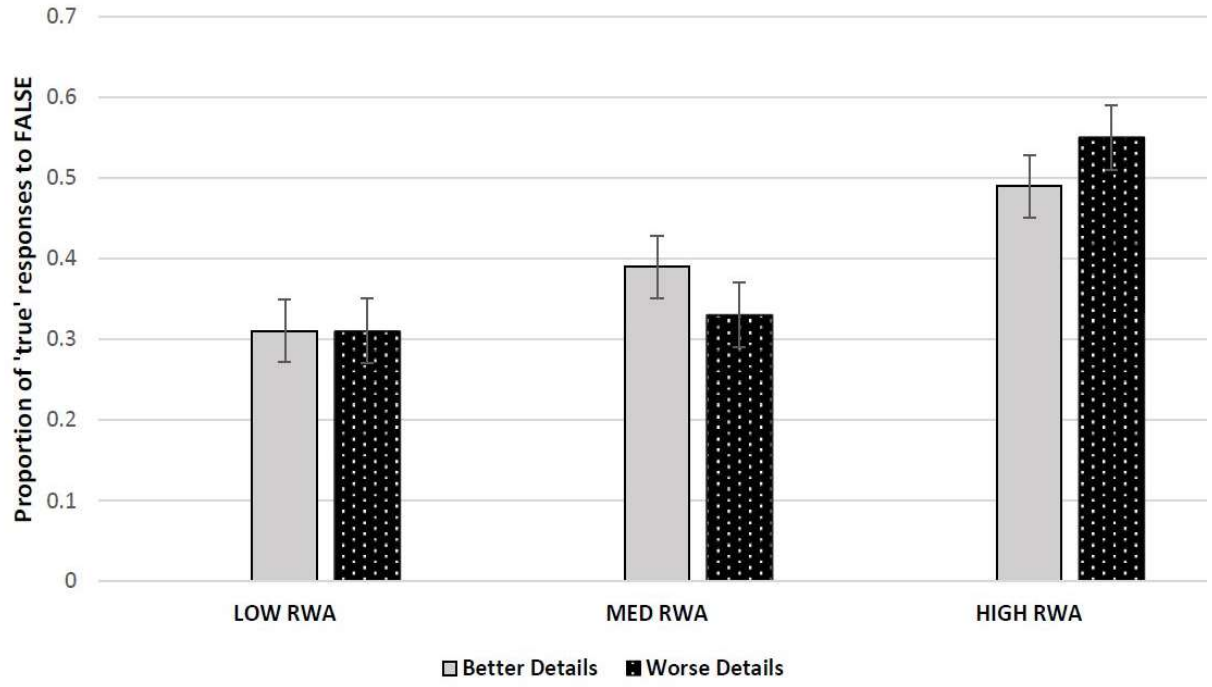
To further investigate whether one's level of RWA might predict the length of prison sentence assigned, a multiple regression analysis was performed setting the levels of RWA as categorical predictor variables and the prison sentence assigned when the false information exacerbated the crime as the outcome variable. The regression model was significant, $F(2, 222) = 3.52$, $MSE = 85.75$, $p = .03$, $\eta^2 = .031$, and indicated that *High* RWA significantly predicted participants' sentencing for the associated crime, $R^2 = .031$, $\beta = 0.19$, $b = 2.10$, $p = .01$, 95% CI [.49, 3.7]. A follow-up regression analysis revealed that RWA level was not a significant predictor of the average length of prison sentence assigned, $F(2, 222) = 1.25$, $MSE = 17.25$, $p = .29$, $\eta^2 = .011$, or the length of prison sentence when the false information extenuated the crime $F(2, 222) = 0.22$, $MSE = 3.32$, $p = .80$, $\eta_p^2 = .002$ (see **Figure 3**).

RWA level also had a significant effect on memory performance. First, a one-way ANOVA revealed significant differences between levels of RWA and overall memory accuracy, $F(2, 222) = 12.76$, $MSE = 0.43$, $p < .001$, $\eta_p^2 = .103$*: Low* RWAs had higher accuracy scores than *Med* RWAs ($M_{low} = .77$, $M_{med} = .69$, $SE = .03$, $p = .03$), and *High* RWAs ($M_{high} = .62$, $SE = .03$, $p < .001$). To explore these differences further, a one-way MANOVA was performed setting the average proportion of incorrect 'true' responses to *False* statements, the proportion of 'true' to *False* statements when the details extenuated the crime, and the proportion of 'true' to *False* statements when the details exacerbated the crime as dependent variables. This test revealed significant main effects of RWA level on overall tendency to misremember *False* items as 'true', $F(2, 222) = 8.44$, $MSE = 0.87$, $p < .001$, $\eta^2 = .071$, regardless of the severity of the false information involved (*worse*: $F(2, 222) = 5.60$, $MSE = .62$, $p = .004$, $\eta^2 = .048$; *better*: $F(2, 222) = 10.26$, $MSE = 1.25$, $p < .001$, $\eta^2 = .085$. Post-hoc pairwise comparisons further explained that

High RWA's were significantly more likely to misremember false statements as 'true' compared to *Low* RWA's overall, $M_{high} = .52$, $M_{low} = .31$, $SE = .05$, $p = < .001$, $\eta^2 = .071$, when the false information exacerbated the crime, $M_{high} = 0.55$, $M_{low} = .31$, $SE = .06$, $p < .001$, $\eta^2 = .085$, and when the false information extenuated the crime, $M_{high} = .49$, $M_{low} = .31$, $SE = .058$, $p < .001$, $\eta^2 = .048$ (see **Figure 4a**).

Figure 4a

Level of RWA Predicts Likelihood of Misremembering False Details as True



Note. RWA = Right-Wing Authoritarianism.
Low RWA: 0-3.27 (n = 98); *Med RWA*: 3.28 -3.87 (n = 88); *High RWA*: 3.88 -5.46 (n =84).
Error bars represent standard errors.
Includes warned participants only.

To determine whether RWA level might predict the likelihood of misremembering false information as true, a final multiple regression analysis was performed assigning the average proportion of incorrect 'true' responses *to False* statements as the dependent variable. All RWA levels were significantly different in their tendency to misremember false details as true. Correlational analysis revealed a significant positive correlation between *High* RWAs and the tendency to misremember *False* statements as 'true', $r(68) = .26$, $p < .001$, and a significant negative correlation between *Low* RWAs and this tendency, $r(78) = -.18$, $p = .004$. The regression model was significant, $F(2, 222) = 8.44$, $MSE = 0.87$, $p = < .001$, and revealed that *High* RWA significantly predicted participants' tendency to misremember false information as true, $R^2 = .07$, $\beta = .29$, $b = .21$, $p < .001$, 95% CI [0.11, 0.32]. This analysis revealed that for *High* RWAs, every one-point increase in RWA increased the likelihood of misremembering *False* items as 'true' by approximately 21% as compared to *Low* RWAs who received identical information.

Analysis of Social Dominance Orientation Scores

The exact procedures described above were performed for levels of SDO scores to determine whether this construct might also predict one's tendency to assign longer prison sentences, or to misremember *False* statements as 'true'. Removing the unwarned participants from analysis left roughly equal sample sizes in each of the levels of SDO; $n_{low} = 84$, $n_{med} = 77$, $n_{high} = 64$. First, a one-way MANOVA was conducted using the three categorical levels of SDO as my independent variables, and assigning average prison sentence, prison sentences assigned when the false information exacerbated the crime, and prison sentences assigned when the false information extenuated the crime as dependent variables. Unlike the case with RWA, there were no significant differences detected between any level of SDO and the average length of prison

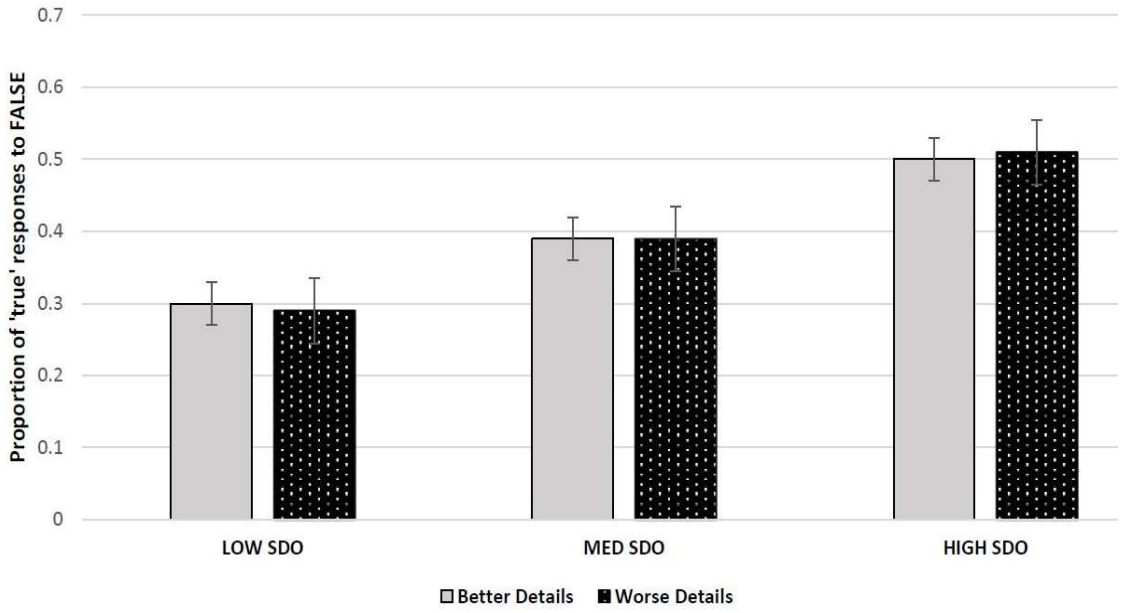
sentence assigned to the suspects, $F(2, 222) = 0.58$, $MSE = 8.14$, $p = .56$, $\eta_p^2 = .005$. This was true whether the false information extenuated, $F(2, 222) = 0.28$, $MSE = 4.27$, $p = .76$, $\eta_p^2 = .003$, or exacerbated the crime, $F(2, 222) = 1.53$, $MSE = 37.96$, $p = .22$, $\eta_p^2 = .014$.

As was found with RWA levels, SDO levels had a significant influence on memory performance (see **Figure 4b**). A one-way ANOVA revealed significant differences in overall memory accuracy between *Low, Med* and *High* SDO individuals, $F(2, 222) = 15.44$, $MSE = 0.51$, $p < .001$, $\eta^2 = .122$. Pairwise comparisons indicate that *High* SDOs ($M_{high} = .59$, $SE = .03$) had significantly less accurate overall memory performance than *M*ed SDOs ($M_{med} = .71$, $SE = .03$, $p < .001$), and *L*ow SDOs ($M_{low} = .76$, $SE = .03$, $p < .001$). There was no significant difference in memory accuracy between *Med* SDOs and *Low* SDOs. To assess whether there were also differences between SDO levels and misremembering *False* items as 'true', a one-way MANOVA was performed using the same three dependent variables that were used when testing RWA scores. Again, there was a significant main effect of SDO level on participant's overall tendency to incorrectly respond 'true' *to False* statements, $F(2, 222) = 7.95$, $MSE = 0.83$, $p < .001$, $\eta^2 = .067$, and this occurred when the false information in the report extenuated, $F(2, 222) = 6.68$, $MSE = 0.73$, $p = .002$, $\eta^2 = .057$, and when it exacerbated the crime, $F(2, 222) = 7.46$, $MSE = 0.93$, $p < .001$, $\eta^2 = .063$. Additional pairwise comparisons revealed that *High* SDOs were significantly more likely to respond 'true' to *False* statements than their *Low* SDO counterparts ($M_{high} = .51$, $M_{low} = .30$, $SE = .04$, $p < .001$), but revealed no significant differences in this tendency between *High* SDOs and *Med* SDOs, or *Med* SDOs and *Low* SDOs. An identical multiple regression analysis as that performed with RWA levels revealed that *High* SDO was positively correlated with average proportion of 'true' responses to *False* statements ($r = .23$, $p <$

.001), *Low* SDO had a significant negative association with proportions of this response (r = -.22, p < .001), and *Med* SDO was uncorrelated with this tendency (r = .01, p = .44). The regression analysis was significant, $F(2, 222) = 7.95$, MSE =0.83, p < .001, another indication that *High* SDO was a significant predictor of participants' tendency to misremember false information as true, R^2 = .07, β = .290, b = .22, p < .001, 95% CI [0.11, 0.32].

Figure 4b

Level of SDO Predicts Likelihood of Misremembering False Details as 'True'



Note. SDO = Social Dominance Orientation
Low SDO: 0-1.91 (n = 84); *Med SDO*: 1.92 -2.73 (n = 77); *High SDO*: 2.74 -4.27 (n = 64).
Error bars represent standard error.
Includes warned participants only.

Discussion

Experiment 1 was an attempt to replicate findings of Gilbert et al. (1993) who report that people are inclined to believe whatever they first encounter as true, even if explicitly made aware that at least some of the information they are processing is false and should be disregarded accordingly. In that study, this tendency was demonstrated by participants assignment of longer prison sentences and harsher character assessments toward the suspect described by false exacerbating details as compared to the suspect described by false extenuating circumstances. Gilbert also reports that experiencing a concurrent interruption task at the time that information was encoded directly influenced the details that people remembered and reported as true—participants misremembered significantly more *False* items as 'true' and assigned longer prison sentences to suspects when the participant read false exacerbating details and experienced a concurrent task compared to when participants read this same information while uninterrupted.

