Confirmation Bias for Degraded Forensic Audio Evidence

by

Megan E. Giroux

M.A., Simon Fraser University, 2016

B.S. (Honors), B.A., Ursinus College, 2014

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Psychology
Faculty of Arts and Social Sciences

Declaration of Committee

Name:	**Megan Giroux**
Degree:	**Doctor of Philosophy (Psychology)**
Title:	**Confirmation Bias for Degraded Forensic Audio Evidence**
Committee:	**Chair: Timothy Racine** Professor, Psychology
	Deborah Connolly Supervisor Professor, Psychology
	Daniel Bernstein Committee Member Adjunct Professor, Psychology
	Ronald Roesch Committee Member Professor, Psychology
	David MacAlister Examiner Associate Professor, Criminology
	Norbert Schwarz External Examiner Professor, Psychology University of Southern California

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Abstract

As recording devices such as cell phones and body cams become more accessible, audio recordings are increasingly being used as evidence in legal cases. The goal of this research was to test the (1) applied circumstances under which people misinterpret audio recordings; (2) mechanisms that contribute to confirmation bias in these cases; and (3) methods for mitigating confirmation bias for degraded audio recordings. In Study 1, I found that incriminating contextual information biased participants' interpretations of innocuous degraded audio recordings. Specifically, participants who learned that they were listening to wiretapped conversations with criminal suspects or that they were listening to wiretapped conversations with criminal suspects in cases also involving eyewitness evidence made more incriminating misinterpretations than participants who learned no context about the recordings. In Study 2, I tested whether fluency misattribution is a mechanism for confirmation bias by including reaction time as an independent measure of fluency. I manipulated perceptual fluency within-subject by presenting non-degraded recordings, minimally degraded recordings, and moderately degraded recordings. I manipulated conceptual fluency between-subject by varying the amount of context participants received about the recordings: (1) no context; (2) the recordings came from criminal suspects' conversations (criminal suspect); (3) incriminating written transcripts. Across degradation levels, participants in the incriminating written transcripts condition made more incriminating misinterpretations than participants in the criminal suspect condition, who in turn made more incriminating misinterpretations than participants in the no context condition. Additionally, reaction time partially mediated the effect of context on incriminating misinterpretations for minimally and moderately degraded recordings, suggesting that fluency misattribution contributes to confirmation bias. In Study 3, I tested whether the Canadian Model Jury Instructions for audio recordings and written transcripts effectively reduce participants' tendency to make incriminating misinterpretations after reading incriminating transcripts. When there was no delay between participants' exposure to incriminating transcripts and when they interpreted the recordings, participants who received instructions made fewer incriminating misinterpretations than participants who did not receive instructions. However, when there was a one-week delay, there were no differences between those who received instructions and those who did not. These studies have methodological, theoretical, and applied implications.

Keywords: Confirmation bias; law; fluency; cognitive bias; context

Acknowledgements

This work was supported by a Vanier Canadian Graduate Scholarship from the Social Sciences and Humanities Research Council of Canada. This work was also supported by an Insight Development Grant awarded to Drs. Daniel Bernstein, Deborah Connolly, and Itiel Dror from the Social Sciences and Humanities Research Council of Canada.

This work would have been impossible without the dedication, guidance, and support of my supervisors, Dan Bernstein and Deb Connolly. Dan expects a lot from his students, but with every high expectation comes equal levels of genuine patience, encouragement, and generosity. Dan is a tremendous scholar and has been absolutely pivotal in my development as a researcher. He has pushed me to think independently and has shaped me into a better methodologist, writer, and collaborator. I am lucky to have benefitted from Dan's extreme dedication to his students. Deb is an incredible mentor. I have always appreciated the depth of expertise that she brings to every project. She has similarly pushed me to think more critically about the methodological, theoretical, and applied implications of my work. Deb is also hilarious and caring, and I have equally enjoyed the personal and professional conversations we have shared over the years. I was uniquely lucky to be part of two powerhouse labs and am extremely grateful for all the experiences I was afforded, connections I built, and skills that I learned. I would also like to thank Ron Roesch for his support and insightful feedback on my dissertation.

Thank you to my fabulous research assistants Angela Pelletier, Kiran Dogra, Elisabeth Avery, Karleigh Williams, Tanya Tyrsa, and Zhina Fullad for their conscientious work coding participants' responses and managing data collection. Thank you to my lab mates, Daniel Derksen, Patricia Coburn, Camille Weinsheimer, Madison Harvey, Shelbie Anderson, Dayna Woiwod, Katie Berens, and Nikola Edge. It is important to have friends to share the successes and struggles with, and you all have been a major source of support throughout grad school. I appreciate all the opportunities we have had to work together and know that I have made lifelong friends.

I would like to thank my family, and especially my parents, for being my biggest fans throughout this journey. Mom, you have been an incredible role model to me. Your work ethic, dedication to your goals, and devotion to our family inspire me. Thank you for all the phone calls, academic insights, and reminders to find a healthy balance between work and fun. Dad, you are simply the most loving father and your pride in me has been a constant source of reassurance at times when I felt like things were too challenging. You’ve shown me the value of hard work, tenacity, and a bit of humor. I am so fortunate that you two have always encouraged me to follow my dreams. Thank you for all your love and support.

Finally, thank you to my husband. Randy, throughout this entire process, you have been my grounding force. I can’t ever fully express my gratitude for your love, support, and endless patience. Thank you for the all the dinners you made, the countless practice presentations you watched, the late nights you stayed up with me while I worked, and for calming me down when I was stressed. Thank you for always encouraging me to celebrate the milestones—whether they were big or small—and for your valiant attempts to explain my research. You sacrificed a lot for me to achieve this dream and I am so lucky that you have been my cheerleader throughout this entire journey. I love you. Thank you.

Table of Contents

List of Tables

List of Figures

Chapter 1.

Introduction

Humans constantly make decisions. From the seemingly insignificant decisions such as choosing what to eat for breakfast, to the life-changing decisions such as finding a new job, we face several decisions daily. Fortunately, our brains have developed several cognitive heuristics, or “short-cuts,” to decision-making (Tversky & Kahneman, 1974; Shah & Oppenheimer, 2008). These heuristics help us filter through the immense amount of information we encounter daily to efficiently make decisions and make decisions in the face of uncertainty. Although mostly helpful, these heuristics may lead to cognitive biases, or systematic deviations from rational information processing and decision-making (Haselton et al., 2015). While there are several cognitive biases that may lead to flawed decision-making, one bias appears to be particularly pervasive. In fact, Nickerson (1998) said, “If one were to attempt to identify a single problematic aspect of human reasoning that deserves attention above all others, the *confirmation bias* would have to be among the candidates for consideration” (p. 175).

1.1. Confirmation Bias

Confirmation bias is a cognitive heuristic that occurs when individuals seek, interpret, recall, or disregard evidence in ways that are consistent with their pre-existing beliefs (Nickerson, 1998). For example, when individuals are led to believe that a child and adult are genetically related before viewing photographs of child-adult pairs, they tend to give higher facial resemblance ratings to the pairs than those who do not learn such information (Bressan & Dal Martello, 2002). Thus, individuals’ pre-existing beliefs can influence their evaluations of subsequently presented information, such that they assess this information in a manner consistent with those pre-existing beliefs. While confirmation bias has only been formally recognized in the psychological literature over the past six decades, humans have long understood the tendency to make decisions consistent with one’s current beliefs or expectations. Consider, for example, Julius

Caesar, who said “Men generally believe quite freely that which they want to be true” (Risinger et al., 2002, p. 6) or Francis Bacon (1620/1939), who said “The human understanding when it has once adopted an opinion…draws all things else to support and agree with it” (p. 36). Thus, people have recognized the pervasive consequences of confirmation bias for centuries.

A classic experiment on hypothesis-testing demonstrated the propensity for individuals to seek information which would confirm, rather than contradict, their pre-existing beliefs. Wason (1960) gave participants a series of three numbers, or triplets, and asked participants to figure out the rule that was used to determine the set of numbers. Participants could either ask about additional triplets that would follow the rule or triplets that would defy the rule. Wason found that participants typically only asked about triplets that were consistent with their hypothetical rule. Thus, he believed that this was evidence that individuals exhibit a confirmation bias. Though Wason did not specifically use the term “confirmation bias” to explain this effect in his original empirical article on this phenomenon, his 1960 work is revered as the seminal study on confirmation bias in the literature (Gale & Ball, 2002).

Other research found that this effect extends to social interactions, demonstrating that individuals often seek hypothesis-consistent information about other people. For example, Snyder & Campbell (1980) asked participants to select questions that would test whether the person they would be interacting with was a prototypical introvert or extravert. When participants were asked to test the introvert hypothesis, they asked more introvert-consistent questions (e.g., “do you like spending time alone?”). Alternatively, when participants were asked to test the extravert hypothesis, they asked more extravert-consistent questions (e.g., “do you enjoy socializing?”). Thus, participants appeared to use a so-called “confirmatory hypothesis-testing” strategy, selecting more questions that would confirm their hypothesis rather than contradict it. These classic experiments prompted years of research on the tendency to seek information consistent with one’s pre-existing hypothesis (Bruner & Potter, 1964; Mynatt et al., 1977; Rosenthal & Jacobson, 1968; Skov & Sherman, 1986; Snyder & Swann, 1978; Wason, 1968).

As research on confirmation bias continued to evolve, the construct extended beyond this initial demonstration of people's tendency to engage in a *positive hypothesis testing strategy* (i.e., the tendency to ask more questions that would confirm, rather than contradict, one's hypothesis; Klayman & Ha, 1987). For example, Jonas et al. (2001) showed that when participants form a preliminary belief on a particular issue (e.g., health policy), they *selectively seek* information which would confirm their belief. In a series of studies, the experimenters presented participants with articles which would either support or contradict their preliminary beliefs on health policy. Jonas et al. found that when they presented supporting information simultaneously with contradictory information, participants tended to select the supporting information. They also found that when they presented supporting and contradictory information sequentially, participants selected supporting articles to an even greater extent than when the options appeared simultaneously. They suggest that sequential selection of information may have more implications for decision-making in the real world, where individuals will most often search for information in a sequential manner. They also provide evidence that increases in confirmation bias resulting from sequential presentation of information occur because individuals become increasingly committed to their initial beliefs over time (Jonas et al., 2001; see also Knobloch-Westerwick et al., 2020 for evidence that political views lead to selective exposure to political information).

While confirmation bias occurs when individuals selectively seek information that is consistent with their pre-existing beliefs, it also occurs when individuals *interpret* information they subsequently encounter as being consistent with their beliefs. For instance, people who hold pre-existing expectations about another person may interpret that person's actions in a manner that is consistent with these expectations (Duncan, 1976; Langer & Abelson, 1974). Darley and Gross (1983) found that individuals interpret a child's actions differently based on their pre-existing beliefs about the child's socioeconomic status (SES). Specifically, the researchers told half the participants that they would be watching a video of a child from either a low- or high-SES family. Then, all participants watched a video of the child taking an academic ability test. Those participants who believed that the child came from a low-SES family rated the child's abilities as below-grade level while participants who believed that the child came from a high-SES family rated the child's abilities as above-grade level, despite both groups watching the same video. Thus, individuals interpreted the same actions in opposite ways depending on what they were led to believe about the child's abilities. Several other studies have demonstrated this same tendency for individuals to distort information once they've formed a decision (see e.g., Cvencek et al., 2011 for evidence that the stereotype that boys are better than girls at math influences young girls' self-concepts about their own mathematic abilities; see also Elliot & Devine, 1994; Festinger, 1957; Hodgins & Zuckerman, 1993). In fact, Russo et al. (1996) found that even a developing preference for a particular hypothesis (as opposed to a strong preference) can cause individuals to evaluate new information in a distorted manner.