The results of Experiment 1 are in agreeance with Gilbert and colleagues' underlying theoretical argument; across all measures and conditions, people consistently judged suspects described by exacerbating false details more harshly than suspects described by more trivial wrongdoings, despite receiving specific instruction to disregard these details. As was true in Gilbert's study there was no difference found between interrupted and uninterrupted participants' tendency to misremember *True* statements (i.e., either responding 'never' or 'false'). Additionally, there was a strong positive relationship found between the proportion of 'true' responses to *False* statements and the length of prison sentence assigned. This relationship was mirrored by the strong positive correlations found between this response tendency and all character assessment ratings for crimes with exacerbating details, as well as a strong *negative* correlation between this incorrect response tendency and the sentencing, liking, and danger

ratings when the false details extenuated the crime. As Gilbert notes, participants were explicitly told to ignore any statements presented as false, and so criticism that participants may have simply been responding 'true' to all *False* statements are weakened when one considers that this pattern is exactly as one would expect if false statements were remembered, believed, and ultimately referenced in the assessments of the suspects. Unlike Gilbert, however, this experiment did not reveal an increased likelihood of interrupted participants' misremembering *Foil* or *False* items as 'true' compared to their uninterrupted counterparts. This was a rather unexpected result as Gilbert reports quite a large difference between interrupted and uninterrupted groups in this regard, though it should be noted that Gilbert did not directly compare interrupted and uninterrupted groups in terms of any character assessments, which do not appear to differ as a result of this manipulation.

Another finding that is unique to Experiment 1 is the significant interaction effect found between the severity of the false information and whether the participant was warned about it. Across all character assessment measures, the presence of the warning served to alleviate the judgments assigned to the exacerbated suspect, however when the suspect was falsely absolved by the details, the warned participants judged him more harshly than the unwarned. This interaction suggests that those who received the warning did heed it to some extent, but that participants were unable to completely disallow these details from influencing their opinion and ultimately the punishment they assigned to each of the suspects.

The novel disfluency manipulation implemented here does appear to have had some effect on memory outcomes. Responding 'true' to *False* statements was by far the most common memory error (M = .39), which is in accordance with the proportions reported by Gilbert ($M_{interrupt}$ = .44, $M_{nointerrupt}$ = .23). Interestingly, however, those who encoded the false details in

the disfluent format were about 9% more likely to provide this incorrect response compared to participants who had encoded the details fluently. I suspect that this deviation from the original prediction is due to the alternating cases format being more difficult to process, meaning that those participants had to exert greater cognitive effort when reading through the initial reports. When a similar detail was presented at test, the heightened experience of fluent remembering was attributed to the veracity of the detail, without consideration of the source of its familiarity – that is, that it was a detail that was more challenging to encode and was to be disregarded. In order for this disfluency manipulation to be successful at deterring belief in the false details, the participants must also recall the source of the fluently experienced stimuli at test.

Experiment 1 also revealed the novel finding that *High* RWA significantly predicted the prison term participants assigned when the false details exacerbated the crime, but did not predict this outcome overall, or when the details extenuated the offense. This finding contributes evidence supporting the finding that *High* RWA individuals divert more attentional resources to negatively valenced stimuli, but also that this bias directly affects their decision-making. This finding was further supported in the results of the memory test; *High* RWAs were less accurate overall and had the greatest tendency to misremember *False* items as 'true', while *Low* RWAs demonstrate the opposite relationship with these outcomes. Though the relationship between prison sentencing and SDO levels was insignificant, the identical pattern of memory errors was found among the different levels of SDO. That is, the tendency to misremember *False* statements as 'true' increased with level of SDO, and level of RWA. These are novel findings as at the time of this writing, I am unaware of any previous studies specifically testing whether either of these personality constructs has any bearing on the likelihood using false information that is misremembered as true.

Experiment 2

Experiment 2 was an extension of Experiment 1, with the aim of investigating whether immediate repetition of fluently or disfluently presented false statements could serve as an effective cue to disregard these details in later tests of memory, and in participants' sentencing and character ratings of the two suspects. That is, perhaps repeating the false details in a disfluent manner might deter people from remembering details that were presented as false as being true, essentially functioning to 'correct' the detail as it is encountered by providing an additional reminder that this information should be disregarded. This experiment also identified the false information in the *disfluent* condition utilizing the same novel alternating-cases disfluency manipulation tested in Experiment 1, however in the *fluent* conditions, all red text was removed, and participants encoded all details of the crime reports in default 12-pt black text.

Method

Participants

Individuals who participated in Experiment 1 were not permitted to participate in Experiment 2. Two-hundred and ninety-four participants were recruited through the Department of Psychology Online Research Participation System at the University of Manitoba (Winnipeg, Canada) and received partial credit in their Introductory Psychology course (PSYC-1200) in exchange for their participation. The recruitment message was identical to that used in Experiment 1, and again participants were only told that they were participating in a study of memory for details of crime reports. Prior to participation, each participant signed an online consent form as required for approval by the University of Manitoba Fort Garry Research Ethics Board. Following completion of the study, seven participants (4 female, 2 male, and 1 unknown, $M_{age} = 20.46$) were excluded for having incomplete data. Prior to any data analysis, the same

exclusion criteria used in Experiment 1 was applied to the sample in Experiment 2. Adhering to said criteria removed an additional 14 participants resulting in a final sample of 273 participants in Experiment 2 (170 female, 101 male, 2 'other', M_{age} = 21.96, *SD* = 5.88, *range* = 17 – 52 years).

Design

As in Experiment 1, the presentation order of each suspect's crime (*Tom then Kevin* versus *Kevin then Tom*), the severity of the false statements (*better* versus *worse*), the presence of instruction to disregard any text indicated to be false (*no warning* versus *warning*), and whether the participant experienced the interruption task (*no interrupt* versus *interrupt*) was counterbalanced across all conditions. Unique to Experiment 2, 149 participants were randomly assigned to a *repeat* condition. In this condition, each item of false information was first presented in regular black text and then immediately repeated in either identical regular black text *(fluent-repeat* condition, *n* = 74), or the false information was repeated in the alternating case format introduced in Experiment 1 (*disfluent-repeat* condition, *n* = 75). The remaining participants were randomly assigned to the *no repeat* condition (n = 124). In this condition, all text was presented in regular black text, meaning that those assigned to the *fluent-no repeat* condition were never warned of any false information *(n* = 57), as for these participants there was no indication that any details should be encoded any differently from any other details in the report (i.e., there was no red or alternating-case text, and there was no repetition). This created a unique control condition untested in Experiment 1. For those participants assigned to *disfluent-no repeat* condition (*n* = 67), the experimental conditions were identical to those in the *disfluent* condition in Experiment 1. For these participants, the false details were presented in alternating cases, but were not repeated. As in Experiment 1, half of the participants assigned to the *disfluent*

condition received instruction that the differently presented information was false and was to be disregarded, while the others were not provided with this instruction. The random assignment of participants across all experimental conditions resulted in a complex mixed design treating severity of the false information and the order that the suspects were presented as within-subjects variables while implementing the same between-subjects variables as Experiment 1, with the additional between-subjects variable of repetition tested in Experiment 2.

Materials

All materials used in Experiment 2 were identical to those described in Experiment 1. Due to the restrictions of the COVID-19 pandemic at the time of data collection, all participants completed Experiment 2 online and using their personal desktop or laptop computer. The experiment was administered through Qualtrics online platform (2005), was not mobile-friendly, and all responses were made by participants in the exact same manner as Experiment 1.

Independent Variables

Repetition Condition

As mentioned, all independent variables from Experiment 1 remained unchanged but for the addition of the between-subjects variable of repetition. Following the initial introduction instruction delivered to all participants (see Conditional Instructions, Experiment 1), those participants assigned to the *disfluent–repeat -warn* condition received a slightly altered message from that provided in Experiment 1 which read; “You will notice that some of the statements are repeated in an odd-looking format. Any statement that is repeated in this alternating case format (e.g., ThE rObBeR hAd A gUn) should be considered FALSE. These details are from an unrelated crime report and were accidentally included, as sometimes can occur in a court of law.” Those assigned to the *fluent–repeat-warn* condition read the following instruction; “You

will notice that some of the statements are repeated. Any statement that is repeated in the crime reports should be considered FALSE…". The remaining instructions were identical to those provided in Experiment 1, according to the randomly assigned experimental condition of the participant. To illustrate, for a participant in the *disfluent-repeat* condition, the first sentence from Tom's crime report with extenuating false details read; "On the evening of January 16th, 2019, Tom left a neighbor's Winnipeg apartment where he had been playing card games PlAyInG cArD gAmEs most of the evening." For those in the *fluent-repeat* condition, the repeated details were presented in the same font and format as the rest of the sentence (i.e., "…where he had been playing card games playing card games most of the evening"). The dependent variables tested in Experiment 2, as well as the experimental procedure and total time to complete the study were identical to that described in Experiment 1.

Results

Data and Variable Exclusion

The same data exclusion criteria applied in Experiment 1 was also followed in Experiment 2, giving primary focus to identifying and removing multivariate outliers and problematic test-takers. First, a mahalanobis distance test identified two extreme multivariate outliers ($p < .001$), and both were excluded from analyses. Next, 10 participants were excluded from analyses for scoring below 30% (fewer than 9 of 30 correct responses) on *both* recognition memory tests. Lastly, examination of boxplots and extreme scores for univariate outliers identified and excluded two problematic test-takers, each of whom assigned 20 years in prison to both suspects and one also responding 'true' to all recognition memory test items.

As in Experiment 1, several mixed ANOVAs with repeated measures were conducted and confirmed that the order of report presentation nor the specifics of either crime report had any significant effect on the character ratings concerning the likeability of the suspect, $F(1, 272)$

= .71, *MSE* = 2.30, *p*= .40, η_p^2 = .003, how dangerous the participants perceived him to be, *F*(1, 272) = 2.48, *MSE* = 5.78, *p* = .12, η_p^2 = .009, or how likely they felt the suspect could be helped with counselling, *F*(1, 272) = 0.67, *MSE* = 2.78, *p* = .42, η_p^2 = .002. In addition, there was no indication that these variables had any significant effect on the prison sentences assigned to the perpetrators, *F*(1, 272) = .006, *MSE* = 0.18, *p* = .94, η_p^2 = .000. Collapsing across these variables resulted in approximately even groups across conditions. Ideally the entire sample size would have been comparable to that of Experiment 1, however time constraints and likely the time of year (i.e., data for this experiment was collected from mid-April until late-July of 2022, a term in which research participation can be slower and less predictable than during the regular academic terms) resulted in a final sample size that was notably smaller than originally intended. However, considering Gilbert et al.'s study only involved 71 participants in total, the smaller sample size achieved in this study is still comparable ($n_{warning}$ = 108), though the increased possibility of Type II error is acknowledged. Due to the smaller sample size and this increased susceptibility to error, several effects that approached significance (i.e., $p < .10$) are noted in the results that follow (Pituch & Stevens, 2015).

In terms of recognition memory test outcomes, again neither order nor suspect had any significant influence on overall memory performance, *F*(1, 272) = 3.26, *MSE* = 0.22, *p* = .07, η_p^2 = .012, or participants' tendency to misremember *False* or *Foil* items as 'true' (*False*: *F*(1, 106) = 0.13, *MSE* = 0.02, *p* = .72, η_p^2 = .001; *Foil*: *F*(1, 272) = 1.34, *MSE* = 0.04, *p* = .25, η_p^2 = .005). To determine whether severity of the false information had the same negligent effects on recognition memory outcomes, I again limited my analysis to include only those in the *warn* condition and conducted three additional repeated-measures ANOVAs, each confirming that there was no significant effect of severity on the same memory outcomes noted above (*accuracy*:

$F(1, 271) = 1.96$, $MSE = 0.02$, $p = .16$, $\eta_p^2 = .007$; *'true' to False*: $F(1, 107) = 1.14$, $MSE = 0.04$, $p = .06$, $\eta_p^2 = .003$; *'true' to Foils*: $F(1, 106) = 1.86$, $MSE = 0.01$, $p = .18$, $\eta_p^2 = .02$), though the effect of severity on the tendency to respond 'true' to *False* items approached significance.

Character Assessments and Prison Sentencing Scores

Likeability Ratings

As in Experiment 1, I conducted a series of one-way mixed ANOVAs, again treating the severity of the false information as the sole within-subjects factor and the fluency format, whether the participant received a warning, whether they experienced the interruption task, and whether the false information was repeated (*no repeat* versus *repeat*) as between-subjects factors. Once more keeping in mind that higher scores for all character assessment measures indicate a more negative perception of the associated suspect, the first mixed ANOVA of this series again revealed a significant effect of severity on liking ratings of the suspects, $F(1, 259) = 63.94$, $MSE = 175.34$, $p < .001$, $\eta_p^2 = .198$. Participants judged the suspect described by exacerbating false details as less likeable than the suspect described by extenuating details ($M_{worse} = 5.50$, $M_{better} = 4.28$, $SE = .14$, $p < .001$). Results also demonstrate a main effect of interruption, $F(1, 259) = 5.59$, $MSE = 16.96$, $p = .02$, $\eta_p^2 = .021$; participants who were not interrupted while reading the reports perceived the suspect as less likeable than those who were given the concurrent task ($M_{interrupt} = 4.71$, $M_{nointerrupt} = 5.07$, $SE = .15$, $p = .02$).

Danger Ratings

An identical mixed ANOVA also indicated a significant main effect of the severity of the false details on how dangerous participants perceived the suspects to be, $F(1, 259) = 146.13$, $MSE = 284.61$, $p < .001$, $\eta_p^2 = .361$. Again, the suspect who was described in a less positive light was rated as more dangerous than the suspect described by extenuating details ($M_{worse} = 5.42$,

M_{better} = 3.87, *SE* = .12, $p < .001$). The analysis also identified a main effect of interruption on perceptions of danger, $F(1, 259) = 6.10$, $MSE = 13.99$, $p = .01$, $\eta_p^2 = .023$. In a pattern similar to that reported for likeability ratings, those who did not experience an interruption rated suspects as more dangerous than those who did ($M_{nointerrupt} = 4.79$, $M_{interrupt} = 4.50$, $SE = .13$, $p = .03$).

The mixed ANOVA also indicated a significant interaction between the severity of the false information and whether the participant was warned, $F(1, 259) = 10.06$, $MSE = 19.60$, $p = .002$, $\eta_p^2 = .037$. Again, when the false details extenuated the crime, those in the *warning* condition judged the suspect as more dangerous than those in the *no warning* condition, however when the false details exacerbated the crime, this pattern was reversed (*better*: $M_{warning} = 4.08$, $M_{nowarning} = 3.71$, $SE = .20$, $p = .06$; *worse*: $M_{warning} = 5.11$, $M_{nowarning} = 5.65$, $SE = .16$, $p = .001$).

Likelihood to Benefit from Counselling Ratings

An identical mixed ANOVA indicated that the severity of the false information had a significant impact on participants' assessment of the extent that each suspect might be helped by counselling, $F(1, 259) = 11.72$, $MSE = 35.33$, $p < .001$, $\eta_p^2 = .043$. The suspect that was reported for the crime with exacerbating false details was perceived as less likely to benefit from counselling compared to the suspect that was reported for the crime with extenuating false details ($M_{worse} = 3.21$, $M_{better} = 2.73$, $SE = .15$, $p = .001$). An interaction effect identified between the severity of the false information, its fluency format, and whether participants were warned approached significance, $F(1, 259) = 3.59$, $MSE = 10.81$, $p = .06$, $\eta_p^2 = .014$. In all conditions, participants perceived the suspect described by the exacerbating details as less likely to be helped by counselling, however this difference was only significant when the participants were not warned and the difference in ratings was greater when the false information had been presented

disfluently (*fluent*: M_{worse} = 3.15, M_{better} = 2.63, *SE* = .26, *p* = .05; *disfluent*: M_{worse} = 3.49, M_{better} = 2.66, *SE* = .29, *p* = .005).

Prison Sentencing

To little surprise, the final analysis revealed a significant effect of severity on the length of prison sentence assigned, $F(1, 259) = 114.89$, $MSE = 1541.69$, $p < .001$, $\eta_p^2 = .307$. Participants assigned significantly longer sentences to suspects whose crimes were described by exacerbating false details than to suspects whose crimes were described by false extenuating details (M_{worse} = 8.91, M_{better} = 5.35, *SE* = .32, *p* < .001). There were no other main effects, though an effect of the fluency format approached significance, $F(1, 259) = 3.17$, $MSE = 87.29$, $p = .08$, $\eta_p^2 = .012$. Post-hoc pairwise comparisons revealed that those who encoded the false information in the *fluent* format assigned longer sentences, on average, than those who encoded the false details disfluently, (M_{fluent} = 7.72, $M_{disfluent}$ = 6.69, *SE* = .46, *p* = .02, $\eta_p^2 = .019$). Interestingly, a main effect of false format approached significance, $F(1, 259) = 3.17$, MSE = 87.29, $p = .08$, $\eta_p^2 = .012$. In addition, a two-way interaction between the fluency of the false details and whether the participant experienced an interruption approached significance, $F(1, 259) = 2.95$, $MSE = 81.26$, $p = .09$, $\eta_p^2 = .01$. To investigate this relationship, post-hoc comparisons revealed that there was no difference in prison sentence assigned between uninterrupted *fluent* and *disfluent* groups, however participants in the *interrupt* condition assigned significantly longer sentences when they encoded the false information fluently (M_{fluent} = 8.48, $M_{disfluent}$ = 6.41, *SE* =.64, *p* < .001). *Fluent* encoders also assigned longer prison sentences when they were interrupted as compared to when they were not ($M_{interrupt}$ = 8.48, $M_{nointerrupt}$ = 6.96, *SE* = .66, *p* = .02); no difference in prison sentence was found between interrupted and uninterrupted *disfluent* encoders ($M_{interrupt}$ = 6.41, $Mno_{interrupt}$ = 6.96, *SE* = .63, *p* = .38).

As in Experiment 1, there was a significant interaction between the severity of the false information and whether a warning was provided on the prison sentences assigned. $F(1, 259) = 5.21$, $MSE = 69.92$, $p = .02$, $\eta_p^2 = .020$. Further post-hoc comparison analysis revealed that in both the *no warning* and *warning* conditions, participants always assigned longer prison terms to the suspect whose report was exacerbated by the false details compared to the suspect whose crime was extenuated by them (*no warning*: $M_{worse} = 9.44$, $M_{better} = 5.08$, $SE = .41$, $p < .001$; *warning*: $M_{worse} = 8.20$, $M_{better} = 5.71$, $SE = .50$, $p < .001$). The nature of the interaction was also consistent; those in the *no warning* condition assigned longer prison sentences for the exacerbated crime and shorter sentences for the extenuated crime than warned counterparts.

Recognition Memory Test Performance

Overall accuracy on the recognition memory tests was again adequate ($M = .68$, $SD = .19$), though notably lower than the memory test accuracy reported in Experiment 1. In addition, overall accuracy on the interruption task was also high, ($M = .83$, $SD = .11$). Although the memory test score analyses that follows was also limited to only those who received a warning about the false details, results of an independent t-test prior to their exclusion indicated a significant difference in likelihood of reporting 'false' to *Foil* statements between the *warning* and *no warning* participants, $t(273) = 5.16$, $p < .001$, $d = .64$. That is, in the *no warning* condition, participants incorrectly responded 'false' in 24% ($SD = .19$) of the *Foil* cases which certainly contributed to the lower accuracy score overall. In comparison, the proportion of 'false' responses to *Foils* was only 13% ($SD = 0.15$) for those who received the warning instruction.

In Comparison to Gilbert's Interruption Findings

To identify whether the interruption had similar effects on recognition memory scores as those reported by Gilbert et al. (1993), I conducted an identical MANOVA to that performed in

Experiment 1, testing the same dependent variables and assigning interruption as the between-subjects factor. The analysis did not reveal any significant effect of interruption on participants' likelihood of incorrectly responding to *True* statements (*'never'* response: $F(1, 106) = 3.69$, $MSE = 0.11$, $p = .06$, $\eta_p^2 = .034$; *'false'* response: $F(1, 106) = 2.10$, $MSE = 0.03$, $p = .15$, $\eta_p^2 = .019$), their likelihood of responding 'true' to *False* statements, $F(1, 106) = 0.14$, $MSE = 0.01$, $p = .71$, $\eta_p^2 = .001$), or the likelihood of responding 'true' to *Foil* statements, $F(1, 106), = 3.36$, $MSE = 0.06$, $p = .07$, $\eta_p^2 = .031$. In addition, a repeated-measures ANOVA revealed no significant difference in the length of prison sentence assigned between the interrupted and uninterrupted groups, whether the false details extenuated, $F(1, 106) = 0.09$, $MSE = 1.31$, $p = .77$, $\eta^2 = .001$, or exacerbated the crime, $F(1, 106) = 1.39$, $MSE = 36.76$, $p = .24$, $\eta^2 = .013$. Gilbert's reported effects of interruption on memory for false information are not matched here, however I did find consistency with Gilbert's reported correlational findings between the average tendency to misremember *False* items as 'true' and the length of the prison sentence, $r(108) = .27$ ($p = .005$), the liking ratings, $r(108) = .41$ ($p < .001$), and the danger ratings, $r(108) = .44$ ($p < .001$), in cases where the false details exacerbated the crime, $r(108) = .27$, ($p = .005$).

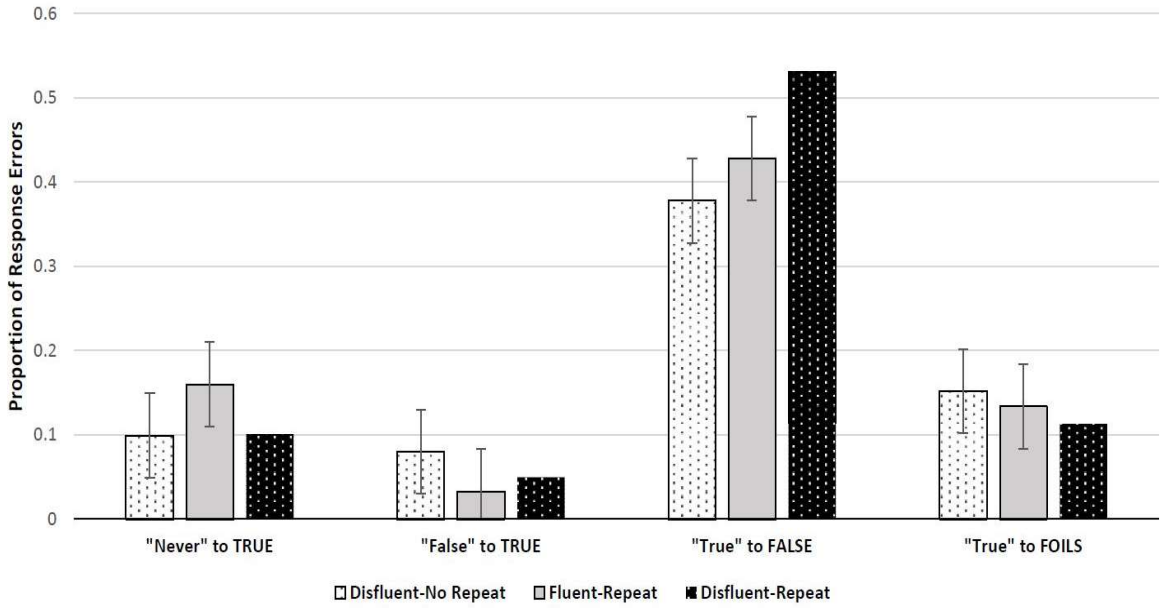
Effects of Repetition and False Format on Memory and Sentencing Outcomes

To determine any main effects of either the novel manipulations introduced in this study, I conducted a two-way MANOVA assigning repetition and the false format as between-subjects variables and tested the same four memory outcome dependent variables, with the addition of overall memory accuracy. Given there were no warned participants in the *fluent-no-repeat* condition, the exclusion of unwarned participants and also of interruption as a between-subjects variable resulted in three experimental groups: *fluent-repeating* ($n = 38$), *disfluent-repeating* ($n = 38$), and *disfluent-not-repeating* ($n = 32$). The MANOVA revealed a significant main effect of

repetition on participants' likelihood of responding 'true' to *False* statements, $F(1, 105) = 4.51$, $MSE = 0.40$, $p = .04$, $\eta_p^2 = .041$. Further post-hoc pairwise comparisons indicated that those who read the false information *disfluently* and *repeatedly* had a significantly higher tendency of misremembering *False* items as 'true' than when the false information was disfluent and was not repeated ($M_{repeat} = .53$, $M_{norepeat} = .38$, $SE = .07$, $p = .04$). In addition, among participants who received repeating misinformation, those who encoded the false information *fluently* were 10% less likely to respond 'true' to *False* statements than their *disfluent* encoding counterparts, however this difference was not significant ($M_{fluent} = .43$, $M_{disfluent} = .53$, $SE = .07$, $p = .14$). This finding suggests that repeating the false information disfluently actually results in a greater

Figure 5

Interaction Effects Between False Format and Repetition on Tendency to Misremember False Details as 'TRUE'



Note. Proportions represent average proportion of responses.
Perfect memory score: 19 correct 'never' responses, 4 correct 'true' responses, 7 correct 'false' responses.
Error bars represent standard errors.
Includes warned participants only.

proportion of critical memory errors compared to fluent repetition, or only encoding the disfluent false details a single time (see **Figure 5**).

Assessing RWA and SDO as Potential Moderating Variables

The same hypotheses involving RWA and SDO scores was also tested in Experiment 2. As in Experiment 1, both the RWA scale (15 items, $\alpha = .79$) and the SDO scale (22 items, $\alpha = .92$) demonstrate high internal consistency and were significantly correlated, $r(270) = .36$ ($p < .001$) (Zakrisson, 2005).

Gender Differences

To determine whether the RWA and SDO data collected in Experiment 2 revealed a similar gender difference to that reported in previous research (and identified in Experiment 1), I conducted an identical MANOVA which again indicated a main effect of gender on SDO scores, $F(1, 267) = 12.28$, $MSE = 9.01$, $p < .001$, $\eta_p^2 = .08$. Noting a difference in group sizes ($n_{female} = 168$, $n_{male} = 100$), post-hoc pairwise analysis demonstrated the expected pattern among SDO scores; male SDO scores were significantly higher than female SDO scores ($M_{males} = 2.79$, $M_{females} = 2.27$, $SE = .11$). There was no main effect of gender on RWA scores found in Experiment 2, $F(1,267) = 1.60$, $MSE = .97$, $p = .20$, $\eta_p^2 = .012$.

Analysis of Right-Wing Authoritarianism Scores

Prior to further data analysis, I implemented tercile splits to each dataset resulting in three approximately equal groups (*Low, Med* and *High*) for each set of scores. The ranges for RWA scores were as follows: *Low*: 0 – 3.27 ($n = 98$), *Med*: 3.28 – 3.87 ($n = 88$), *High*: 3.88 – 5.46 ($n = 84$). Ranges for SDO scores were as follows: *Low*: 0 – 2.00 ($n = 94$), *Med*: 2.01 – 2.77 ($n = 88$), *High*: 2.78 – 4.63 ($n = 91$). As in Experiment 1, these cut-offs were well within range of those implemented in previous studies (e.g., Son Hing, 2007). Limiting the analysis of RWA scores to

include only warned participants resulted in roughly equal groups across all RWA levels (n_{low} = 41, n_{med} = 38, n_{high} = 27).