Confirmation bias also occurs when individuals fail to seek or utilize information which would support an alternative hypothesis, or even *disregard* information that is inconsistent with their pre-existing beliefs (Mynatt et al., 1977; Skov & Sherman, 1986). Some research shows that when individuals receive evidence that contradicts their beliefs, they may evaluate that evidence as flawed and unreliable (Lord et al., 1979). For example, Lord et al. provided proponents and opponents of capital punishment with two studies on the deterrent effects of capital punishment—one which supported their initial beliefs about deterrence, and one which contradicted these beliefs. Participants who were proponents of capital punishment found the study which supported the deterrent efficacy of capital punishment to be more convincing and of a higher quality than the study that was anti-deterrent. Alternatively, participants who were opponents of capital punishment found the anti-deterrent study to be higher in quality and convincingness than the pro-deterrent study. Furthermore, participants' decisions to accept the study's findings or to search for flawed procedures and consider alternative explanations depended heavily on whether the study was consistent with their initial beliefs (Lord et al., 1979). Thus, individuals appear to consider the relevance of information differently depending on whether it supports or contradicts their beliefs (Darley & Gross, 1983).

In addition to the individual tendency to selectively seek hypothesis-consistent information and erroneously evaluate post-decision information, there is evidence that confirmation bias can also occur in group decision-making. In a series of experiments, Schulz-Hardt et al. (2000) compared confirmation bias exhibited by individuals to that exhibited by groups of people with either homogenous or heterogenous pre-existing beliefs about a financial investment case. In their first experiment, there were five members of each group and heterogenous groups consisted of either one or two people with opposing pre-existing beliefs. They found that individuals, groups with members who all held homogenous initial beliefs, and heterogenous groups with a one-person minority sought and evaluated information consistent with the dominating pre-existing beliefs. Thus, individuals, as well as groups with a majority of homogenous beliefs, exhibited confirmation bias. However, heterogenous groups with a two-person minority did not exhibit confirmation bias. In a follow-up study, the authors also found that confirmation bias in group decision-making reflected group-level processes rather than the aggregate of individual processes. That is, homogenous groups (but not heterogenous groups) collectively decided to search for belief-consistent information rather than each individual group member independently selecting consistent information. This finding has implications for several fields where important decisions are often made by groups of people, including business, politics, and the law.

Confirmation bias can also influence how individuals recall information that they previously encountered (Kassin et al., 2013). For example, imagine that a police investigator encounters ambiguous evidence before forming any beliefs about the suspect in a criminal case. Later, that investigator learns from an informant that a particular person has confessed to committing the crime. This person subsequently becomes the suspect in the case. The police officer may now recall the once-ambiguous evidence as evidence which clearly demonstrates that the suspect identified by the informant committed the crime. While confirmation bias can and does affect how people recall previously encountered information (also see work on *hindsight bias* for discussion of similar processes; Roese & Vohs, 2012), I primarily focus on cases in which pre-existing beliefs influence how individuals evaluate subsequently encountered information.

The term “bias” may raise a particular connotation of conscious or even prejudicial intent. However, it is important to note that confirmation bias is typically not a deliberate process. Rather, confirmation bias is often an automatic process that occurs when individuals unwittingly seek belief-consistent information, interpret evidence in a way that supports their beliefs, and disregard evidence that contradicts their beliefs (Nickerson, 1998). Thus, beyond conscious awareness, there are certain mechanisms that may cause individuals to exhibit confirmation bias in their decision-making processes.

1.2. Mechanisms for Confirmation Bias

In general, there has been a lack of empirical research formally testing the mechanisms that contribute to confirmation bias (Lidén et al., 2018). However, researchers have proposed several cognitive, motivational, and metacognitive mechanisms that may underlie confirmation bias (see Table 1). These mechanisms are not mutually exclusive; confirmation bias likely stems from an interplay between multiple mechanisms.

In the information gathering stage, confirmation bias may occur when individuals engage in a *selective information search.* In other words, once people have formed a belief, they may only seek out information that is consistent with this belief (Darley & Gross, 1983; Nickerson, 1998; Schulz-Hardt et al., 2000). This theory is limited to the information search process and does not account for cases in which individuals encounter potentially inconsistent evidence. Thus, there are other mechanisms that affect how individuals evaluate evidence they encounter after forming their initial beliefs.

First, it could be that confirmation bias arises from a *primacy effect*, when the information upon which individuals form their initial beliefs affects how they evaluate information they subsequently encounter (Anderson & Jacobson, 1965; Bruner & Potter, 1964; Lingle & Ostrom, 1981; Sherman et al., 1983). Asch (1946) was the first to demonstrate the primacy effect. He found that individuals formed a more positive impression about a person when they learned positive characteristics about the person first, and a more negative impression about a person when they learned negative characteristics first. Thus, the information that participants first encountered influenced their conclusions more strongly than information they later encountered.

A closely related mechanism called *belief persistence* or *belief perseverance* may also underlie confirmation bias (Hayden & Mischel, 1976; Ross et al., 1975). As the phrase suggests, belief persistence occurs when individuals deliberately or unconsciously resist changing their initial beliefs. This may influence how people evaluate information they subsequently acquire, causing them to interpret ambiguous information as consistent with their pre-existing beliefs. This may also lead people to give little or no weight to information that is inconsistent with their pre-existing beliefs. Confirmation bias may also arise from a *positive test strategy*, in which people tend to test hypotheses that confirm, rather than contradict, pre-existing beliefs, expectations, or desires (Klayman & Ha, 1987; Wason, 1960).

Other research suggests that *selective attention* underlies confirmation bias. For example, Talluri et al. (2018) found that after viewing an array of randomly moving dots, participants selectively attended to features of a second array of randomly moving dots that were consistent with their previous decision. In this study, researchers presented participants with a perceptual task in which they viewed an array of randomly moving dots and reported whether the dots were moving clockwise or counterclockwise with respect to a reference line. Then, participants viewed a second array of randomly moving dots and estimated the average direction of both the first and second display of dots by moving a dial as far to the right or left of a 0° reference line as they wished. The researchers found that when participants estimated the average direction of both stimuli, their average was consistently biased in the direction of their first decision. That is, if they indicated that the dots were moving clockwise in the first array of dots, they were more likely to move the dial to the right of the 0° reference line in the second array of dots. Alternatively, if they indicated that the dots were moving counterclockwise in the first array of dots, they were more likely to move the dial to the left of the reference line in the second array of dots. The authors suggested that this occurred because, when viewing the second stimulus, participants selectively attended to the dots moving in the direction they identified in the first stimulus rather than the dots moving in the other direction (Prat-Ortega & de la Rocha, 2018; Talluri et al., 2018).

Neuroscientists are also beginning to study the neural correlates of confirmation bias. In one study, participants in an fMRI scanner received information that either confirmed or contradicted their previous decisions about the asking price of several properties on an international real-estate site. The researchers found that when individuals received disconfirming evidence, there was reduced neural activity in the posterior medial prefrontal cortex. This brain region is believed to play a role in error and performance monitoring upon receiving post-decision information (Kappes et al., 2020). Furthermore, activity in the posterior medial prefrontal cortex mediated participants' final decisions when they received confirmatory information, but not when they received contradictory information.

There is some evidence that confirmation bias might arise from motivational bases as well (Nickerson, 1998). First, it could be that once people have made a decision, they circumvent or even distort contradictory information to avoid *cognitive dissonance* (Elliot & Devine, 1994; Russo et al., 1996). Cognitive dissonance occurs when conflicting attitudes or beliefs produce a feeling of discomfort (Festinger, 1957). In the seminal research on cognitive dissonance, Festinger suggested that individuals have a need for internal consistency. When this consistency is challenged, people are motivated to reduce the resulting inconsistency. It is possible that to reduce this inconsistency, people selectively seek information which would confirm rather than contradict their current beliefs. Additionally, people may distort information which contradicts their current beliefs to reduce cognitive dissonance. This could lead to both a biased search for belief-consistent information, as well as a biased interpretation of information they subsequently encounter.

People may also be motivated by a *need for cognitive closure,* or the need to reach definitive conclusions about certain pieces of information or hypotheses (Ask & Granhag, 2005; Kruglanski & Webster, 1996). This particular mechanism may apply more specifically to circumstances in which there is high pressure to draw a conclusion, such as in police investigations where detectives typically work under urgent time pressures to find and charge criminals. When people are motivated by a dispositional or situational need for closure, they tend to focus their attention on a tentative solution and fail to consider alternative solutions. Consequently, they may selectively attend to information which supports this tentative solution, show heightened primacy effects, or disregard information which contradicts their solution (Ask & Granhag, 2005; Kruglanski et al., 1993; Webster et al., 1996). Other research has found that participants are more likely to distort or disregard correlations between two variables (e.g., individualism and relationship success) when that correlation threatens their self-concept, suggesting that *self-esteem needs* may underlie confirmation bias as well (Munro & Stansbury, 2009).

Confirmation bias might also occur because individuals process information that is consistent with their pre-existing beliefs more easily (Darley & Gross, 1983; Zadny & Gerard, 1974). This explanation is consistent with the *fluency misattribution* theory. Fluency refers to the ease with which people process information (Jacoby & Dallas, 1981; Whittlesea, 1993). It is a metacognitive experience that accompanies our reasoning and can influence our decision-making in many ways (Schwarz, 2004). For example, people are more likely to believe a statement is true when it is presented in an easy-to-read font (e.g., in a high contrast blue font) compared to a difficult-to-read font (e.g., in a low contrast yellow font; Reber & Schwarz, 1999). Research has found that processing fluency affects a range of judgments including truth, confidence, liking, and pleasure (Alter & Oppenheimer, 2009).

When people form an initial hypothesis, it may be easier for them to generate, evaluate, and recall information consistent with that hypothesis than information that contradicts their hypothesis. That is, they experience fluency when processing hypothesis-consistent information. Consequently, they may make a fluency misattribution, in which they mistakenly assume that the ease with which they processed the hypothesis-consistent information means that the information is correct (Alter & Oppenheimer, 2009; Jacoby & Dallas, 1981). Little research has empirically tested whether fluency is a mechanism for confirmation bias, but there is some initial evidence that fluency plays a role. For example, one study found that fluency mediates confirmation bias in decisions about financial experts' authority (Zaleskiewicz & Gasiorowska, 2021). Specifically, when participants received recommendations from a financial advisor that were consistent with their own opinions, they rated the advice as easier to process and understand (i.e., higher fluency ratings). These fluency ratings subsequently mediated participants' perceptions of the financial experts' authority.

Furthermore, some research has shown that *disfluency* mitigates confirmation bias (Hernandez & Preston, 2013; O'Brien, 2009; O'Brien & Ellsworth, 2006). For example, Hernandez and Preston (2013) demonstrated that presenting ambiguous information in a disfluent format reduces confirmation bias after participants have formed initial beliefs. The researchers experimentally manipulated participants' pre-existing beliefs in an accused's guilt by providing "witness testimony" about the accused's past behavior. Then, they gave participants the facts of the case in which the accused's guilt was ambiguous. Half the participants read the case in a fluent (i.e., easy to read) font, while the remaining participants read the case in a disfluent (i.e., hard to read, but still legible) font. Participants only exhibited confirmation bias when they read the case in a fluent font. That is, participants with pre-existing beliefs in the accused's guilt were more likely to find the accused guilty after reading the case in a fluent font than participants who read the case in a disfluent font. This is consistent with the notion that presenting information in a disfluent font caused participants to evaluate the case more analytically and critically (Alter et al., 2007).

Table 1. Mechanisms for Confirmation Bias

Cognitive	
Selective Information Search	Individuals only seek information which is consistent with their pre-existing beliefs
Primacy Effect	Information upon which individuals form their initial beliefs affects how they evaluate subsequent information
Belief Persistence	Individuals deliberately or unconsciously resist changing the initial belief they formed
Positive Test Strategy	People tend to test hypotheses that confirm, rather than contradict, pre-existing beliefs
Selective Attention	People selectively attend to information or perceptual features of stimuli that support their pre-existing beliefs
Neural Inputs	There is reduced neural activity in the posterior medial prefrontal cortex in response to disconfirming evidence

	Motivational
Motivation to Avoid Cognitive Dissonance	People selectively seek information consistent with their pre-existing beliefs or distort information inconsistent with their beliefs to reduce feelings of internal inconsistency arising from conflicting beliefs/attitudes
Need for Closure	Individuals with a strong desire for predictability in their world are more likely to select information that supports their initial decision because they are motivated to reach a definitive conclusion
Self-Esteem Needs	Individuals are more likely to distort or disregard evidence that threatens their self-concept
	Metacognitive
Fluency Misattribution	People misattribute the ease with which they process information that is consistent with their pre-existing beliefs to the information being correct

1.3. Confirmation Bias and Law

Despite confirmation bias's clear implications for decision-making across several applied domains, researchers have only recently begun to empirically examine the effects of confirmation bias in the law. Yet, it is clear from this growing literature that confirmation bias can contribute to erroneous decision-making across several stages of the legal process. First, confirmation bias can impact the investigative process, influencing not only the evidence that investigators seek, but the way they interpret the evidence they encounter. For instance, Charman et al. (2017) found that police officers' initial beliefs in a suspect's guilt influenced their later evaluations of ambiguous evidence. Specifically, police officers who initially received inculpatory evidence, or evidence which tends to incriminate the accused, subsequently rated four pieces of ambiguous evidence (alibi evidence, facial composite evidence, handwriting evidence, and informant evidence) as being more incriminating than officers who initially received either exculpatory evidence (evidence which tends to exonerate the accused) or neutral evidence. This suggests that officers who initially made higher likelihood of guilt ratings systematically evaluated the ambiguous evidence as evidence of guilt. Furthermore, those police officers who interpreted the ambiguous evidence as more incriminating given their pre-existing beliefs in the suspect's guilt provided even higher likelihood of guilt ratings after evaluating the ambiguous evidence. In other words, their final likelihood of guilt ratings were bolstered by their biased evaluations of the ambiguous evidence. This type of bias may also influence interrogations and eyewitness line-up identifications (Hasel & Kassin, 2009; Kassin et al., 2003).

Confirmation bias also appears to be particularly prevalent within the forensic science field. In fact, in a report on the current issues within the forensic science field, the United States (U.S.) National Academy of Sciences noted that "biasability" was among the forefront (National Academy of Sciences, 2009). This is because forensic scientists are often exposed to several sources of contextual information. For example, they often know about other lines of evidence, who the target suspect in a crime is, or whether a suspect confessed to the crime (Dror, 2016; Dror & Cole, 2010; Dror et al., 2017). Thus, despite the fact that forensic scientists are expected to provide reliable conclusions about whether forensic evidence provides proof of a particular suspect's involvement in a crime, contextual information about a case may bias their evaluations and conclusions (Dror & Cole, 2010).

Dror et al. (2006) found evidence that contextual information influences forensic experts' evaluations. In their study, the researchers presented five fingerprint examiners with a set of prints they had previously examined and deemed to be a match. However, the examiners were unaware that they had previously examined the prints, and instead were led to believe that the prints came from a high-profile case of erroneous identification. The examiners were asked to ignore the background information about the case, examine the prints, and make a conclusion about whether the prints matched. With this new contextual information in mind, four out of five of the examiners (i.e., 80%) concluded that the fingerprints were a definite non-match, despite previously concluding that the fingerprints were a definite match. Thus, working within the context of this biasing information influenced the experts' evaluations, and ultimately, the conclusions that they drew about the fingerprint evidence (Dror et al., 2006). Several other studies have found that forensic experts are biased by information about the case, including in polygraph examinations, risk assessments of offenders, and even in DNA analyses (Dror & Hampikian, 2011; Elaad et al., 1994; Murrie et al., 2013).

Researchers have also demonstrated that confirmation bias may affect evaluations that triers of fact (i.e., judges and/or juries; hereafter referred to as "triers") make at trial. In one study, researchers presented participants with two handwriting samples—one that was purportedly left at the crime scene by the perpetrator and one that came from the accused's handwritten *Miranda* waiver. Before evaluating the handwriting samples, half the participants learned that the accused had confessed, but later retracted his confession because it was coerced. The remaining participants learned that the accused had maintained his innocence the entire time. Those participants who learned that the accused had initially confessed perceived the handwriting samples to be more similar than participants who believed the accused had maintained his innocence (Kukucka & Kassin, 2014). Thus, participants' pre-existing beliefs about the suspect's guilt influenced their evaluations of ambiguous evidence that was presented later in the case.

Overall, research has demonstrated the effects of this so-called "forensic confirmation bias" on decision-making at several stages of the legal process, including in investigations, interrogations, eyewitness identifications, forensic analyses, and at trial (Kassin et al., 2013). Confirmation bias might be particularly likely to occur within this context given what Kassin (2012) has deemed a *corroboration inflation.* This occurs when legal decision-makers encounter apparently incriminating evidence (e.g., confession evidence) and it influences the way in which they perceive and evaluate subsequently encountered ambiguous evidence. That is, they tend to perceive other ambiguous evidence as being highly indicative of guilt. This makes it appear as though the ambiguous evidence corroborates the incriminating evidence.

Importantly, this is most likely to occur when legal decision-makers must evaluate *ambiguous* evidence in the face of pre-existing beliefs. In other words, when evidence is more "elastic," it is more easily influenced by contextual or extraneous information (Ask et al., 2008). Therefore, the ambiguity of evidence moderates the confirmation bias effect. Thus, while it is generally unlikely that individuals would interpret clearly exculpatory DNA evidence as consistent with their pre-existing belief in a suspect's guilt (although, see Dror & Hampikian, 2011 for evidence that confirmation bias can even affect DNA analyses in some cases), they might interpret facial composite evidence or hard-to-hear audio recordings as incriminatory (Charman et al. 2009, Fraser, 2018).

1.4. Audio Recordings as Evidence in Legal Cases

Due to the growing prevalence of recording devices that are easily accessible and readily available (e.g., cell phones, computers, body cams, “smart speakers” like Amazon Echo or Google Home, etc.), it has become increasingly common for investigators to seek, recover, and present evidence from these technologies. Consider, for example, a 2015 case in the U.S. After Victor Collins was found dead in James Bates’ hot tub in Arkansas, the prosecution requested recordings from Bates’ Amazon Echo (McLaughlin, 2017). The Amazon Echo is a smart speaker that constantly records everything in its vicinity, waiting to hear its “wake word” (“Alexa” or “Amazon”). Thus, even when the speaker is not prompted to listen to the user’s request, it is recording. These recordings are then stored remotely by Amazon. This is the first case in which lawyers have sought the recordings from Amazon. The company refused the request, stating that “unless the Court finds that the State has met its heightened burden for compelled production of such materials,” the privacy implications at stake were too great. Though Bates eventually consented to Amazon producing the recordings, nothing came of the recordings and the case was dropped. However, this case represents a first in what may be a future ripe with evidence recorded by smart speakers such as the Amazon Echo. It is unclear how the legal system will deal with future requests for evidence recorded from these devices. Yet, audio recorded by other devices has served as evidence in legal cases for decades (Fishman, 2006).

The first U.S. Supreme Court case to address the admissibility of covertly-recorded conversations occurred in 1963 (*Lopez v. United States*). The court held that a recording that was covertly obtained by a U.S. Internal Revenue Agent was properly admitted in evidence. In their decision, the court stated that so long as one party in the recording consents to the conversation being recorded in advance, law enforcement officers do not need a warrant to covertly record the conversation. Since this case, U.S. federal law, 38 states, and the District of Columbia have accepted this “one-party consent" criterion (Matthiesen et al., 2018). Furthermore, an increasing number of criminal cases in several countries have included audio recordings as key evidence against the accused, including U.S. and Canadian cases (see e.g., *Golden v. State, 1983*; *R. v. Demeter*, 1975; *R. v. Fegan*, 1993; *R. v. Strano*, 2001; *R. v. Randall and Weir,* 1983; *State v. Trask*, 2007). Importantly, in some of these cases, the audio recordings were degraded.

Canadian courts have handled audio evidence in much the same way as U.S. courts. According to Section 184(1) of the Canadian Criminal Code, it is illegal to willfully intercept a private conversation. However, exceptions to this rule exist, including cases in which one of the parties in the communication is aware that they are being recorded. Thus, so long as the person recording the communication is part of the conversation, the recording is legal (s. 184(2)(a)). The Supreme Court of Canada upheld that intercepting a private communication is lawful as long as one of the parties consents to the recording in the *R. v. Goldman* (1980) case. They noted that "the consent may be express or implied and may be given by either the originator of the private communication or the intended recipient." Police officers can also obtain a warrant to intercept private communications expected to reveal evidence of a criminal act if there are reasonable grounds to believe that a crime has been, is being, or will be committed (*R. v. Madrid*, 1994).

While audio recordings will typically be procured as evidence in criminal cases, they can appear in civil cases as well. For example, audio recordings sometimes appear in child custody cases in which one parent records the other parent's conversations with their child (see, e.g., *Cacciarelli v. Boniface*, 1999). So long as the parent can establish that they recorded the conversations out of concern for the child's welfare, most courts will admit such evidence. In some child custody cases, courts may even admit illegally obtained recordings. For example, courts may decide that the probative value of the audio recordings in helping to determine the best interests of the child outweighs the prejudicial effect of such evidence (*Droit de la famille,* 2015). Thus, the admissibility of covertly-recorded conversations as evidence has implications for both criminal and civil cases.