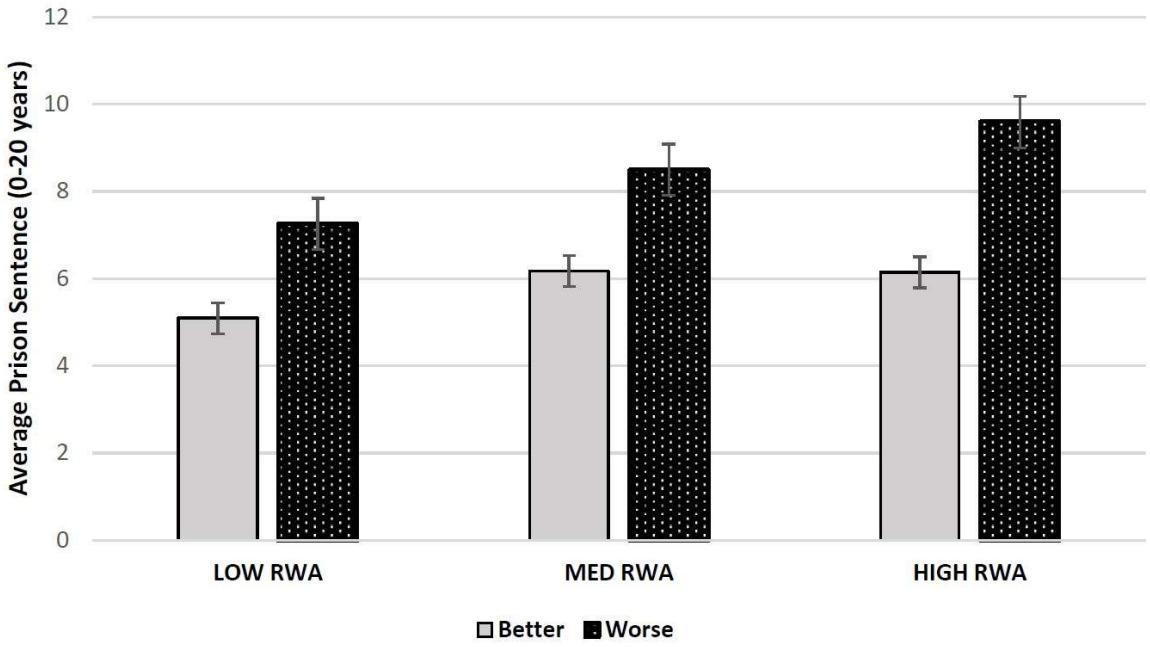
It is noted that implementing tercile splits on continuous data is not recommended as these methods inevitably result in a decrease of power due to the large amount of data loss at the cut-off values. For instance, in this study, those who scored an average of 3.27 in RWA would be categorized as *Low* RWA, while those scoring less than a full point higher would be categorized as having *High* RWA. While not ideal, future investigations should consider multivariate fractional polynomials (Sauerbrei, 2006) for analysis of RWA and/or SDO scores as opposed to utilizing categorical splits as I have here.

As in Experiment 1, I was interested in understanding the extent to which RWA might predict the length of prison sentence assigned, and by extension might also predict one's tendency to misremember *False* items as 'true'. I first conducted a one-way MANOVA setting RWA as the independent variable with three categorical levels (*Low, Med* and *High* RWA) and testing the outcomes of average length of prison sentence assigned, prison sentences given to suspects when false details exacerbated the crime, and sentences when these details extenuated the crime. Unlike Experiment 1, results of this MANOVA indicated no significant differences among RWA levels and the average length of prison sentence assigned, $F(2, 103) = 1.75$, $MSE = 26.42$, $p = .18$, $\eta_p^2 = .033$, nor the length of sentence when details extenuated, $F(1, 103) = 1.00$, $MSE = 14.53$, $p = .37$, $\eta_p^2 = .019$, or exacerbated the crime, $F(1, 103) = 1.75$, $MSE = 45.66$, $p =$

.18, $\eta_p^2 = .033$. Although non-significant, *High* RWAs did assign longer prison sentences than *Low* RWAs in all cases mentioned (see **Figure 6**).

Figure 6

Level of RWA Influences Length of Prison Sentence for Exacerbated Crimes (Exp 2)



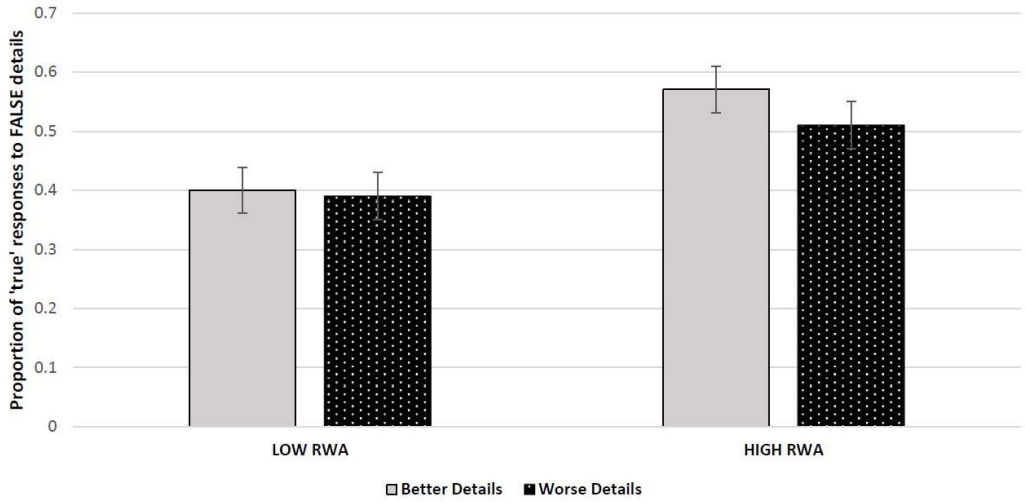
Note. RWA = Right-Wing Authoritarianism.
Low RWA: 0-3.27 (n = 41); *Med RWA:* 3.28 -3.87 (n = 38); *High RWA:* 3.88 -6.07 (n =27).
Error bars represent standard errors.
Includes warned participants only.
p (worse) = .18

A separate MANOVA revealed no significant differences among RWA levels and the likelihood of misremembering false details as true, however once again, those in the *High* RWA group were approximately 10 -15% more likely than *Low* RWA to respond 'true' to *False* test items overall, when the details exacerbated the crime, and when false details extenuated the crime. Correlational analysis revealed no significant relationship between RWA level and the average tendency to respond 'true' to *False* statements, though *High* RWA was positively

associated with this response, $r(108) = .17$, $p = .08$, and *Low* RWA was negatively correlated with this tendency, $r(108) = -.15$, $p = .13$ (see **Figure 7**).

Figure 7

Level of RWA Influences Tendency to Misremember False Details as True



Note. RWA = Right-Wing Authoritarianism.
Low RWA: 0-3.27 (n = 41); *High RWA*: 3.88 -6.07 (n =27).
Error bars represent standard errors.
Includes warned participants only.
p(avg) = .15; *p*(better) = .10; *p*(worse) = .34

Analysis of Social Dominance Orientation Scores

An identical MANOVA to that performed for levels of RWA was conducted on SDO levels and similarly revealed no significant effect of SDO level on average prison sentence assigned, $F(2, 105) = 0.63$, MSE = 9.76, $p = .53$, $\eta_p^2 = .012$, or length of sentence assigned when the false details exacerbated, $F(2, 105) = 0.80$, $MSE = 21.33$, $p = .45$, $\eta_p^2 = .015$, or extenuated the crime, $F(2, 105) = 0.94$, $MSE = 13.56$, $p = .40$, $\eta_p^2 = .018$. A second MANOVA revealed no significant effect of SDO level on proportion of 'true' responses *to False* test items, though as was true for RWA scores, *High* SDOs were approximately 10 – 15% more likely to respond

'true' to *False* statements than *Low* SDOs on average, and regardless of whether the false details extenuated or exacerbated the crime. Similar to RWA, correlational analysis revealed that SDO level was positively associated with the tendency to respond 'true' to *False* items, but only when the false details were exacerbating, $r(108) = .19$, $p = .045$.

Discussion

Experiment 2 introduced the variable of repetition as a possible additional cue for participants to disregard details presented as false. In agreeance with Gilbert's original argument and the results of Experiment 1, Experiment 2 also demonstrates that people are unable to entirely disallow influence from previously encountered false details as evidenced by the main effects of severity across all measures and conditions. Again, participants assigned significantly harsher prison sentences and character assessments to the suspect whose crime report included exacerbating false details compared to the suspect described by extenuating circumstances. Experiment 2 also exhibits the exact same significant interaction pattern between the severity of the false details and whether the participant was warned that these details were false; those who were warned generally provided less harsh assessments for crimes with exacerbating false details and harsher assessments to the suspects with exacerbating crimes than unwarned participants. Again, this suggests that although false details ultimately informed participants' final assessments, the warning was likely heeded to some extent by those in that condition.

This second experiment also revealed slightly more influence from the interruption task compared to tests conducted in Experiment 1. For instance, a main effect of interruption was found for both the liking and danger ratings assigned to suspects; those who were uninterrupted rated the suspects as less likeable and more dangerous than their interrupted counterparts, a pattern opposite to Gilbert's claim. In addition, the interaction between the fluency of the false

details and the interruption on prison terms suggests that the interruption primarily affected fluent responders; interrupted *fluent* encoders assigned longer sentences than interrupted *disfluent* encoders, as well as longer sentences than uninterrupted *fluent* encoders. No difference in prison sentencing was found between interrupted and uninterrupted *disfluent* encoders.

In terms of recognition memory test results, the significant positive correlations identified between the tendency to respond 'true' to *False*, the length of prison term assigned, and the character assessments again suggest that the greater the likelihood of misremembering false information as true, the greater the tendency for these details to influence judgments toward the suspects. Contrary to Gilbert's argument that the interruption task resulted in participants misremembering a greater proportion of *False* items as true, Experiment 2 found little evidence of any effect of interruption on memory performance, regardless of the severity of the false details. Interestingly, however, the interaction between the format of the false information and whether it was repeated did seem to affect this memory error; those who encoded the false details disfluently and repeatedly were significantly more likely to misremember *False* items as 'true' compared to when details were only presented disfluently a single time. In addition, of participants for whom the false details were repeated, those who read the details disfluently were approximately 10% more likely to misremember *False* statements as true compared to those who processed the details fluently. This finding was not significant, though it suggests that disfluency, coupled with repetition, may not have the corrective influence initially proposed.

Although the findings concerning RWA and SDO personality constructs were underwhelming in comparison to Experiment 1, the patterns among *High* RWA individuals and *High* SDO individuals compared to *Low* RWA and *Low* SDO groups are consistent. Overall, those high in these traits assigned longer prison sentences and were 10-15% more likely to

incorrectly respond 'true' to *False* test items compared to their *Low* RWA and *Low* SDO counterparts, regardless of the severity of the false details.

General Discussion

The present study introduces a novel method to inform that information is false, while simultaneously ruining the processing fluency with which false information is processed and, hopefully, individuals' later reliance on this information in memory and judgment tasks.

As this study confirms, people are generally not very good at disregarding information that they are told is false. As Gilbert et al. (1993) assert, "believing what one reads seems so easy…that people must take great pains to guard against it" (p. 221). Two experiments found that in all conditions participants assigned harsher character assessments and longer prison terms to suspects when they were exposed to false details that, if believed, worsened the suspects' crime considerably compared to the story without any false details. Similarly, in all conditions, participants judged the suspect aligned with extenuating details less negatively.

In both experiments, these patterns were mirrored in the strong correlations identified between the likelihood of misremembering *False* items as 'true' and the character assessments of the suspects; exacerbating details shared a positive relationship with this tendency while a negative association was found between this tendency and extenuating false details. These findings suggest that these memory errors have a direct influence on the decisions made—namely, the punishment assigned to criminals. The memory test results are particularly interesting as each of the participants included in those analyses received explicit warning of the presence of false information, specific instruction to disregard it, *and* all were informed of the upcoming memory test—three aids working in favor of discounting the false information still were not enough to deter people from misremembering false information as true.

Another novel and consistent finding was the recurring interaction between receiving a warning and instruction to disregard false details and the severity of those details included in the crime report. Overall, warned participants judged the suspect described by the exacerbating details less harshly than unwarned participants. However, when the details extenuated the crime, the warned participants judged the associated suspect more harshly than the unwarned participants. If these warnings are indeed heeded to some extent, then it is curious to note that those in the warning condition were *harsher* on the suspects described by extenuating details than those who were unwarned. However, if participants who were warned that the more distinctively presented details were false were also drawn to the theme that these details were painting the criminal in a more flattering light, then this finding is consistent with the idea that participants opted to assign harsher critiques to suspects to compensate for this realization, compared to those who were unwarned. Warned participants recalling and actively working to disregard the false information would follow this response pattern, under the assumption that the *actual* details must be worse than those recalled as false, in a sense intuitively correcting for the plausible implicit understanding that the false details were an effort to 'excuse' the suspect.

According to Gilbert, those who were uninterrupted were twice as accurate when identifying false statements, and half as likely to respond 'true' to *False* statements compared to their uninterrupted counterparts. These differences were not identified in either Experiment 1, or Experiment 2. In Gilbert's study, as in both experiments presented here, the most common memory error was responding 'true' to *False* statements (M_{Exp1} = .39, M_{Exp2} = .45), though notably these proportions were more similar to the *interrupt* condition than the *no interrupt* condition in Gilbert's study. It is also curious that Gilbert et al. would find such a stark difference among the groups, especially when the fact that the 'uninterrupted' participants in this

study still witnessed the scrolling digits beneath the crime report text (but were told to ignore them) is considered. Since Gilberts' uninterrupted group *also* had false detail warnings to heed while not having the option to truly avoid the concurrent digit-scan task, one could make the argument that the 'uninterrupted' group was still distracted (though perhaps not to the same extent as the 'interrupted' group). In the present study, however, there was a definite difference between the encoding experiences of interrupted and uninterrupted participants.

The novel disfluency manipulation tested in these experiments also appears to have an effect on memory performance for false details. Though the effect was small Experiment 1 revealed that those who encoded false information disfluently were 9 - 10% more inclined to incorrectly respond 'true' to *False* judgments than fluent encoders. Similarly, in Experiment 2, those in the *disfluent-repeat* condition were more likely to respond 'true' to *False* statements compared to both *disfluent-no repeat* condition and the *fluent-repeat* condition. These findings are intriguing and warrant further investigation. It appears that the disfluency manipulation made the false details slightly more memorable to participants, though there was no real significant difference identified between disfluent and fluent responders in terms of prison sentences or character assessments assigned, and so the extent that disfluent false details affect decisions and behaviour remains unclear. It does appear, however, that repeating false statements in the disfluent format as tested here does not function well as a corrective technique for misinformation, but instead rather the opposite. This is useful knowledge in the context of developing techniques to effectively fact-check in real time, and a rich starting point for future investigations testing disfluency manipulations as a strategy to steer attention toward the veracity of claims at the same time that they are initially encountered.

This study offered a unique opportunity to test whether those scoring high in RWA and SDO personality constructs might be more inclined to misremember *False* statements as 'true', and whether they might also assign harsher judgments to the suspects as a result. These hypotheses were generally supported, particularly for individuals scoring high in RWA. Though Experiment 2 results were insignificant, the pattern of memory errors was consistent across both experiments—*High* RWA individuals misremembered more *False* memory test items as 'true' than their *Low* RWA counterparts, and this pattern was matched in their sentencing and harsher character assessment judgements toward the suspects.

To my knowledge, this is another novel finding within the misinformation literature as well as adds to the RWA and SDO research available. Though it is apparent that no participant was immune to the influence of false details, scoring high on the RWA scale was a significant predictor of how likely a person was to misremember a false detail as being true, and consequently how long of a prison sentence would be assigned, particularly for crimes worsened by the false information. These findings also potentially help to explain some of the misinformed beliefs that took hold during the pandemic, and those that rage on today. As mentioned, the combination of personality and ideological beliefs (e.g., RWA), threat and uncertainty, the echo chambers of 'news' stories rampant on social media, and a world leader determined to shout blatantly false claims from the rooftops is a Molotov cocktail for the spread of and belief in misinformation which, as we have witnessed, can lead to extreme and devastating outcomes.

This finding is also interesting in the context of Gilbert's original experiments, which were conducted in Austin, TX in the early 1990's, a state that hadn't elected a democratic presidential candidate since Jimmy Carter in 1976 (Montgomery, 2020). Given that these experiments found no differences resembling anything reported by Gilbert, it raises the question

of whether there may have been other variables affecting participant's responses that were unaccounted for, though this consideration could still not account for the large differences in prison sentences and memory errors reported between Gilbert's interrupted and uninterrupted groups.

There were several limitations to this study—conducting both experiments entirely virtually as opposed to in the lab during the two years of the global pandemic is one drawback. Having participants complete the study without any supervision makes it difficult to gauge how seriously participants took the study, and how distracted they truly were while completing the experiment at home. In addition, the time of year that the data was collected for Experiment 2 may have been a factor in the resulting smaller-than-ideal sample size, though fortunately the sample size obtained still fell within a reasonable range of other comparable studies reported here (e.g., Gilbert et al., 1993; Jalbert et al., 2020).

In future studies, it may be beneficial to update the crime reports to add more relevance to the current time. For instance, it would likely cause quite an uproar to hear that a crime described by the exacerbating details used here— a man robbing a convenience store for $210 dollars or mugging a disabled veteran and uttering insulting words before fleeing the scene—had sent a person to prison for 20 years, at least in Winnipeg, Manitoba Canada in 2022. Also, the rather ambiguous, and potentially culturally biased question regarding the suspects' '*likelihood to be helped by counselling'* could be altered to instead ask how likely participants felt the suspect were to re-offend following the punishment they assigned. This would be an improvement to the measure of the perception of the suspect's rehabilitation potential.

Additionally, the *interrupt* condition could be adjusted to increase the level of distraction by instead involving a more ecologically valid interruption that people experience everyday as

they consume information, such as having to respond to text messages as they read the reports, for example. Having to report whether a digit repeated is a rather simple task, particularly for skilled readers. Changes to the interruption task, as well as adjusting the crime details and the punishments available for participants to assign to the suspects would help to improve the ecological validity and generalizability of these findings.

Similarly, adjustments to the disfluency manipulation such as intensifying it by implementing alternating-alternating cases (i.e., changing the text from TeSt to tEsT with each changing second of the words' presentation time) might help to better understand its effects, if any, not only for memory outcomes and assessments of the criminals, but also its effectiveness (or lack thereof) in deterring people from believing things that they are told to disregard as false. As noted, the research presented in this study provides a rich starting point for investigating many further questions about the influence that disfluency, particularly that embedded within a message itself, might have on our ability to discern fact from fiction.

In conclusion, the two experiments reported here support the claim that people experience great difficulty when trying to remember whether a piece of information was presented as true or false, regardless of the degree of fluency in which it is presented. Warnings seem to help to an extent, though only modestly and not when the false information is geared towards alleviating judgments by painting an accused in a more flattering light. These findings have many implications relevant to the challenging times we currently find ourselves in—a world where imperfect memory systems and global information machines designed to take advantage of our collective cognitive vulnerabilities have come to a head. While the cards may seem to be stacked against the truth-seekers, there is hope that the more knowledge and

awareness we can build concerning these vulnerabilities in relation to misinformation, the better chance there is of a more well-informed public, no matter what we read.

References

Altemeyer, B. (1981). *Right-Wing Authoritarianism*. University of Manitoba Press.

Altemeyer, B. (1998). The other "authoritarian personality". *Advances in Experimental Social Psychology, 30*, 47-92. https://doi.org/10.1016/S0065-2601(08)60382-2

Altemeyer, B. (2004). Highly dominating, highly authoritarian personalities. *The Journal of Social Psychology, 144*(4), 421-447. https://doi.org/10.3200/SOCP.144.4.421-448

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning; *Journal of Experimental Psychology: General, 136*(4), 569-576. https://doi.org/10.1037/0096-3445.136.4.569

Alter, A. L. (2013). The benefits of cognitive disfluency. *Current Directions in Psychological Science, 22*(6), 437 – 442. https://doi.org/10.1177/0963721413498894

Bakshy, E., Rosenn, I., Marlow, C., Adamic, L. (2012). The role of social networks in information diffusion. *Proceedings of the 21st International Conference on World Wide Web,* 519-528. https://doi-org.uml.idm.oclc.org/10.1145/2187836.2187907

Berger, J. (2011). Arousal increases social transmission of information. *Psychological Science, 22*(7), 891-893. https://doi.org/10.1177/0956797611413294

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,* 991-1013. https://www.nber.org/papers/w9873

Bilewicz, M., Soral, W., Marchlewska, M., & Winiewski, M. (2017). When authoritarians confront prejudice. Differential effects of SDO and RWA on support for hate-speech prohibition. *Political Psychology, 38*(1), 87-99. https://doi.org/10.1111/pops.12313

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. J. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56-64). Worth Publishers.

Brashier, N. M., & Marsh, E. J. (2020). Judging truth. *Annual Review of Psychology, 71*, 499-515. https://doi.org/10.1146/annurev-psych-010419-050807

Bresnahan, M., Zhu, Y., Hooper, A., Hipple, S., & Savoie, L. (2022). The negative health effects of anti-Asian stigma in the U.S. during COVID-19. *Stigma and Health.* Advance online publication. https://doi.org/10.1037/sah0000375

Bode, L. & Vraga, E. K. (2015). In related news, that was wrong: the correction of misinformation through related stories functionality in social media. *Journal of Communication, 65*(4), 619-638. https://doi.org/10.1111/jcom.12166

Capps, J. S. (2002). Explaining punitiveness: Right-wing authoritarianism and social dominance. *North American Journal of Psychology, 4*(2), 263 – 278. https://uml.idm.oclc.org/login

Carraro, L., Castelli, L., Macciella, C. (2011). The automatic conservative: Ideology-based attentional asymmetries in the processing of valenced information. *PloS One, 6*(11), e26456. https://doi.org/10.1371/journal.pone.0026456

Conway, M. A. (1995). *Flashbulb memories.* Erlbaum Associates.

Daley, D. J., & Kendall, D. G. (1964). Epidemics and rumors. Retrieved Aug 5, 2022 from https://www.nature.com/articles/2041118a0

Dechene, A., Stahl, C., Hansen, J., & Wanke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review, 14*(2), 238-257. https://doi-org.uml.idm.oclc.org/10.1177/1088868309352251

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58,* 17-22. https://doi.org/10.1037/h0046671

De Keersmaecker J, Dunning D, Pennycook G, et al. (2020). Investigating the Robustness of the Illusory Truth Effect Across Individual Differences in Cognitive Ability, Need for Cognitive Closure, and Cognitive Style. *Personality and Social Psychology Bulletin*. 2020;46(2):204-215. doi:10.1177/0146167219853844

Dharawat, A., Lourentzou, I., Morales, A., Zhai, C. (2022). Drink bleach or do what now? A study of risk-informed health decision making in the presence of COVID-19 misinformation. *Proceedings of the Sixteenth International AAAI Conference on Web and Social Media (ICWSM).* https://arXiv.org

Dodd, M. D., Balzer, A., Jacobs, C. M., Gruszczynski, M. W., Smith, K. B., Hibbing, J. R. (2012). The political left rolls with the good and the political right confronts the bad: connecting physiology and cognition to preferences. *Philosophical Transactions of the Royal Society B, 367*, 640-649. https://doi.org/doi:10.1098/rstb.2011.0268

Dunlosky, J., & Mueller, M. L. (2016). Recommendations for exploring the disfluency hypothesis for establishing whether perceptually degrading materials impacts performance. *Metacognition Learning, 11,* 123-131. https://doi:10.1007/s11409-016-9155-9

Dyrendal, A., Kennair, L. E. O., & Bendixen, M. (2021). Predictors of belief in conspiracy theory: The role of individual differences in schizotypal traits, paranormal beliefs, social dominance orientation, right wing authoritarianism and conspiracy mentality. *Personality and Individual Differences, 173,* 110645. https://doi.org/10.1016/j.paid.2021.110645

Ecker, U. K. H., Lewandowsky, S., Chang, E. P., Pillai, R. (2014). The effects of subtle misinformation in news headlines. *Journal of Experimental Psychology: Applied, 20*(4), 323-335. https://doi.org/10.1037/xap0000028

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews, 1*, 13 -29. https://doi.org/10.1038/s44159-021-00006-y

Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition, 38*, 1087-1100. https://doi.org/10.3758/MC.38.8.1087

Evans, J. St. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning, 11*(4), 382-389. https://doi.org/10.1080/13546780542000005

Fazio, L. K. (2020). Repetition increases perceived truth even for known falsehoods. *Collabra: Psychology, 6*(1): 38. https://doi.org/10.1525/collabra.347

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General, 144,* 993-1002. https://doi.org/10.1037/xge0000098

Fazio, L. K., Rand, D. G., Pennycook, G. (2019). Repetition increases perceived truth equally for plausible and implausible statements. *Psychonomic Bulletin Review, 26,* 1705 -1710. https://doi.org/10.3758/s13423-019-01651-4

Fogg, B.J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J. & Trauber, E. R. (2003). *How do users evaluate the credibility of Web sites? A study with over 2,500 participants.* Paper presented at the Proceedings of the 2003 Conference on Designing for User Experiences, San Francisco. https://doi.org/10.1145/997078.997097.

Forgas, J. P., Laham, S. M., & Vargas, P. T. (2005). Mood effects on eyewitness memory: Affective influences on susceptibility to misinformation. *Journal of Experimental Social Psychology, 41*(6), 574-588. https://doi.org/10.1016/j.jesp.2004.11.005

Frenda, S. J., Nichols, R. M., & Loftus, E. F. (2011). Current issues and advances in misinformation research. *Current Directions in Psychological Science, 20*(1), 20-23. https://doi.org/10.1177/0963721410396620

Frenkel, S., & Alba, D. (2020). Misleading virus video, pushed by the Trumps, spreads online. *The New York Times*. https://www.nytimes.com/2020/07/28/technology/virus-video-trump.html

Friedman, J. P., & Jack, A. I. (2018). What makes you so sure? Dogmatism, fundamentalism, Analytic thinking, perspective taking and moral concern in the religious and nonreligious. *Journal of Religion and Health, 57*(1), 157-190. https://doi.org/10.1007/s10943-017-0433-x

Geller, J. & Peterson, D. (2021). Is this going to be on the test? Test expectancy moderates the disfluency effect with sans forgettica. *Journal of Learning, Memory, and Cognition, 47*(12), 1924-1938. https://doi.org/10.1037/xlm0001042

Ghorbani, H. (2019). Mahalanobis distance and its application for detecting multivariate outliers. *Facta Universitatis, 34*(3), 583-595. https://doi.org/10.22190/FUM11903583G

Gilbert D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology, 59,* 601-613. https://doi.org/10.1037//0022-3514.59.4.601

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology, 65*(2), 221-233. https://doi.org/10.1037/0022-3514.65.2.221

Gottfried, J., & Shearer, E. (2016). *News use across social media platforms 2016*. http://journalism.org/2016/05/26/news-use-across-social-media-platforms-2016

Gover, A. R., Harper, S. B., Langton, L. (2020). Anti-Asian hate crime during the COVID-19 pandemic: Exploring the reproduction of inequality. *American Journal of Criminal Justice, 45*, 647-667. https://doi.org/10.1007/s12103-020-09545-1

Green, T. C., & Jame, R. (2013). Company name fluency, investor recognition and firm value. *Journal of Financial Economics, 109,* 813-834. https://doi.org/10.1016/j.jfineco.2013.04.007.