It should also be noted that audio recordings may serve as a source of evidence in cases involving "earwitnesses." Unlike an eyewitness who identifies a suspect based on their visual memory of a crime, an earwitness identifies a suspect based on their auditory memory of a crime (Clifford, 1980; Yarmey, 1995). In some cases involving earwitnesses, individuals are tasked with identifying voices that they previously heard through recording devices such as a wiretap or phone call (Kerstholt et al., 2006). In other cases, earwitnesses who originally heard the voice of a perpetrator in person must try to identify a suspect from audio-recorded voices (Sherrin, 2015). Thus, audio recordings can also form the basis of earwitness identifications, though this is beyond the scope of the current research.

Additionally, with many court proceedings shifting to virtual formats throughout the COVID-19 pandemic, this presented even further potential for poor-quality audio to be introduced into the legal process. Technological challenges such as Wi-Fi lags or poor recording conditions (e.g., background noise, inadequate audio equipment, etc.) may have led to low-quality audio. Unfortunately, Bild et al. (2021) found that poor audio quality caused participants to rate witnesses as less credible, reliable, and trustworthy. Furthermore, participants were less likely to consider the witness' testimony in their final guilt decisions when the audio was poor in quality. Again, this is beyond the scope of the current investigation but has important implications for how audio evidence may be handled in the future.

1.4.1. Degraded Audio Recordings

Audio recordings are particularly probative given that few forms of evidence are as convincing as hearing an accused incriminate himself. However, there are several issues that make audio recordings challenging to handle as evidence. Some of the issues concern listeners' ability to distinguish the identity of the speakers, interpret slang or code words, or understand foreign accents. Other issues include the quality of the recording itself. If the recording device is concealed or far away from the speaker, the recordings may be muffled, distorted by background noises, or otherwise hard to hear (Fishman, 2006; see also Sherrin, 2015 for evidence that many of these issues also apply to earwitness identifications). Despite the potential for these issues, if there is suspicion that the degraded recordings may contain probative evidence, they may ultimately serve as evidence in a legal case. Consequently, the party offering the recordings as evidence (most typically the prosecution) has the burden of proof of establishing that the recordings are *audible* and *intelligible*. Part of establishing these criteria is demonstrating that the party can provide transcripts of the recordings (Fishman, 2006). Thus, when recordings are hard to interpret, transcripts might be very influential on individuals' evaluations of the recordings.

For example, in the Canadian case *R. v. Bennett* (2011), the judge noted that the audio recordings that the Crown presented were poor in quality and that written transcripts of the recordings were "valuable and necessary tools" for interpreting the recordings. The judge even went as far to say that "it is absolutely necessary for proper disclosure and for the reasonable conduct of any trial" to produce written transcripts (*R. v. Bennett*, 2011 at para. 53). This shows how influential transcripts can be in cases involving degraded audio evidence. Additionally, in the U.S. case *Golden v. State* (1983), the judge ruled that written transcripts were necessary for the jurors to be able to interpret the recordings. Therefore, written excerpts of the recorded statements were projected on a screen as the degraded recordings were played aloud to the jury. The written excerpts that were displayed represented the most incriminating content that was purported to be in the recordings. These transcripts, presented along with the degraded audio recordings, undoubtedly influenced the jury's belief in the accused's guilt.

In the Australian case *R. v. Clark* (2008), a man was convicted of being an *accessory to murder before the fact* largely based on degraded audio recordings and incriminating police transcripts that a forensic phonetics expert later showed to be unreliable (after the final appeal was denied; Fraser, 2018). Before arresting the grandson of an elderly man that was murdered, the police put a covert recording device in the family's home. The police asserted that one of the recorded statements proved that the father of the son who committed the murder was an accessory to murder *before* the fact. They claimed that the father said, "at the start we made a pact," and therefore, masterminded the murder. The father admitted that his son had confessed to him after committing the murder and that he did not report this to the police. In this case, the father was an accessory to murder *after* the fact, a much lesser charge. However, based on their knowledge that the son had confessed, the police interpreted the father's statement as "at the start we made a pact," and further alleged that this statement was evidence that the father and son planned the murder together. Thus, the prosecution charged him with being an accessory to murder *before* the fact based on the police transcripts of the degraded audio recordings. He was sentenced to 30 years in prison, 10 years longer than the 20-year sentence his son received.

While presenting written transcripts to listeners would be less problematic if the transcripts accurately reflect what is being said in the recordings, there will almost always be subjectivity inherent in interpreting degraded audio recordings. This is especially true when transcriptionists know the context from which the recordings were taken (Lange et al., 2011). Consequently, issues arise when there are discrepancies between the recordings and the transcripts. Individuals may assume that the transcripts accurately reflect what is being said in the recordings and subsequently, the transcripts may influence how individuals interpret the recordings. This is particularly problematic if the transcripts contain incriminating errors which increase individuals' beliefs in an accused's guilt.

1.4.2. Confirmation Bias and Degraded Audio Recordings

Audio recordings are often used as critical evidence in criminal cases (Fraser, 2018; Lange et al., 2011). However, forensic audio evidence frequently appears degraded for many reasons, including wiretaps, background noise, and poor-quality connections in phone calls. Given the ambiguous nature of such degraded audio recordings, people's pre-existing beliefs or knowledge may influence how they interpret this evidence. Support for this assertion comes from research on the *phonemic restoration effect*, which demonstrates that context or knowledge can influence how we interpret words when there are sounds, or phonemes, missing from those words. Whether sounds are missing due to background noise or other distortion, the brain can oftentimes fill in those missing sounds automatically. For example, if people hear the sentence "The *eel was on the orange" with the phoneme prior to the fragment "eel" missing, they will likely hear the word "peel." However, if they hear "The *eel was on the axle," they will be more likely to report hearing "wheel" (Warren, 1984). Thus, people can interpret the words despite the missing sounds, and sometimes will even fail to notice that the sounds are missing (Samuel, 1981, 1996; Warren, 1970). This occurs because of top-down processing, in which context or knowledge guides our perception of sensory stimuli (Warren, 1970). While much is known about phonemic restoration, few studies have investigated how cognitive biases, including confirmation bias, affect individuals' evaluations of degraded audio recordings within a forensic context. Because audio recordings are becoming increasingly prevalent as evidence in criminal cases, it is essential to understand the implications of presenting this evidence to people with contextual knowledge about the cases, particularly given our understanding of top-down processing.

Of the limited research in this area, studies have shown that individuals' contextual knowledge can influence their interpretations of degraded audio recordings. In one study, Lange et al. (2011) found that participants who believed they were listening to statements from criminal suspects' interviews were more likely to misinterpret non-incriminatory statements (e.g., "I never meant to *charm* her") as incriminating (e.g., "I never meant to *harm* her") compared to participants who believed they were listening to job candidates' interviews or who were given no contextual information. Thus, individuals interpret audio recordings in a way that reflects their expectations about the information contained in the recordings.

When transcriptionists learn contextual information about the degraded audio recordings they are transcribing, they may misinterpret the evidence and produce inaccurate transcripts. If these inaccurate transcripts accompany the degraded audio recordings at trial, triers will use the transcripts and likely misinterpret the audio recordings. Indeed, in a follow-up study, Lange et al. (2011) also found that compared to participants who read accurate transcripts or no transcripts of degraded audio recordings, those who read transcripts containing incriminating errors were more likely to misinterpret the statements as incriminating. Additionally, Fraser (2018) found that when participants were told to listen for a particular phrase in a degraded audio clip (either

"Adelaide bank account" or "at the start we made a pact"), they were more likely to transcribe that particular phrase than when they transcribed the statements without explicitly listening for that phrase. While this provides preliminary evidence that confirmation bias influences interpretations of degraded audio recordings in a forensic context, more research is needed.

1.5. The Current Research

There is a tenet in the criminal justice system that different lines of evidence in a case are independent (Hasel & Kassin, 2009). Thus, different lines of evidence presented at trial are presumed to have been collected and evaluated independently, uninfluenced by irrelevant contextual information. However, research shows that cognitive biases arising from contextual information such as pre-existing beliefs about a suspect's guilt affect decision-making at several stages of the legal process (Kassin et al., 2013). Confirmation bias is especially prevalent and consequential in forensic settings because investigators, forensic examiners, lawyers, and triers are particularly likely to be exposed to contextual information about a case that may influence how they evaluate evidence, though it should not (e.g., knowing that a suspect confessed may influence how they evaluate fingerprint evidence; however, these lines of evidence should be evaluated independently). Given the potential for erroneous legal decisions that implicate innocent people, it is vital that we understand how confirmation bias operates in legal proceedings so that we can develop approaches to minimize its impact. There is little research in this area, particularly regarding how confirmation bias affects people's evaluations of degraded forensic audio evidence.

The goal of the current research was to provide a more comprehensive understanding of when and why confirmation bias operates for audio recordings as well as explore methods for mitigating this bias. Thus, I conducted three studies designed to meet each of these research objectives. In Study 1, I replicated and extended previous research (Lange et al., 2011) to explore what factors lead to confirmation bias when individuals interpret degraded audio recordings. In Study 2, I explored whether fluency misattribution is a mechanism underlying confirmation bias for degraded audio recordings. In Study 3, I examined whether the Canadian Model Jury Instructions for audio recordings and written transcripts can effectively mitigate confirmation bias. Prior to conducting these studies, I pilot tested a large set of innocuous degraded audio recordings.

Overall, this work has methodological, theoretical, and applied implications for the study of confirmation bias. I contributed to methodology by developing a large set of degraded audio recordings that elicit confirmation bias given incriminating contextual information. Across three studies, I demonstrated that both a general criminal context and written transcripts containing incriminating errors caused participants to make incriminating misinterpretations of the audio recordings. I advanced theory by including an independent measure of fluency to reveal that fluency misattribution contributes to confirmation bias. Finally, I contributed to practice by exploring the different types of forensically-relevant context that affect participants' evaluations of degraded audio recordings. I also demonstrated that, under some circumstances, the existing Canadian Model Jury Instructions for audio recordings and written transcripts can mitigate the biasing effects of context in these cases.

Chapter 2.

Pilot Studies

2.1. Pilot 1

Prior to conducting Studies 1-3, I developed a set of innocuous degraded audio recordings and performed pilot testing to ensure that my materials were suitable. There were two primary objectives of Pilot 1. First, I wanted to verify that participants would misinterpret the innocuous recordings as incriminating in the presence of incriminating contextual information. Thus, half the participants learned that the recordings came from wiretapped conversations with criminal suspects; the remaining participants did not receive any contextual information about the recordings. The second goal of Pilot 1 was to develop a set of *minimally* degraded recordings (i.e., participants correctly transcribe roughly 60-85% of the statements) and *moderately* degraded recordings (i.e., participants correctly transcribe roughly 30-55% of the statements) to be used in Studies 1-3.

2.1.1. Stimuli Generation

I developed a set of 84 innocuous statements (see Appendix A). Forty-eight of these statements contained a target word that could be misinterpreted as incriminating given forensically-relevant contextual information (e.g., "Next thing I knew, there was *mud* everywhere" could be misinterpreted as "Next thing I knew, there was *blood* everywhere"). The remaining 36 statements were filler statements in which incriminating misinterpretations were not expected (e.g., "We both liked to work on cars together"). I recruited eight male speakers to record each of the 84 statements. I chose to record only male speakers because males are more likely to commit violent crimes (Statistics Canada, 2021). Speakers were instructed to speak clearly into a microphone in an emotionally neutral tone.

Once I gathered clear versions of all 84 statements from the eight speakers, I used the audio-editing software Audacity® to edit the recordings. First, I normalized the volume of all the statements. Then, to create the perception that the recordings came from a larger group of speakers, I changed the pitch of the recordings to make the voices sound different. Depending on the speaker's normal pitch, I changed their pitch to be between 5-15% higher or lower. I adjusted the percent change in pitch to the point at which the voice sounded like a normal male voice rather than a voice distortion. Some voices did not lend themselves well to pitch change (e.g., they sounded like manipulated voices rather than genuine voices) and therefore, I did not include these additional "voices" in the recordings. Overall, changing the pitch of the voices resulted in six additional "speakers" (i.e., voices that sounded like genuine people rather than voice distortions).

Next, I degraded all the statements using low-pass filters (Bernstein et al., 2012; Lange et al., 2011). Low-pass filters simulate distorted audio obtained from imperfect recordings by allowing sounds with frequencies lower than a specific cut-off frequency to pass through the filter, while blocking sounds with frequencies higher than the specific cut-off frequency. For example, when using a low-pass filter, if the cut-off frequency is 1,000 Hz, sounds with a frequency below 1,000 Hz will "pass" through the filter, while sounds with a frequency above 1,000 Hz will be blocked by the filter. As a result, people will only hear the sounds that are below 1,000 Hz which have passed through the filter. Cutting off certain frequencies of sounds associated with consonants in language makes it more challenging to identify what consonants are being said. People typically begin having trouble identifying consonant sounds at a cut-off frequency of 2,000 Hz (Sher & Owens, 1974). The lower the cut-off frequency, the harder it is to identify the words.

Using low-pass filters, I degraded the recordings at six levels: 1,600 Hz (least degraded), 1,400 Hz, 1,200 Hz, 1,000 Hz, 800 Hz, 600 Hz (most degraded). The degradation levels were selected based on Lange et al.'s (2011) findings. They found that 1,000 Hz was the level of degradation at which the stimuli were ambiguous enough to elicit incriminating misinterpretations but not too degraded that the recordings were completely uninterpretable. Thus, I created the six different degradation levels by setting one of the middle levels of degradation at 1,000 Hz and creating equal increments of 200 Hz between each level. In total, there were 504 recordings from each speaker (including the six additional "speakers" I generated) and 7,056 recordings in total.

2.1.2. Methodology

Design

I conducted a 6 (degradation level: 1,600 Hz; 1,400 Hz; 1,200 Hz; 1,000 Hz; 800 Hz; 600 Hz) x 2 (contextual information: no context; criminal suspects' conversations) mixed design with degradation level as the within-subject factor.

Participants

I conducted a power analysis in G*Power 3.1 to determine the sample required for Pilot 1 (Faul et al., 2009). I assumed a medium effect based on previous research that has demonstrated a medium effect of incriminating contextual information on participants' incriminating misinterpretations of audio recordings ($OR = 4.56$; Lange et al., 2011, Experiment 1). Thus, with $f = .25$, $\alpha = .05$, Power = .95, number of groups = 2 and number of measurements = 6, I required a sample of $N = 124$.

I recruited 168 participants from Amazon Mechanical Turk (MTurk). MTurk is an online subject pool through which individuals participate in online experiments in exchange for payments. Studies have demonstrated that the data gathered from MTurk participants is comparable to that from laboratory participants (Buhrmester et al., 2011; Germine et al., 2012; Mason & Suri, 2012). To qualify for the study, participants had to live in the United States (U.S.) or Canada, have a 95% approval rating, and have successfully completed at least 100 tasks on MTurk. I used these latter two inclusion criteria to filter out any "bots" that are on MTurk (Chmielewski & Kucker, 2020). Participants received $3.00 USD for their participation.

Procedure

Once participants signed up for the study, they were re-directed from MTurk to Qualtrics. They first completed a Completely Automated Public Turing Test to Tell Computers and Humans Apart (Captcha) Verification. Then, participants completed a quick hearing test. First, they listened to a clear audio recording ("I spent all weekend outside working in my garden") and typed what they heard. Then, across 8 trials, participants listened to tones at each of the 6 frequencies used in the experiment and indicated if they heard the tone. A tone played in 6 of the 8 trials and 2 of the trials were silent.

Each participant then heard all 84 statements at one of the 6 degradation levels (i.e., they heard 14 statements per degradation level). The 6 degradation levels were counterbalanced across participants such that all 14 statements were presented equally at each level of degradation. The statements were played in a fixed order that was randomly determined. Half the participants received no contextual information about the recordings before they played. The remaining participants learned that the recordings came from wiretapped conversations with criminal suspects that have been used as evidence in criminal trials. Participants were instructed to transcribe each statement to the best of their ability.

To determine whether participants were paying attention, I included a variant of the Instructional Manipulation Check (IMC; Oppenheimer et al., 2009). The IMC measures whether participants are paying attention to and following the instructions. To test this, experimenters embed a question within the experimental materials that resembles the other questions in terms of length and response format. However, instead of responding as they have to the other questions, participants confirm that they have read the instructions. Oppenheimer et al. (2009) found that using an IMC improves statistical power by eliminating noise due to participants who respond randomly. Some research has also found that MTurk participants pass the IMC at much higher rates than traditional university samples (Hauser & Schwarz, 2016). While there are currently no guidelines on the number of IMCs that researchers should include among their experimental items, Berinsky et al. (2014, 2016) demonstrate the importance of using multiple attention check questions throughout an experiment. They argue that including/excluding participants based on only one item assumes that there are participants who "*always* pay attention" and participants who "*never* pay attention" (Berinsky et al., 2014, p. 747). Thus, I included five attention check questions throughout the experiment to ensure that participants were completing the study diligently.

Participants saw the following instructions before hearing each statement: "Please listen to the following statement. When the statement is finished, type the statement as you heard it in the box. Once you have finished typing, click continue." On the first IMC question, participants instead saw the following instructions before listening to the statement: "Please listen to the following statement. When the statement is finished, type the letter "a" 3 times in the box. Once you have finished typing, click continue." On the second, third, fourth, and fifth IMC questions, instead of being prompted to type the letter "a" three times, they were prompted to type the letter "h," the letter "e," the letter "m," and the letter "x" three times, respectively. I presented the IMC questions in a fixed random order throughout the experiment. Finally, participants completed demographic questions (see Appendix B). The study took approximately 30 minutes for participants to complete.

2.1.3. Results

I originally intended to exclude participants who failed three or more of the IMC questions. However, I received feedback from multiple participants that the questions were "unfair" because throughout the study, the instructions generally did not vary from page to page and participants became accustomed to listening to the statements and transcribing what they heard. Because the instructions varied only slightly on the IMC questions and an audio recording still played, many participants did not catch that they were supposed to type three letters instead of transcribing the statement. Participants noted that even though they failed to type the letters on the screen, this did not mean that they were failing to pay attention to the task. Indeed, the very nature of the task (i.e., transcribing audio recordings) required that participants were paying attention to the task. Thus, I ultimately chose to retain participants who failed three or more of the IMC questions so long as they transcribed the statements. However, I excluded participants who provided nonsense responses (e.g., responded "good" on every trial; 5.4%) or who withdrew from the study early (23.8%). Overall, 119 participants were included in the analyses.

For my research, it was most important to identify the degradation levels at which participants in the criminal suspects' conversations condition made more incriminating misinterpretations of the target statements than participants in the no context condition. Therefore, I first examined differences in the proportion of participants who made incriminating misinterpretations for each of the target statements in the no context and criminal suspects' conversations conditions (see Appendix C). I found that for several of the target statements, participants in the criminal suspects' conversations condition did not make many incriminating misinterpretations at any of the degradation levels. Additionally, they made several accurate interpretations across the degradation levels.

To proceed with Studies 1-3, I needed to ensure that my target statements would elicit incriminating misinterpretations in the presence of incriminating information. Thus, I decided to conduct a second pilot study with modified materials. Specifically, given the relatively high proportion of accurate interpretations across degradation levels, I chose to further degrade the audio recordings. Additionally, I chose to add more target statements to increase the likelihood of identifying 48 suitable target statements. Finally, given that I did not reach my target *N* of 124 in Pilot 1 after applying my exclusion criteria, I aimed to power the second pilot appropriately.

2.2. Pilot 2

There were four major aims of Pilot 2. First, I increased the degradation of the recordings to determine whether this increased participants' tendency to make incriminating misinterpretations when they received incriminating contextual information about the recordings. Second, I aimed to identify the degradation level at which participants made more incriminating misinterpretations in the criminal suspects' conversations condition than the no context condition for each of the statements. Third, I aimed to develop a set of minimally degraded recordings (i.e., those for which participants were able to identify roughly 60-85% of the statements) and moderately degraded recordings (i.e., those for which participants were able to identify roughly 30-55% of the statements) to use in Study 2, where I intended to manipulate perceptual fluency at different levels. Finally, I aimed to select 48 target statements to use as stimuli in Studies 1-3.

2.2.1. Stimuli Generation

After Pilot 1, I made the decision to drop the filler statements from my studies. I originally intended to use the filler statements in Study 3, but I modified the Study 3 paradigm and no longer required the filler statements. Thus, in Pilot 2, I only included target statements in the materials. I generated 13 additional target statements for Pilot 2, which replaced 13 of the original target statements used in Pilot 1 (see Appendix D). Most of these additional statements replaced statements with similar target words (e.g., "I made sure he was *buying* [dying]" replaced "I looked in his eyes and I could tell he was *lying* [dying]."

The original male speakers recorded these additional statements in the manner described above. Again, I used Audacity® to normalize the new recordings and change the pitch of the voices to create a larger set. Then, I degraded all the new statements using low-pass filters ranging from 1,600 Hz (least degraded) to 600 Hz (most degraded). Next, I aimed to further degrade all the audio recordings. However, using low-pass filters with cut-off frequencies lower than 600 Hz made some of the statements relatively indecipherable by cutting off too many sounds. Thus, I chose to further degrade the audio recordings by using the distortion filter in Audacity®. This filter distorts the waveform of the audio so that the recordings sound distorted. I applied the "Soft Overdrive" distortion filter to all recordings.

2.2.2. Methodology

Design

I conducted a 6 (degradation level: 1,600 Hz; 1,400 Hz; 1,200 Hz; 1,000 Hz; 800 Hz; 600 Hz) x 2 (contextual information: no context; criminal suspects' conversations) mixed design with degradation level as the within-subject factor. The distortion filter was applied to the recordings at all levels of degradation.

Participants

Once again, I conducted a power analysis in G*Power 3.1 to determine the sample required for Pilot 1 (Faul et al., 2009). With f = .25, α = .05, Power = .95, number of groups = 2 and number of measurements = 6, I required a sample of N = 124. I recruited 219 participants from MTurk. To participate in the study, participants were required to live in the U.S. or Canada, have a 95% approval rating, and have successfully completed at least 100 tasks on MTurk. Participants were compensated $3.00 USD for their participation. Sixty-three participants were excluded for the following reasons: (1) they provided nonsense responses (5.02%); (2) they withdrew early (23.3%); or (3) they had more than 25% missing data (0.46%). Thus, data from 156 participants were included in the analyses.