Guo, Z., Schlichtkrull, M., & Vlachos, A. (2022). A survey on automated fact-checking. *Transactions of the Association for Computational Linguistics, 10*, 178-206. https://doi.org/10.1162/tacl_a_00454

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Behavior, 16,* 107-112. https://doi.org/10.1016/S0022-5371(77)80012-1.

Heaven, P. C. L., & Bucci, S. (2001). Right-wing authoritarianism, social dominance orientation and personality: analysis using the IPIP measure. *European Journal of Personality, 15*(1), 49-56. https://doi.org/10.1002/per.389.

Henderson, E. L. Simons, D. J., Barr, D. J. (2021). The trajectory of truth: A longitudinal study of the illusory truth effect. *Journal of Cognition, 4*(1), 1-23. https://doi.org/10.5334

Henderson, E. L., Westwood. S. J. & Simons, D. J. (2022). A reproducible systematic map of research on the illusory truth effects. *Psychonomic Bulletin & Review, 29,* 1065-1088. https://doi.org/10.3758/s13423-021-01995-w

Hintzman, D. L. (2010). How does repetition affect memory? Evidence from judgments of recency. *Memory & Cognition, 38*, 102-115. https://doi.org/10.3758/MC.38.1.102

Hughes, H. C., & Waismel-Manor, I. (2020). The Macedonian Fake News Industry and the 2016 US Election. *PS: Political Science & Politics, 54*(1), 19-23. https://doi.org/10.1017/S104909650000992

Jalbert, M., Schwarz, N., & Newman, E. (2020). Only half of what I'll tell you is true: Expecting to encounter falsehoods reduces illusory truth. *Journal of Applied Research in Memory and Cognition, 9,* 602-613. https://doi.org/10.1016/j.jarmac.2020.08.010.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110*(3), 306-340. https://doi.org/10.10373445.110.3.306

Jacques, P. J., Dunlap, R. E., & Freeman, M. (2008). The organisation of denial: Conservative think tanks and environmental scepticism. *Environmental Politics, 17*(3), 349-385. https://doi.org/10.1080/09644010802055576

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1420-1436. https://doi.org/10.1037/0278-7393.20.6.1420

Kim, H. & Walker, D. (2020). Leveraging volunteer fact checking to identify misinformation about COVID-19 in social media. *The Harvard Kennedy School Misinformation Review: Special Issue on COVID-19 and Misinformation, 1.* https://doi.org/10.37016/mr-2020-021

Knowlton, B. J., & Squire, L. R. (1995). Remembering and knowing: Two different expressions of declarative memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 699-710. https://doi.org/10.1037/0278-7393.21.3.699

Koch, A., & Forgas, J. P. (2012). Feeling good and feeling truth: The interactive effects of mood and processing fluency on truth judgments. *Journal of Experimental Social Psychology, 48*(2), 481-485. https://doi.org/10.1016/j.jesp.2011.10.006

Lanius, C., Weber, R., Mackenzie, W. I. (2021). Use of bot and content flags to limit the spread of misinformation among social networks: a behavior and attitude survey. *Social Network Analysis and Mining, 11*(1), 32-32. https://doi.org/10.1007021-00739-x.

Laham, S. M., Koval, P., & Alter, A. (2012). The name-pronunciation effect: Why people like Mr. Smith more than Mr. Colquhoun. *Journal of Experimental Social Psychology, 48*, 752-756. https://doi.org/10.1016/j.jesp.2011.12.002

Larsen, A. O. (2014). The trouble with amicus facts. *Virginia Law Review, 100*(8), 1757-1818. https://heinonline.org

Lee, A. Y., & Labroo, A. A. (2004). The effect of conceptual and perceptual fluency on brand evaluation. *Journal of Marketing Research, 41*(2), 151-165. https://doi.org/10.1509/jmkr.41.2.151.28665

Levine, T. R. (2022) Truth-default theory and the psychology of lying and deception detection. *Current Opinion in Psychology, 47*(101380). https:doi.org/10.1016/j.copsyc.2022.101380

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition, 6,* 353-369. https://doi.org/10.1016/j.jarmac.2017.07.008

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*(3), 106-131. https://doi.org/10.1177/1529100612451018

Lewandowsky, S., Holford, D., & Schmid, P. (2022). Public policy and conspiracies: The case of mandates. *Current Opinion in Psychology* [in press]. https://doi.org/10.1016/j.copsyc.2022.101427

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12*, 361-366. http://doi.org/ 10.1101/lm.94705

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human, Learning, and Memory, 4*, 19-31. https://doi.org/10.1037/0278-7393.4.1.19

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behaviour, 13*, 585-589. https://doi.org/10.1016/S0022-5371(74)80011-3

Love, J. S., Blumenberg, A., & Horowitz, Z. (2020). The parallel pandemic: Medical Misinformation and COVID-19. *Journal of General Internal Medicine, 35*(8), 2435-2436. https://doi.org/10.1007/s11606-020-05897-w

Marsh, E. J., & Yang, B. W. (2018). Believing things that are not true: A cognitive science perspective on misinformation. In B. G. Southwell, E. A. Thorson, & L. Sheble (Eds.), *Misinformation and mass audiences* (pp. 15-34). University of Texas Press.

Meese J, Frith J, Wilken R. COVID-19, 5G conspiracies and infrastructural futures. *Media International Australia*. 2020;177(1):30-46. doi:10.1177/1329878X20952165

Metzger, M. J., Flanagin, A. J., Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication, 60*(3), 413-439. https://doi-org.uml.idm.oclc.org/10.1111/j.1460-2466.2010.01488.x

Meyer, A., Frederick, S., Burnham, T. C., Guevara Pinto, J. D., Boyer, T. W., Ball, L. J., Pennycook., G., Ackerman, R., Thompson, V. A., Schuldt, J. P. (2015). Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology: General*, *144*(2), e16-e30. http://dx.doi.org/10.1037/xge0000049

Mian, A., & Khan, S. (2020). Coronavirus: The spread of misinformation. *BMC Medicine, 18*(89). https://doi.org/10.1186/s12916-020-01556-3

Montgomery, D. (2020, November 1). The last democratic presidential candidate to win Texas, Jimmy Carter, seeks to 'turn Texas blue'. *The New York Times* Retrieved from: https://www.nytimes.com/2020/10/30/us/elections/the-last-democratic-presidential-candidate-to-win-texas-jimmy-carter-seeks-to-turn-texas-blue.html

Morgan, C. A. III., Southwick, S., Steffian, G., Hazlett, G. A., & Loftus, E. F. (2013). Misinformation can influence memory for recently experienced, highly stressful events. *International Journal of Law and Psychiatry, 36*, 11-17. https://doi.org/10.1016/j.ijlp.2012.11.002

Motyka, S., Suri, R., Grewal, D., & Kohli, C. (2016). Disfluent vs. fluent price offers: Paradoxical role of processing fluency. *Journal of Academic Marketing and Science, 44*, 627-638. https://doi.org/10.1007/s11747-015-0459-0

Newman, E. J., Sanson, M., Miller, E. K., Quigley-McBride, A., Foster, J. L., Bernstein, D. M., Garry, M. (2014). People with easier to pronounce names promote truthiness of claims. *PloS One, 9*(2), pp. e88671–e88671, https://doi.org/10.1371/journal.pone.0088671

Oliver, E. J., & Wood, T. J. (2014). Conspiracy theories and the paranoid style(s) of mass opinion. *American Journal of Political Science, 58*(4), 952-966. https://doi.org/10.1111/ajps.12084

Ost, J., Vrij, A., Costall, A., & Bull, R. (2002). Crashing memories and reality monitoring: Distinguishing between perceptions, imaginations and false memories. *Applied Cognitive Psychology, 16*, 125-134. https://doi.org/10.1002/acp.779

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General, 147*(12), 1865-1880. https://doi.org/10.1037/xge0000465

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality, 88*, 185-200. https://doi.org/10.1111/jopy.12476

Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2016). Atheists and agnostics are more reflective than religious believers: Four empirical studies and a meta-analysis. *PloS One*, *11*(4), e0153039–. https://doi.org/10.1371/journal.pone.0153039

Pezdek, K. (2003). Event memory and autobiographical memory for the events of September 11, 2001. *Applied Cognitive Psychology, 17*1033-1045. https://doi.org/10.1002/acp.984

Pillai, R. M., & Fazio, L.K. (2021). The effects of repeating false and misleading information on belief. *WIRES Cognitive Science, 12*, e1573. https://doi.org/10.1002/wcs.1573

Pituch, K. A., & Stevens, J. P. (2015). Applied Multivariate Statistics for the Social Sciences: Analyses with SAS and IBM's SPSS (6th ed.). Routledge. https://doi.org/10.4324/9781315814919

Posetti, J., & Matthews, A. (2018). A short guide to the history of 'fake news' and disinformation: A learning module for journalists and journalism educators. *International Center for Journalists (ICFJ)*, https://www.icfj.org/sites/default/files/2018-07/A%20Short%20Guide%20to%20History%20of%20Fake%20News%20and%20Disinformation_ICFJ%20Final.pdf

Pratto, F., Sidanius, J., Stallworth, L. M., and Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes.

Qualtrics software, Version Jan-Aug, 2022 of Qualtrics. Copyright © 2005 Qualtrics. Qualtrics and all other Qualtrics product or service names are registered trademarks or trademarks of Qualtrics, Provo, UT, USA. https://www.qualtrics.com https://www.qualtrics.com

Read, J. D. (1996). From a passing thought to a false memory in 2 minutes: confusing real and illusory events. *Psychonomic Bulletin & Review, 3*, 105-111. https://doi-org.uml.idm.oclc.org/10.3758/BF03192934

Reber, R. Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science, 9*(1), 45-48. https://doi.org/10.1111/1467-9280.00008

Roccato, M., & Ricolfi, L. (2005). On the correlation between right-wing authoritarianism and social dominance orientation. *Basic and Applied Social Psychology, 27*(3), 187-200. https://doi.org/10.1207/s15324834basp2703_1

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803-814. https://doi.org/10.1037//0278-7393.21.4.803

Sanchez, C. A., & Naylor, J. S. (2018). Disfluent presentations lead to the creation of more false memories. *Public Library of Science, 13*(1), e0191735. https://doi.org/10.1371/journal.pone

Schmidt, S. R. (2004). Autobiographical memories for the September 11th attacks: Reconstructive errors and emotional impairment of memory. *Memory & Cognition, 32,* 443-454. https://doi.org/10.3758/BF03195837

Scholl, S. G., Greifeneder, R., & Bless, H. (2014). When fluency signals truth: Prior successful reliance on fluency moderates the impact of fluency on truth judgments. *Journal of Behavioral Decision Making, 27,* 268-280. https://doi.org/10.1002/bdm.1805

Shimane, D., Matsui, H., & Itoh, Y. (2020). False memory for faces is produced by the Deese-Roediger-McDermott paradigm based on the morphological characteristics of a list. *Current Psychology, 41*(5), 3247-3257. https://doi.org/10.1007/s12144-020-00830-0

Sidi, Y., Ophir, Y., & Ackerman, R. (2016). Generalizing screen inferiority—does the medium, screen versus paper, affect performance even with brief tasks? *Metacognition and Learning, 11*(1), 15-33. https://doi.org/10.1007/s11409-015-9150-6.

Sinclair, A. H., Stanley, M. L., & Seli, P. (2020). Closed-minded cognition: Right-wing authoritarianism is negatively related to belief updating following prediction error. *Psychonomic Bulletin & Review, 27*, 1348 – 1361. https://doi.org/10.3758/s13423-020-01767

Song, H., & Schwarz, N. (2009). If it's difficult to pronounce, it must be risky: Fluency, familiarity, and risk perception. *Psychological Science, 20*(2), 135-138. https://doi.org/10.1111/j.1467-9280.2009.02267.x

Son Hing, L. S., Bobocel, D. R., Zanna, M. P., & McBride, M. V. (2007). Authoritarian dynamics and unethical decision making: high social dominance orientation leaders and high right-wing authoritarianism followers. *Journal of Personality and Social Psychology, 91*(1), 67 – 81. https://doi.org/10.1037/0022-3514.92.1.67

Spinoza, B. (1982). *The Ethics and selected letters* (S. Feldman, Ed., and S. Shirley, Trans.). Indianapolis, IN: Hackett. (Original work published 1677).

Strube, M. J., & Rahimi, A. M. (2006). "Everybody knows it's true": Social dominance orientation and right-wing authoritarianism moderate false consensus for stereotypic beliefs. *Journal of Research in Personality, 40*, 1038-1053. https://doi.org/doi:10.1016/j.jrp.2005.10.004

Swire, B., & Ecker, U. (2018). Misinformation and its correction: Cognitive mechanisms and recommendations for mass communication. In B. G. Southwell, E. A. Thorson, & L. Sheble (Eds.), *Misinformation and mass audiences* (pp. 195-209). University of Texas Press.

Swire, B., Ecker, U. K. H., & Lewandowsky, S. (2017). The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(12), 1948-1961. https://doi.org/10.1037/xlm0000422

Swire-Thompson, B., & Lazer, D. Public health and online misinformation: Challenges and recommendations. *Annual Review of Public Health*, *41*(1), Annual Reviews, 2020, pp. 433–51. https://doi:10.1146/annurev-publhealth-040119-094127

Tessler, H., Choi, M., & Kao, G. (2020). The anxiety of being Asian American: Hate crimes and negative biases during the COVID-19 pandemic. *American Journal of Criminal Justice.* https://doi.org/10.1007/s12103-020-09541-5

Thompson, V., & Morsanyi, K. (2012). Analytic thinking: Do you feel like it? *Mind & Society, 11*, 93-105. https://doi.org/10.1007/s11299-012-0100-6

Trippas, D., Handley, S. J., Verde, M. F., & Morsanyi, K. (2016). Logic brightens my day: Evidence for implicit sensitivity to logical validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(9), 1448-1457. https://doi.org/10.1037/xlm0000248

Tufecki, Z. (2015). Facebook said its algorithms do help form echo chambers, and the tech press missed it. *New Perspectives Quarterly, 32*(3), 9-12. https://doi.org/10.1111/npqu.11519

Visoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*, 1146-1151. https://doi.org/10.1126/science.aap9559

Voss, J. F., Perkins, D. N. & Segal, J. W. (Eds.). (1991). *Informal reasoning and education*. New York, NY: Routledge.

Vraga, E. K., Bode, L., & Tully, M. (2022). Creating news literacy messages to enhance expert corrections of misinformation on Twitter. *Communication Research, 49*(2), 245-267. https://doi.org/10.1177/0093650219898094

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1235-1253. https://doi.org/10.1037/0278-7393.19.6.1235

Whittlesea, B. W. A., & Leboe, J. P. (2000). The heuristic basis of remembering and classification: Fluency, generation, and resemblance. *Journal of Experimental Psychology: General, 129*(1), 84-106. https://doi.org/10.1037/0096-3445.129.1.84

Write, C. L., & Duong, H. (2021). COVID-19 Fake news and attitudes toward Asian Americans. *Journal of Media Research, 14*(39), 5 – 29. https://doi.org/10.24193/jmr.39.1

Yelderman, L. A., & Miller, M. (2017). Religious fundamentalism, religiosity, and priming: effects on attitudes, perceptions, and mock jurors' decisions in an insanity defense case. *Psychology, Crime & Law, 23*(2), 147 -170. https://doi.org/10.1080/1068316X.2016.1239097

Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is—and is not—a desirable difficulty: The influence of typevace clarity on metacognitive judgments and memory. *Memory and Cognition, 41,* 229-241. https://doi.org/10.3758/s1342-012-0255-8

Zakrisson, I. (2005). Construction of a short version of the Right-Wing Authoritarianism scale. *Personality and Individual Differences, 39,* 863-872. https://doi.org/10.1016/j.paid.2005.02.026

Zhao, H. Fu, S., & Chen, X. (2020). Promoting users' intention to share online health articles on social media: The role of confirmation bias. *Information Processing & Management, 57*(6), 102354. https://doi:10.1016/j.ipm.2020.102354

Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., et al. (2010). Individual differences in false memory from misinformation: Personality characteristics and their interactions with cognitive abilities. *Personality and Individual Differences, 48*, 889-894. https://doi.org/10.1016/j.paid.2010.02.016

Appendix A

Letter of Invitation/Informed Consent

Prior to their registration to participate in this research, the following letter of invitation was provided to participants of both Experiment 1 and Experiment 2:

Title of Experiment: An investigation of memory for details of crime reports.

The general purpose of this study is to better understand how people remember details of a crime, and how memory for these details affects their attitudes and sentencing judgments of suspects. This experiment will take approximately 30 minutes to complete, and you will be granted 1 research credit in exchange for your participation. The experiment will involve assuming the role of a judge reviewing the details of two separate crime reports. You will then be asked to make judgments about the perpetrators. In addition, you will be asked to complete a memory test of the various details you remember about each crime.

Exclusion/Inclusion Criteria: All participants must have normal or corrected-to-normal vision, including normal color vision. You may not participate more than once. All participants must be fluent in the English language.

Important Note: This study is not mobile-friendly. You will need to complete the experiment on a computer or laptop.

Principal Investigator:
Erin J. Dowling
umdowlie@myumanitoba.ca

Faculty Supervisor:
Dr. Launa Leboe-McGowan
launa.leboe-mcgowan@umanitoba.ca

Informed Consent Document

Research Project Title: An investigation of memory for details of crime reports

Principal Investigator: Erin Dowling

Supervisor: Dr. Launa Leboe-McGowan

Purpose: The general purpose of this study is to better understand how people remember details of a crime, and how their memory for these details affects their judgments and sentencing of suspects.

Description: The following experiment is an investigation into how people remember details of two separate crimes. You will be asked to assume the role of a judge and read through two different reports that include the details of two separate crimes committed in Winnipeg, Manitoba. Once you have read through each of the crime reports, you will be asked to indicate your attitudes toward the perpetrator, as well as assign a prison sentence to each criminal. In addition, you will be asked to complete a recognition memory test based on the details of the crimes you read. Finally, we are also asking you to complete two personality measures before ending the experiment. This experiment will take approximately 30 minutes to complete. In exchange for your participation, you will receive one (1) research participation credit toward your Introductory to Psychology (PSYC-1200) course.

Risks: There are no risks to participating in this study.

Benefits: This research will provide a better understanding of the basic processes that influence how people remember the details of crimes, and how these remembered details can affect attitudes and judgments towards the people these details involve.

Confidentiality: Confidentiality will be respected. No information that discloses your identity will be released or published. Any identifiers, such as your name, will not be used and will be destroyed a month after the experiments are completed. Other than the data you have contributed during your session, only the number of total participants of each gender along with mean and range of ages of participants will be included in the publication of the results of this study.

Some data and information from this study may be sent outside of the University of Manitoba to other researchers, organizations, or made publicly available. This is for further analysis, testing, as part of the research study, or a requirement by a granting or publication agency. Any information sent out of the University of Manitoba will not show your name or any other identifying personal information about you. However, despite efforts to keep your personal information confidential, absolute confidentiality cannot be guaranteed. Your personal information may be disclosed if required by law.

If you would like to be informed of the main findings of this research via email once the data has been analyzed, you will be given an opportunity to provide your UM email address to the Principal Investigator at the end of the experiment. Only your name and email address will be kept for this purpose and will be kept secure until the main results have been distributed to interested participants, at which point this information will be destroyed. Your name and email

address will never be connected to your responses provided in this experiment. The earliest these results will be distributed is June 2022.

Participation: Your participation in this study is completely voluntary. If you choose to participate, you have the right to discontinue your participation at any time during this experiment, even after consenting to participate. Should you choose not to participate or choose to stop once you have begun, you will still receive your experimental credits. Your email address will only be used to assign research credits, and personal information such as your name will never be connected to your individual responses. Because responses are kept anonymous, you will not be able to withdraw your responses once you have completed the experiment.

Dissemination: The results of this study will provide the data for my Master's thesis research at the University of Manitoba. The findings of this work will form the basis of journal publications, conference presentations, and applications for future research grants and other funding opportunities.

The University of Manitoba Research Ethics Board(s) may also require access to research records for safety and quality assurance purposes.

This research has been approved by the Psychology/Sociology Research Ethics Board at the University of Manitoba. If you have any concerns or complaints about this project you may contact any of the above-named persons or the Human Ethics Secretariat at 474-7122, or e-mail humanethics@umanitoba.ca.

Clicking "I consent to participate" from the options below indicates that you have understood to your satisfaction the information regarding participation in the research project and agree to participate as a subject. In no way does this waive your legal rights nor release the researchers, sponsors, or involved institutions from their legal and professional responsibilities. You are free to withdraw from the study at any time, and/or refrain from answering any questions you prefer to omit, without prejudice or consequence. Your continued participation should be as informed as your initial consent, so you should feel free to ask for clarification or new information throughout your participation.

O Yes, I consent to participate

O No, I do NOT consent (ends the experiment)

Appendix B

ORIGINAL TOM Crime Reports

** *Underlined and italicized details indicate information that was presented as false.*

Tom's Crime – Extenuating False Information

Sentence 1: On the evening of January 16th, 1988, Tom left a neighbor's Austin apartment, where he had been *playing card games* most of the evening.

Sentence 2: He began walking east down a busy main street toward I-35, *stopping to chat with some old friends,* while occasionally turning to face traffic and extending this thumb in the hope of hitching a ride.

Sentence 3: Before reaching the highway, he was picked up by Larry, *a well-known drug dealer,* who was heading north *to join his friends at a strip club.*

Sentence 4: Tom indicated that he wanted to get off downtown.

Sentence 5: Travelling along the highway, Larry recognized that Tom was nervous, preoccupied, and strangely silent.

Sentence 6: To remedy the tension, Larry began recounting a humorous incident that he had recently witnessed.

Sentence 7: Moments into the story, Tom pulled out a knife, *said he was ashamed of what he had to do,* and demanded Larry's wallet, wristwatch, and rings.

Sentence 8: Larry complied.

Sentence 9: With the valuables in his possession, Tom began muttering *that he thought drug dealers were really despicable.*

Sentence 10: Tom then ordered Larry to take an early exit and directed him through a residential area.

Sentence 11: He forced him to stop at a deserted corner.

Sentence 12: As Tom turned to get out of the car, Larry began pleading with him to return his wedding ring.

Sentence 13: Tom continued getting out of the car.

Sentence 14: Before running off, he leaned in through the window toward Larry *and tossed the wedding ring back into Larry's lap.*

Tom's Crime – Exacerbating False Information

Sentence 1: On the evening of January 16th, 1988, Tom left a neighbor's Austin apartment, where he had been *drinking beer and scotch* most of the evening.

Sentence 2: He began walking east down a busy main street toward I-35, *shouting obscenities at pedestrians,* while occasionally turning to face traffic and extending this thumb in the hope of hitching a ride

Sentence 3: Before reaching the highway, he was picked up by Larry, *a disabled veteran,* who was heading north *to join his family at a theater.*

Sentence 4: Tom indicated that he wanted to get off downtown.

Sentence 5: Travelling along the highway, Larry recognized that Tom was nervous, preoccupied, and strangely silent.

Sentence 6: To remedy the tension, Larry began recounting a humorous incident that he had recently witnessed.

Sentence 7: Moments into the story, Tom pulled out a knife, *held it tight against Larry's throat,* and demanded Larry's wallet, wristwatch, and rings.

Sentence 8: Larry complied.

Sentence 9: With the valuables in his possession, Tom began muttering *that he thought crippled people were really disgusting.*

Sentence 10: Tom then ordered Larry to take an early exit and directed him through a residential area.

Sentence 11: He forced him to stop at a deserted corner.

Sentence 12: As Tom turned to get out of the car, Larry began pleading with him to return his wedding ring.

Sentence 13: Tom continued getting out of the car.

Sentence 14: Before running off, he leaned in through the window toward Larry *and threatened to slit Larry's throat if he tried to follow.*

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology, 65*(2), 221-233.

Appendix C

ORIGINAL KEVIN Crime Reports

Kevin's Crime – Extenuating False Information

Sentence 1: On the evening of November 28th, 1989, Kevin left his south Austin home abruptly, after his wife warned she would call the police.

Sentence 2: During an argument that had erupted *over his wife's extravagant spending,* Kevin had yelled at his wife several times *and had threatened to divorce her.*

Sentence 3: Neighbors report that loud disputes between Kevin and his wife were common.

Sentence 4: Leaving his home, Kevin was confronted by his brother-in-law, who had just arrived after being summoned by Kevin's wife.

Sentence 5: When his brother-in-law tried to keep him from leaving, Kevin pushed him aside and rushed to his car.

Sentence 6: He then tore out the parking lot in his red Camaro and travelled to the nearby home of a friend, *to whom he tearfully confessed his long-time marital problems.*

Sentence 7: Hours later he headed home, stopping on the way at a convenience store to pick up cigarettes.

Sentence 8: At the counter, Kevin drew a 9-millimeter gun on the clerk, *told her to stay calm,* and demanded money.

Sentence 9: The clerk silently surrendered 210 dollars in small bills, *while Kevin explained that he needed the cash to cover his wife's debts.*

Sentence 10: On the way out, Kevin overturned a large magazine rack *by accidently running into it.*

Sentence 11: As Kevin pulled away from the front of the store, the clerk inside peered out through the window to get a look at the license plate *and thought she could see Kevin crying in the car*.

Kevin's Crime – Exacerbating False Information

Sentence 1: On the evening of November 28th, 1989, Kevin left his south Austin home abruptly, after his wife warned she would call the police.

Sentence 2: During an argument that had erupted *over his affairs with other women,* Kevin had yelled at his wife several times *and had threatened to stab her.*

Sentence 3: Neighbors report that loud disputes between Kevin and his wife were common.

Sentence 4: Leaving his home, Kevin was confronted by his brother-in-law, who had just arrived after being summoned by Kevin's wife.

Sentence 5: When his brother-in-law tried to keep him from leaving, Kevin pushed him aside and rushed to his car.

Sentence 6: He then tore out the parking lot in his red Camaro and travelled to the nearby home of a friend, *where he bought and used a considerable amount of cocaine.*

Sentence 7: Hours later he headed home, stopping on the way at a convenience store to pick up cigarettes.

Sentence 8: At the counter, Kevin drew a 9-millimeter gun on the clerk, *put it directly to her head,* and demanded money.

Sentence 9: The clerk silently surrendered 210 dollars in small bills, *while Kevin leaned closer and threatened to sexually assault her.*

Sentence 10: On the way out, Kevin overturned a large magazine rack *by violently kicking it out of his way*

Sentence 11: As Kevin pulled away from the front of the store, the clerk inside peered out through the window to get a look at the license plate, *and thought she could see Kevin laughing in the car.*

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology, 65*(2), 221-233.