Procedure

Participants were re-directed from MTurk to Qualtrics to complete the study. They first completed a Captcha Verification followed by the same hearing test described in Pilot 1. Then, participants either received no contextual information about the recordings they would hear, or they learned that the recordings came from wiretapped conversations with criminal suspects that have been used as evidence in criminal trials. Participants then heard 48 target statements at one of the 6 degradation levels (i.e., they heard 8 statements per degradation level). The 6 degradation levels were counterbalanced across participants such that all 48 statements were presented equally at each level of degradation. The statements played in a fixed order that was randomly determined. Participants were instructed to transcribe each statement to the best of their ability. There were five attention check questions throughout the experiment that required participants to do simple math problems (e.g., “What is 6 x 2?”). Finally, participants completed a few demographic questions. The study took approximately 30 minutes for participants to complete.

2.2.3. Results

I first examined differences in the proportion of participants who made incriminating misinterpretations for each of the statements in the no context and criminal suspects' conversations conditions (see Appendix E). I used the following criteria for selecting statements and degradation levels to use in Studies 1 and 3: (1) the proportion of incriminating misinterpretations participants made in the no context condition was less than half of what participants made in the criminal suspect condition; and (2) if there were multiple degradation levels for which this criterion was met, I generally selected the degradation level with the largest ratio of incriminating misinterpretations in the criminal suspects' conversations condition compared to the no context condition.[1] The degradation levels I selected for Studies 1 and 3 are highlighted in blue in Appendix E. Across the statements and degradation levels selected for Studies 1 and 3, the average percentage of each statement correctly identified ranged between 29.08% and 95.82% with an average of 70.46%.

Next, I selected a set of minimally and moderately degraded recordings for Study 2. For minimally degraded statements, I selected the degradation levels for which all participants (i.e., in both context conditions) were able to correctly identify roughly 60-85% of the statement. Target words (e.g., "mud") had to be accurately identified in the statements in order to be considered correct and target word misinterpretations (e.g., "blood") were considered incorrect. The minimally degraded statements selected for Study 2 are highlighted in yellow in Appendix E. For moderately degraded statements, I selected the degradation levels for which participants were able to identify roughly 30-55% of the statement. The moderately degraded statements selected for Study 2 are highlighted in green in Appendix E. There were some cases in which no identification percentages fell between 30-55%, and in these cases, only the minimally degraded recordings are highlighted. For these statements, I degraded the recordings to 450 Hz and included this as the degradation level for the moderately degraded condition in Study 2.

---

[1] I occasionally selected recordings that did not meet this second criterion if participants were more likely to make the expected target word misinterpretation (e.g., misinterpret mud for blood) at one degradation level compared to the other.

Chapter 3.

Study 1

Degraded audio recordings are increasingly being used as evidence in civil and criminal trials (Fishman, 2006; Fraser, 2018; Lange et al., 2011). Therefore, it is essential to understand what factors lead to confirmation bias when individuals interpret degraded forensic audio evidence. When degraded audio recordings are presented as evidence at trial, judges may permit written transcripts to accompany the audio evidence. In these cases, individuals may rely on the transcripts when attempting to interpret the degraded audio evidence. Yet, even when written transcripts are not permitted to accompany audio evidence, triers understand that the audio recordings are being used as evidence in a criminal trial. Triers may rely on this contextual information when evaluating the degraded audio evidence and erroneously interpret the recordings as incriminatory.

Support for this assertion comes from a study conducted by Lange et al. (2011). The researchers demonstrated that participants who believed they were listening to recordings from criminal suspects' interviews were more likely than participants given no contextual information to misinterpret innocuous degraded audio recordings as incriminatory. This suggests that when participants have trouble processing what the speaker is saying in degraded recordings, they rely on contextual information about the source of the recordings. Study 1 was designed to replicate this finding using degraded audio recordings that were purportedly being used as evidence in criminal trials.

Furthermore, triers will likely learn about other lines of evidence besides the audio evidence in a criminal trial (e.g., eyewitness evidence, confession evidence, forensic science evidence, etc.). It remains unclear whether these other lines of evidence can also influence individuals' perceptions of the degraded audio evidence. Specifically, research has yet to explore whether there is an additive or multiplicative effect of presenting additional lines of evidence on people's tendency to make incriminating misinterpretations. Thus, Study 1 was also designed to test whether presenting additional lines of evidence increases the likelihood that individuals will misinterpret innocuous degraded audio recordings as incriminatory.

3.1. Methodology

I pre-registered this study on Open Science Framework prior to collecting data (https://doi.org/10.17605/OSF.IO/T7Z4S).

3.1.1. Participants

I conducted an a priori power analysis in G*Power 3.1 to determine the sample size required for Study 1 (Faul et al., 2009). Previous research has demonstrated a medium effect of incriminating contextual information on participants' incriminating misinterpretations of audio recordings (OR = 4.56; Lange et al., 2011, Experiment 1). However, given the new stimuli and different contextual information, I assumed a more conservative, small-to-medium effect. Thus, with f = .20, α = .05, Power = .95, and number of groups = 3, I required a sample of N = 390.

I recruited 718 participants through MTurk (n = 598) and the Simon Fraser University (SFU) research pool (n = 120). The MTurk sample was restricted to those individuals living in Canada or the U.S. Additionally, any individuals who participated in the pilot studies were not permitted to participate in Study 1 given that I used the same stimuli. Finally, MTurk participants must have achieved at least a 95% approval rating and successfully completed at least 500 tasks on MTurk to participate. Participants recruited through MTurk received $3.00 USD in exchange for their participation while participants recruited through SFU received 2 course credits.

I excluded 298 participants from analyses for the following reasons: (1) they withdrew early (100 participants; 13.93%); (2) they didn't follow the instructions to transcribe the recordings or gave nonsense responses (75 participants; 10.45%)[2]; (3) they failed the manipulation check (101 participants; 14.07%); (4) they scored below 50% on

[2]These responses were likely bots or individuals who were not fluent in English.

the LexTALE measure of English language fluency (17 participants; 2.37%); or (5) more than 50% of their data was missing (5 participants; 0.70%). Thus, a total of 417 participants were included in the analyses.

The average age of participants was 32.36 years (*SD* = 12.00 years; *Range* = 18 – 70 years). Additionally, 50.6% of participants identified as female, 48.4% identified as male, and 0.9% identified as non-binary. Furthermore, 70.3% of participants where White, 15.4% were Asian (including 4.6% East Asian, 4.1% South Asian, and 1.9% Southeast Asian), 6.2% were Black, 1.9% were Hispanic, 0.2% were Latinx, 1.0% were Middle Eastern, 0.2% were Indigenous, 3.8% were Mixed Race, and 0.5% did not report their ethnicity.

3.1.2. Design

Study 1 was a single factor design with three levels. The independent variable was the type of contextual information that participants received. Participants in the *no context* condition learned no contextual information about the recordings before listening to the degraded audio and transcribing what they heard. Participants in the *criminal suspect* condition learned that the recordings came from wiretapped conversations with criminal suspects that were used as evidence in criminal trials prior to listening to the recordings. Finally, participants in the *criminal suspect + eyewitness evidence* condition learned that the recordings came from wiretapped conversations with criminal suspects that were used as evidence in criminal trials in which there were also eyewitnesses who identified the accused persons from lineups prior to listening to the recordings.

The dependent variable was the proportion of incriminating misinterpretations that participants made. From this measure, I could determine whether participants exhibited confirmation bias by comparing the interpretations of participants without incriminating contextual information to the interpretations of participants who learned incriminating contextual information. There would be evidence of confirmation bias if participants with contextual information systematically made more incriminating misinterpretations than participants without contextual information. Additionally, the more incriminating misinterpretations participants with contextual information made above and beyond those made in the no context condition, the greater confirmation bias they would exhibit.

3.1.3. Audio Recordings

I selected 48 degraded audio recordings based on the results of the pilot studies (see items with asterisks in Appendices A and D). Each statement contained a target word that could be misinterpreted as incriminating given what contextual information participants learned about the recordings (e.g., "I checked to make sure he wasn't *grieving*" could be misinterpreted as "I checked to make sure he wasn't *breathing*"). Degradation levels ranged from 600 Hz to 1,600 Hz across statements. I determined which degradation level to present for each statement by identifying the degradation levels for which the proportion of incriminating misinterpretations participants made in the no context condition was less than half of what participants made in the criminal suspect condition in the pilot studies. The statements were presented across two randomly determined fixed orders.

3.1.4. Procedure

MTurk and SFU participants completed the study online through Qualtrics. At the beginning of the study, participants completed a Captcha Verification question followed by a short hearing test. As in the pilot studies, participants indicated whether they could hear tones ranging from 600 Hz – 1,600 Hz across several trials and transcribed a clear audio-recorded statement. Then, participants proceeded to the main part of the study. They learned that they would be listening to degraded audio recordings of spoken statements and that these statements might be challenging to hear. One third of participants received no additional contextual information about the recordings. One third of participants also learned that the recordings they were listening to came from wiretapped conversations with criminal suspects that have been used as evidence in criminal trials. The remaining participants also learned that the recordings they were listening to came from wiretapped conversations with criminal suspects that have been used as evidence in criminal trials in which there were also eyewitnesses who identified the accused persons from lineups.

Then, participants heard degraded audio recordings of 48 statements. After hearing each statement, participants attempted to transcribe the statement to the best of their ability. The procedure also included five attention checks (i.e., simple math questions) throughout the experiment. After transcribing all statements, participants completed the LexTALE measure of English language fluency (Lemhöfer & Broersma, 2012). Participants saw a series of 60 stimuli, including 40 words (e.g., "recipient") and 20 nonwords (e.g., "exprate") and simply indicated whether the target was a word by responding "yes" or "no." Participants' responses were scored as the average number of items answered correctly, calculated as: ((number of words correct/40*100) + (number of nonwords correct/20*100)) / 2. Finally, participants completed a demographic questionnaire. This questionnaire also included a manipulation check question, which asked participants to select the instructions they saw at the beginning of the experiment from a multiple-choice list. The study took approximately 45 minutes in total.