Appendix D

CURRENT STUDY – TOM Crime Reports

*** Underlined and italicized details indicate information that was presented as false***

Tom's Crime – Extenuating False Information

Sentence 1: On the evening of January 16th, 2019, Tom left a neighbor's Winnipeg apartment, where he had been *playing card games* most of the evening.

Sentence 2: He began walking east down Portage avenue towards Main street, *stopping to chat with some old friends*, while occasionally turning to face traffic and extend his thumb in the hope of hitching a ride.

Sentence 3: Before reaching Main street, Tom was picked up by Larry, *a well-known drug dealer*, who was heading North *to join his friends at a strip club*.

Sentence 4: As they travelled along Main street, Tom indicated that he wanted to get off in the North end.

Sentence 5: Larry recognized that Tom was nervous, preoccupied, and strangely silent.

Sentence 6: In an effort to remedy the tension, Larry began recounting a humorous incident he had recently witnessed.

Sentence 7: Moments into the story, Tom pulled out a knife, *said he was ashamed of what he had to do*, and demanded Larry's wallet, wristwatch and rings.

Sentence 8: Larry did as he was told, and with the valuables in his possession Tom began muttering *that he thought drug dealers were really despicable*.

Sentence 9: Tom then ordered Larry to take an early exit and directed him through a residential area, forcing him to stop at a deserted corner.

Sentence 10: As Tom turned to get out of the car, Larry began pleading with him to return his wedding ring—but Tom continued to exit the vehicle.

Sentence 11: Before running off, Tom leaned in through the window toward Larry and *tossed the wedding ring back into Larry's lap.*

Tom's Crime – Exacerbating False Information

Sentence 1: On the evening of January 16th, 2019, Tom left a neighbor's Winnipeg apartment, where he had been *drinking beer and scotch* most of the evening.

Sentence 2: He began walking east down Portage Avenue towards Main Street, *shouting obscenities at pedestrians*, while occasionally turning to face traffic and extend his thumb in the hope of hitching a ride.

Sentence 3: Before reaching Main Street, Tom was picked up by Larry, *a disabled veteran*, who was heading North *to join his family at the theater*.

Sentence 4: As they travelled along Main Street, Tom indicated that he wanted to get off in the North end. Larry recognized that Tom was nervous, preoccupied, and strangely silent.

Sentence 5: In an effort to remedy the tension, Larry began recounting a humorous incident he had recently witnessed.

Sentence 6: Moments into the story, Tom pulled out a knife, *held it tight against Larry's throat*, and demanded Larry's wallet, wristwatch, and rings.

Sentence 7: Larry did as he was told, and with the valuables in his possession Tom began muttering *that he thought crippled people were really disgusting*.

Sentence 8: Tom then ordered Larry to take an early exit and directed him through a residential area, forcing him to stop at a deserted corner.

Sentence 9: As Tom turned to get out of the car, Larry began pleading with him to return his wedding ring—but Tom continued to exit the vehicle.

Sentence 10: Before running off, Tom leaned in through the window toward Larry and *tossed the wedding ring back into Larry's lap.*

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology, 65*(2), 221-233.

Appendix E

CURRENT STUDY – KEVIN Crime Reports

** *Underlined and italicized details indicate information that was presented as false.*

Kevin's Crime – Extenuating False Information

Sentence 1: On the evening of November 28th, 2018, Kevin left his home in south Winnipeg abruptly, after his wife warned that she would call the police.

Sentence 2: During an argument that had erupted *over his wife's extravagant spending*, Kevin had yelled at his wife several times *and had threatened to divorce her*.

Sentence 3: Neighbors report that loud disputes between Kevin and his wife were common.

Sentence 4: Leaving his home, Kevin was confronted by his brother-in-law James, who had just arrived after being called by Kevin's wife.

Sentence 5: When James tried to keep Kevin from leaving, Kevin pushed him aside and rushed to his car.

Sentence 6: He then tore out of the driveway in his red Camaro and travelled to the nearby home of a friend, *to whom he tearfully confessed his long-time marital problems*.

Sentence 7: Hours later he headed home, stopping at a convenience store on the way to pick up cigarettes.

Sentence 8: At the counter, Kevin drew a 9-millimeter gun on the clerk, *told her to stay calm*, and demanded money.

Sentence 9: The clerk silently surrendered 210 dollars in small bills, *while Kevin explained that he needed the cash to cover his wife's debts*.

Sentence 10: On the way out, Kevin overturned a large magazine rack *by accidentally running into it.*

Sentence 11: As Kevin pulled away from the front of the store, the clerk inside peered out through the window to get a look at the license plate *and thought she could see Kevin crying in the car* as he drove away.

Kevin's Crime – Exacerbating False Information

Sentence 1: On the evening of November 28th, 2018, Kevin left his home in south Winnipeg abruptly, after his wife warned that she would call the police.

Sentence 2: During an argument that had erupted *over his affairs with other women*, Kevin had yelled at his wife several times *and had threatened to stab her*.

Sentence 3: Neighbors report that loud disputes between Kevin and his wife were common.

Sentence 4: Leaving his home, Kevin was confronted by his brother-in-law James, who had just arrived after being called by Kevin's wife.

Sentence 5: When James tried to keep Kevin from leaving, Kevin pushed him aside and rushed to his car.

Sentence 6: He then tore out of the driveway in his red Camaro and travelled to the nearby home of a friend, *where he bought and used a considerable amount of cocaine.*

Sentence 7: Hours later he headed home, stopping at a convenience store on the way to pick up cigarettes.

Sentence 8: At the counter, Kevin drew a 9-millimeter gun on the clerk, *put it directly to her head*, and demanded money.

Sentence 9: The clerk silently surrendered 210 dollars in small bills, *while Kevin leaned closer and threatened to sexually assault her*.

Sentence 10: On the way out, Kevin overturned a large magazine rack *by violently kicking it out of the way.*

Sentence 11: As Kevin pulled away from the front of the store, the clerk inside peered out through the window to get a look at the license plate *and thought she could see Kevin laughing in the car* as he drove away.

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology, 65*(2), 221-233.

Appendix F

Recognition Memory Measure

Tom's Crime Report

1. Tom had been watching football on T.V. at a neighbour's place.
2. Tom had been drinking beer and scotch at a neighbor's place.
3. Tom had been playing card games at a neighbor's place.
4. Tom's neighbor was named Dave.
5. Tom had been watching baseball on T.V. at a neighbor's place.
6. Tom's neighbor was named John.
7. Tom stopped to chat with some old friends.
8. Tom shouted obscenities at pedestrians.
9. Larry was a disabled veteran.
10. Larry was a well-known drug dealer.
11. Larry was driving a Buick.
12. Larry was driving an Oldsmobile.
13. Larry was planning to join his family at the theatre.
14. Larry was smoking a cigar.
15. Larry was planning to join some friends at a strip club.
16. Tom told Larry he wanted to get off in the north end.
17. Larry was smoking a cigarette.
18. Tom told Larry he wanted to get off at a restaurant.
19. Larry began telling Tom a humorous story.
20. Larry began telling Tom a sad story.
21. Tom held a knife against Larry's throat.
22. Tom said he was ashamed of what he had to do.
23. Larry refused to hand over his wallet, wristwatch, and rings.
24. Larry handed over his wallet, wristwatch and rings to Tom.
25. Tom said that he thought that drug dealers were despicable.
26. Tom said that he thought crippled people were disgusting.
27. Larry asked Tom to please give back the wedding ring.
28. Tom gave Larry back his wedding ring.

29. Larry asked Tom to please give back the wristwatch.
30. Tom threatened to slit Larry's throat if he followed.

Kevin's Crime Report

1. Kevin argued with his wife over her extravagant spending.
2. Kevin's wife was named Julie.
3. Kevin threatened to stab his wife.
4. Kevin's wife was named Mary.
5. Kevin threatened to divorce his wife.
6. Kevin argued with his wife over his affairs with other women.
7. Neighbors report that the couple often threw loud parties.
8. Neighbors report that the couple often had loud arguments.
9. Leaving his home, Kevin pushed his brother-in-law.
10. Leaving his home, Kevin pushed a neighbor.
11. Kevin drove off in a red Camaro.
12. Kevin drove off in a red Cobra.
13. Kevin's friend was also Kevin's wife's friend.
14. Kevin told his friend about his marital problems.
15. Kevin's friend had never met Kevin's wife.
16. Kevin used cocaine at his friend's place.
17. Kevin used a 9-millimeter gun in the robbery.
18. Kevin used a shotgun in the robbery.
19. The clerk was a high school student.
20. The clerk was a university student.
21. Kevin told the clerk to stay calm.
22. Kevin put his gun to the clerk's head.
23. Kevin told the clerk he needed money to cover his wife's debts.
24. Kevin threatened to sexually assault the clerk.
25. Kevin kicked a magazine rack on the way out.
26. Kevin accidentally ran into a magazine rack on the way out.
27. The clerk thought she saw Kevin crying as he drove away.

28. The clerk thought she saw Kevin laughing as he drove away.
29. Kevin drove straight home after the robbery.
30. Kevin drove back to his friend's place after the robbery.

Reference:

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology, 65*(2), 221-233.

Appendix G

Right-Wing Authoritarianism (RWA) Measure

Please indicate your level of agreement with each of the following statements using the scale below:

Strongly Disagree			*Neutral*			*Strongly Agree*
1	***2***	***3***	***4***	***5***	***6***	***7***

1. Our country needs a powerful leader, in order to destroy the radical and immoral currents prevailing in society today.
2. *Our country needs free thinkers, who will have the courage to stand up against traditional ways, even if this upsets many people.*
3. The "old-fashioned ways" and "old-fashioned values" still show the best way to live.
4. *Our society would be better off if we showed tolerance and understanding for untraditional values and opinions.*
5. God's laws about abortion, pornography and marriage must be strictly followed before it is too late, violations must be punished.
6. *Our society needs to show openness towards people thinking differently, rather than a strong leader – the world is not particularly evil or dangerous.*
7. It would be best if newspapers were censored so that people would not be able to get hold of destructive and disgusting material.
8. *Many good people challenge the state, criticize the church and ignore "the normal way of living".*
9. Our forefathers ought to be honored more for the way they have built our society, at the same time we ought to put an end to those forces destroying it.
10. *People ought to put less attention to the Bible and religion, instead they ought to develop their own moral standards.*
11. There are many radical, immoral people trying to ruin things; the society ought to stop them.
12. *It is better to accept bad literature than to censor it.*
13. Facts show that we have to be harder against crime and sexual immorality, in order to uphold law and order.
14. *The situation in society of today would be improved if troublemakers were treated with reason and humanity.*
15. If society so wants, it is the duty of every true citizen to help eliminate the evil that poisons our country from within.

Note: Items in italics were reverse coded.

Reference:

Zakrission, I. (2005). Construction of a short version of the Right-Wing Authoritarianism (RWA) scale. *Personality and Individual Differences, 39*, 863 – 872.

Appendix H

Social Inequity (SI) Scale - SDO Measure

Please indicate your level of agreement with each of the following statements using the scale

Strongly Disagree			*Neutral*			*Strongly Agree*
1	***2***	***3***	***4***	***5***	***6***	***7***

below:

1. Economic equality is a bad idea dreamed up by muddle-headed "do-gooders". It would be a big mistake to pursue it.

2. *Although it may be ultimately impossible, economic equality is a very worthy goal that we should strive for.*

3. People have **no** right to economic equality. All of us should get as much as we can, and if some don't get enough, that's their problem.

4. If the natural forces of supply and demand and power make a few people immensely wealthy and millions of others poor, so be it.

5. *Everyone should have an equal opportunity for economic success. Those born into poor circumstances should be given extra help to make the "playing field" level for them.*

6. *"Access programs" to higher education, which give people from poor backgrounds extra financial support and counseling while in university, are a good idea.*

7. Nobody should get extra help improving his place in society. Everyone should start off with what his family gives him and go from there.

8. *Tax money should be used to make sure everyone has an adequate standard of living.*

9. There is nothing wrong with the fact that powerful people get better treatment by the law than poor people do.

10. *Society should provide poor people with a good attorney in court if they cannot afford one.*

11. If powerful people can get away with illegal acts because they can afford the best lawyers, and because they have "friends in high places", so what? It's just natural.

12. *Since so many members of minority groups end up in our jails, we should take strong steps to make sure prejudice plays no role in their treatment by the legal system.*

13. *It is very important that women and members of visible minorities become police officers and move up in the ranks.*

14. The "one-person-one vote" idea is dumb. People who make bigger contributions to our society should get a lot more votes than those who do nothing.

15. "Equality" is one of those nice-sounding names for suckers. Actually, only fools believe in it.

16. *Equality is one of the fundamental principles of democracy, so we should work hard to increase it.*

17. Some groups just don't care about improving their lot, and don't deserve to be equal with others.

18. It is a huge mistake to treat everyone as equals. That will just encourage the shiftless, lazy people to coast along reaping the benefits of other people's hard work.

19. *The biggest challenge before us is to create greater social justice for the disadvantaged in our country.*

20. Every human life is worth as much as every other human life.

21. If everyone really were treated equally, I would get less and I would not like that.

22. *No racial group is naturally inferior to any other. If a group does poorly, it is usually because of discrimination.*

Note: The anti-equality response for items in italics is to disagree. Italicized items will be reverse coded. Items 1-8 tap economic inequality, items 9-14 tap legal inequality, and items 15-22 tap general ideas about equality.

Reference:

Altemeyer, B. (2004). Highly dominating, highly authoritarian personalities. *The Journal of Social Psychology, 144*(4), 421-447.