3.1.5. Coding Participants' Transcriptions

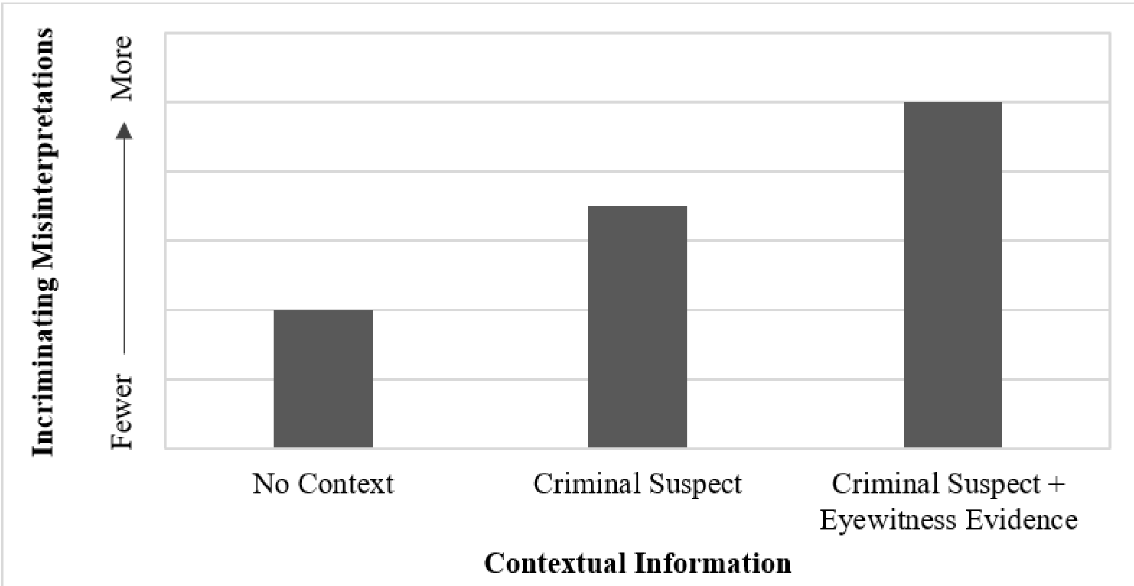
I, along with another independent, condition-blind coder coded participants' transcriptions into the following categories: (a) invalid responses; (b) accurate interpretations; (c) non-incriminating errors; and (d) incriminating misinterpretations. These categories represent all possible responses. Invalid responses included blank responses, don't know responses, and transcriptions with fewer than 3 words. Transcriptions were coded as accurate interpretations if participants correctly identified every word in a given statement. I adopted these strict criteria for accurate interpretations from Lange et al. (2011) because misinterpreting even one word can have serious implications for one's overall interpretation of the statement. Transcriptions were coded as non-incriminating errors if they contained at least one error that was not incriminating to the speaker. Transcriptions were coded as incriminating misinterpretations if participants provided a transcription which appeared to implicate the accused (i.e., the speaker). This occurred either when participants misinterpreted the target word as incriminating (e.g., they interpreted the statement “I made sure he was *buying*” as “I made sure he was *dying*”) or they otherwise misinterpreted the statement in a way that implicated the speaker (e.g., they misinterpreted “I made sure he was buying” as “I shot all of the bodies”).

I instructed the other coder on how to code the responses. Then, for the purposes of training, we each coded the same 10% of participants’ responses. We compared our codes, discussing where we disagreed and settling on agreed-upon codes. We did not track disagreements on this set of codes as the purpose of this training phase was to ensure that we understood how to consistently assign codes to participants’ responses. After this training phase, we independently coded approximately 30% of the responses (31.43%) and then worked together to compare codes for the purposes of inter-rater reliability. We noted and discussed disagreements to determine an agreed-upon code. Inter-rater reliability was high (kappa = .94). We divided the remaining responses among the two of us and individually coded the remaining responses.

3.2. Hypotheses

I expected that participants in the no context condition would make fewer incriminating misinterpretations than participants in the criminal suspect condition, who in turn would make fewer incriminating misinterpretations than participants in the criminal suspect + eyewitness evidence condition (see Figure 1 for hypothetical data pattern).

Figure 1 **Hypothetical Data Pattern for the Proportion of Incriminating Misinterpretations that Participants will Make as a Function of Contextual Information in Study 1**



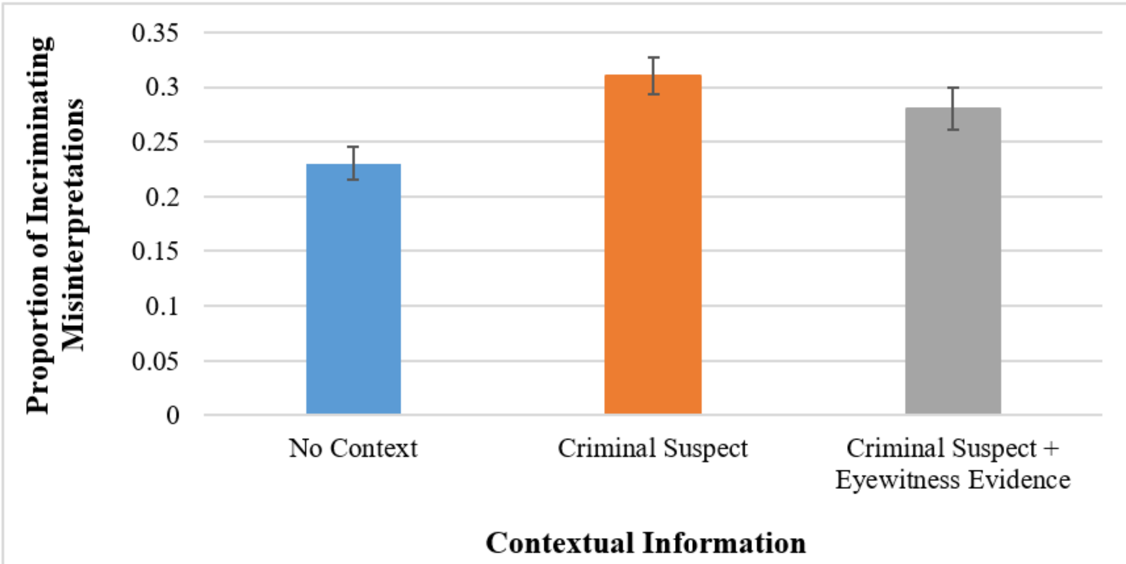
3.3. Results

First, to determine whether there were differences across the SFU and MTurk samples, I conducted an Analysis of Covariance (ANCOVA) with type of contextual information and sample as the independent variables, proportion of incriminating misinterpretations as the dependent variable, and LexTALE score as the covariate. Sample did not interact with contextual information (p = .956), indicating that there were no differences across SFU (n = 93) and MTurk participants' data (n = 324). Thus, I conducted all remaining analyses collapsed across sample. A one-way ANCOVA with type of contextual information as the independent variable, proportion of incriminating misinterpretations as the dependent variable, and LexTALE score as the covariate revealed a main effect of contextual information, $F(2, 413) = 29.06, p < .001, \eta_p^2 = .12$ (see Figure 2).

To supplement this analysis, I also conducted a Bayesian ANCOVA. Bayesian analyses provide a numerical value called the Bayes Factor (BF), which expresses evidence in favor of either the null hypothesis or the alternative hypothesis. BFs can either be expressed as BF_{10}, which indicates the BF in favor of the alternative hypothesis over the null hypothesis, or as BF_{01}, which indicates the BF in favor of the null hypothesis over the alternative hypothesis. A BF_{10} of 1 suggests no evidence for either hypothesis. A BF_{10} above 1 indicates increasing evidence in favor of the alternative hypothesis with values between 1-3 indicating inconclusive evidence, values between 3-10 indicating substantial evidence, and values above 10 indicating strong evidence. A BF_{10} below 1 indicates increasing evidence in favor of the null hypothesis with values between .33-1 indicating inconclusive evidence, values between .10-.33 indicating substantial evidence, and values between .03-.10 indicating strong evidence (Dienes, 2014; Lee & Wagenmakers, 2014; Wetzels et al., 2011). The Bayesian ANCOVA revealed a BF_{10} of 1.47e+9 in favor of the alternative hypothesis, indicating that the data are 1.47 billion times more likely to occur under the alternative model than the null model. Thus, there is strong evidence for there being a true difference in participants' tendency to make incriminating misinterpretations based on the contextual information participants learned about the recordings.

Post-hoc tests with Bonferroni adjustments revealed that, consistent with my hypothesis (Figure 1), participants in the no context condition (M = .23) made fewer incriminating misinterpretations than participants in the criminal suspect condition (M = .31), $p < .001$, $d = 0.76$, 95% CI [0.06, .0.12] (BF_{10} = 1.20e+10). Additionally, participants in the no context condition made fewer incriminating misinterpretations than participants in the criminal suspect + eyewitness evidence condition (M = .28), $p < .001$, $d = 0.50$, 95% CI [0.03, 0.09] (BF_{10} = 3,872.45). However, in contradiction to my hypothesis, I found that participants in the criminal suspect condition made more incriminating misinterpretations than participants in the criminal suspect + eyewitness evidence condition, $p =.039$, $d = 0.29$, 95% CI [0.00, 0.06] (BF_{10} = 2.15).

Figure 2 **Proportion of Incriminating Misinterpretations Participants Made as a Function of Contextual Information in Study 1**



Note. Error bars are 95% Confidence Intervals.

3.4. Discussion

The purpose of Study 1 was to replicate previous research by Lange et al. (2011) which demonstrated that incriminating contextual information biases people's interpretations of degraded audio recordings. Additionally, I explored whether presenting multiple pieces of contextual information increased the amount of bias (i.e., the proportion of incriminating misinterpretations) that participants exhibited. Consistent with Lange et al.'s (2011) findings, my results show that participants who learned incriminating contextual information before listening to a series of degraded audio recordings made more incriminating misinterpretations than participants with no prior contextual information. This was evident in the fact that participants in both the criminal suspect and criminal suspect + eyewitness evidence conditions made more incriminating misinterpretations than participants in the no context condition. This demonstrates that participants are susceptible to confirmation bias when evaluating degraded audio. That is, once they have a pre-existing belief about the context of the recordings, they are more likely to interpret the recordings in a manner that is consistent with this belief. This supports previous research that shows the pervasiveness of confirmation bias in a variety of forensic domains (Charman et al., 2017; Dror & Hampikian, 2011; Findley & Scott, 2006; Kassin et al., 2013).

While the presence of incriminating contextual information elicited confirmation bias for the degraded recordings, including additional contextual information did not increase the confirmation bias effect. In fact, it appeared to suppress it. That is, participants in the criminal suspect + eyewitness evidence condition made significantly fewer incriminating misinterpretations than participants in the criminal suspect condition. However, one important factor to note is that the Bayesian analyses suggest inconclusive evidence for there being a true difference between the criminal suspect and criminal suspect + eyewitness evidence conditions ($BF_{10} = 2.15$). Thus, it is possible that learning that the recordings came from criminal suspects' conversations created enough of a bias that no additional context further increased this effect.

Alternatively, perhaps there is something about the addition of eyewitness evidence specifically that made participants more skeptical about the nature of the recordings. As books, shows, movies, and podcasts featuring wrongful convictions based on faulty eyewitness testimony (e.g., *The Innocence Files*, *When They See Us*, *Making a Murderer*, etc.) have become increasingly popular, it is possible that the general population has developed an increased awareness of the issues associated with eyewitness evidence. Indeed, mistaken eyewitness identifications are the largest contributor to wrongful convictions (Innocence Project, n.d.). Thus, participants may have been more dubious about this type of evidence overall, which could have suppressed their tendency to make incriminating misinterpretations. To determine whether there is something unique about eyewitness evidence that truly suppresses the confirmation bias effect in evaluations of degraded audio recordings, future research should replicate this finding as well as explore whether other types of evidence (e.g., confession evidence) affect participants' tendency to make incriminating misinterpretations in the same manner.

It is also possible that learning about the eyewitness evidence did not increase participants' tendency to make incriminating misinterpretations due to the format in which this information was presented in the current study. For the purpose of this study, I was interested in determining whether contextual information influenced participants' interpretations of a series of audio recordings from various speakers across various circumstances (e.g., types of crimes). Thus, participants were simply told that the recordings came from wiretapped conversations with criminal suspects that had been used as evidence in criminal trials in which there were also eyewitnesses who identified the accused persons from lineups. It is possible that learning more information about the circumstances of the eyewitness identification evidence (e.g., details about the lineup composition) would more substantially impact participants' interpretations. Some research has found that more detailed eyewitness testimony leads to higher perceptions of eyewitness credibility and accused guilt than less detailed testimony (Bell & Loftus, 1988). Thus, perhaps more details about the eyewitness evidence would similarly impact interpretations of other evidence (e.g., audio recordings). To explore this question, future research can investigate whether, in the context of a single case with multiple lines of detailed evidence (e.g., in the form of a case summary with more specific details about the individual case), participants exhibit more confirmation bias. However, this may be challenging to test within the context of a single case given that there must be enough target statements to test for incriminating misinterpretations. Thus, limiting the number of target statements would make it more challenging to experimentally test the confirmation bias effect.

While future research should continue to explore whether learning more contextual information leads to more bias, this study clearly demonstrates that incriminating contextual information elicits robust confirmation bias for degraded audio recordings. This has important implications for audio recordings used as evidence in criminal trials. Even when recordings are completely innocuous, they can be misinterpreted as incriminating given participants' pre-existing beliefs about the source of these recordings. This becomes even more likely in the case of stimuli that are ambiguous, such as degraded audio recordings that are hard to hear (Ask et al., 2008). Given that audio evidence is often collected under suboptimal conditions (e.g., wiretapping, concealed recording devices, etc.), it is plausible that these recordings may be muffled or otherwise hard to hear. Additionally, when recordings are being used as evidence, participants may already have a pre-existing belief that the recordings contain incriminating content. This bias can lead to misinterpretations, and unfortunately, may contribute to wrongful convictions. It is important to reveal the issues that can arise from using this type of evidence in criminal cases.

Overall, Study 1 replicated Lange et al.'s (2011) finding that participants who learn incriminating contextual information make more incriminating misinterpretations of degraded audio recordings than participants without contextual information. This is evidence of confirmation bias. Additionally, there was inconclusive evidence regarding whether including additional contextual information increases confirmation bias. To further investigate why participants exhibit confirmation bias for degraded audio recordings, I designed Study 2 to explore fluency misattribution as a mechanism for confirmation bias in these cases.

Chapter 4.

Study 2

Previous research has demonstrated that participants who learn that recorded statements come from criminal suspect interviews tend to misinterpret the innocuous degraded audio recordings as incriminating (Lange et al., 2011, Experiment 1). Additionally, when participants receive written transcripts of the recordings that contain incriminating errors, they also tend to misinterpret the innocuous recordings as incriminating (Lange et al., 2011, Experiment 2). However, research has not yet explored the mechanism(s) underlying confirmation bias in these cases. Thus, the goal of Study 2 was to investigate possible mechanisms that underlie confirmation bias in cases involving degraded audio evidence.

While there are several mechanisms that may underlie confirmation bias (see Table 1), top-down processing and fluency misattribution are two particularly plausible candidates in cases involving degraded audio evidence. When individuals must process information in the face of uncertainty (e.g., interpret recordings that are hard to hear), they often rely on their existing knowledge or expectations to guide their interpretation (Gregory, 1974). If individuals have been primed with contextual information about the degraded recordings, top-down processing may guide their interpretations, with context driving their perception of the degraded audio (Gregory, 1970). Previous research has demonstrated that this process influences people's interpretations of distorted audio (Lange et al., 2011; Samuel, 1981; Warren, 1970). Yet, the mechanism through which top-down processing leads to confirmation bias for degraded audio remains unclear.

Given that top-down processing occurs when context drives perception, it is possible that ambiguous stimuli are processed more quickly and easily when contextual information is available than when it is not (Whittlesea, 1993). Thus, I propose that top-down processing leads to confirmation bias for degraded audio through *fluency misattribution*. Fluency misattribution occurs when people misattribute the subjective ease with which they process information to an incorrect source (e.g., they misattribute the subjective ease with which an incriminating interpretation comes to mind as objective evidence that this must be the correct interpretation). To explore the rationale for why fluency misattribution may be a mechanism for confirmation bias in cases involving degraded audio recordings, consider a related cognitive bias called *auditory hindsight bias*. In a typical auditory hindsight bias paradigm, there are two knowledge trials. In the naïve identification trials, individuals hear a series of degraded words and attempt to identify these words. In the hindsight estimation trials, individuals hear clear words before hearing degraded versions of the same words. Then, they estimate what percentage of their peers would be able to identify the degraded words had their peers not heard the clear versions of the words. Individuals exhibit auditory hindsight bias when, in hindsight, they estimate that a greater percentage of their peers could identify the degraded words than the percentage of words they were able to correctly identify in the naïve identification trials. Thus, participants are unable to ignore their knowledge of the clear words (Bernstein et al., 2012).

Recent work has demonstrated that auditory hindsight bias results from fluency misattribution: Individuals process the degraded words more fluently, or easily, in the hindsight trials because they have just heard clear versions of the words. However, they wrongly misattribute the ease with which they processed the degraded words in the hindsight trials to the degraded words being easy to identify (Higham et al., 2017). Bernstein et al. (2018) demonstrated that both repetition priming in an exposure phase (i.e., hearing clear words either 0, 1, 3, or 6 times) and clear presentation of the target word immediately before the degraded word independently increased participants' tendency to overestimate how many of their peers could identify degraded words. However, there was not an additive effect of these two sources of fluency. That is, when participants heard clear versions of target words immediately before hearing degraded versions of the same words in the test phase, there was no additional effect of priming presentations in the exposure phase on participants' estimates of how many peers would be able to identify the degraded words. This interactive data pattern suggests that both hindsight bias and repetition priming share fluency as a common mechanism. Although the authors did not include an independent measure of fluency, they demonstrated that the increased number of prime presentations in the exposure phase increased participants' tendency to identify the distorted words at test—thus, there was indirect evidence for increased fluency.

Auditory hindsight bias differs from confirmation bias for degraded audio recordings in terms of the task and ultimate judgment individuals make. In auditory hindsight bias paradigms, individuals must attempt to ignore their current knowledge to imagine how easily they, or someone else, could identify the degraded audio. Thus, participants are explicitly instructed to ignore their current knowledge to consider what was known in the past. Alternatively, confirmation bias for degraded audio recordings arises when individuals with contextual knowledge make decisions about what they currently hear in the audio. They generally do not attempt to suppress their contextual knowledge (although, they could be instructed to do so). However, auditory hindsight bias also resembles confirmation bias for degraded audio recordings. Both biases are consequences of priming; that is, individuals' knowledge influences their perception and evaluation of degraded audio (Tulving & Schacter, 1990). Whether through perceptual priming in the case of auditory hindsight bias (i.e., prior exposure to the clear word) or conceptual priming in the case of confirmation bias for degraded audio recordings (i.e., prior exposure to context), individuals are primed to hear something in the degraded audio recordings, and therefore, they tend to hear it (Vaidya et al., 1997). Thus, it is possible that auditory hindsight bias and confirmation bias for degraded audio recordings share a common mechanism: Fluency misattribution.

Previous research has typically measured fluency through either objective or subjective measures of how effortful it is to process information or generate judgments (Alter & Oppenheimer, 2009; Schwarz, 2004). Several studies have measured fluency as the speed or accuracy with which participants process information (Reber et al., 2004). It is generally believed that the amount of time it takes for people to identify a stimulus (i.e., their reaction time) reflects the ease with which they were able to process the stimulus (Oppenheimer, 2008). For example, Undorf et al. (2017) measured perceptual fluency as the amount of time it took for participants to identify stimuli that gradually clarified over time. They found that the less time it took for participants to identify a stimulus, the more likely they were to believe they would remember the stimulus (i.e., higher judgments of learning). The quicker reaction times suggest that participants processed the stimuli more quickly and easily (i.e., fluently) than stimuli that took them longer to identify. Thus, fluency is often operationalized as reaction time. Other studies have measured fluency through people's subjective perceptions of how quickly or accurately they were able to process information (Schwarz, 2004). For example, participants may provide a rating of how easily they were able to perceive a target stimulus.

There is an interplay between two aspects of fluency that is inherent to evaluating audio evidence. The first aspect is *perceptual fluency*, which refers to the ease with which individuals process stimuli given manipulations to perceptual quality (Jacoby & Dallas, 1980; Johnston et al., 1985; Mandler, 1980). Variables such as figure-ground contrast, font readability, and the duration for which stimuli are presented affect the ease with which participants process perceptual characteristics of the stimuli (Jacoby et al., 1989; Schwarz, 2004). These manipulations to perceptual fluency, in turn, have been shown to affect judgments such as truth, liking, and confidence (Alter & Oppenheimer, 2009). In the current paradigm, perceptual fluency is related to how easily individuals can process the audio recordings. When participants hear clear audio recordings, they should experience high perceptual fluency because they are able to clearly and easily hear the contents of the recordings. However, when participants hear degraded audio recordings, they should experience less perceptual fluency because it becomes more challenging to clearly or easily discern what the speakers are saying.

The second aspect of fluency that may contribute to individuals' judgments is *conceptual fluency,* which refers to the ease with which individuals process the meaning of a stimulus (Whittlesea, 1993). Variables such as priming individuals with semantically related concepts, the consistency between stimuli and their context, and stimulus prototypicality affect individuals' ability to process stimuli (Rhodes et al., 2001; Schwarz, 2004; Tversky & Kahneman, 1973). Overall, individuals seem to process stimuli more easily if they have a better concept of what the stimuli are and what they mean. For example, if participants believe that they are listening to recordings from a criminal suspect's interview, interpretations consistent with what a criminal might say in an interview may come to mind more easily. Conversely, individuals may fail to consider other interpretations of the recordings because they do not come to mind as easily. Thus, when individuals have contextual information about the audio recordings, they should experience greater conceptual fluency.

It is impossible to completely distinguish perceptual from conceptual fluency for degraded audio recordings. This is because even when the recordings are degraded, individuals have a conceptual understanding of the words that they do hear in the recorded statements, and therefore, experience some conceptual fluency. If I were to completely isolate perceptual fluency, I would have to use stimuli such as uncommon words or even non-words, which would be irrelevant to this study. Additionally, even though I expect that participants will experience heightened conceptual fluency when given more contextual information about the recordings, they will still experience some conceptual fluency without contextual information because they have a concept of the words in the recordings. Therefore, for the purposes of my research, I do not try to separate perceptual from conceptual fluency.

Overall, I argue that at least two things must occur for individuals to exhibit confirmation bias for degraded audio recordings. First, individuals must experience some perceptual disfluency when listening to the recordings. This is necessary because stimuli must be ambiguous for participants to rely on contextual information, and thus, exhibit confirmation bias (Ask et al., 2008; Tversky & Kahneman, 1974). If individuals can process the recordings perceptually fluently, they will likely rely on their perceptual experience of the recordings to interpret the statements over the context of the recordings. Second, individuals must rely on the contextual information to derive meaning from the recordings. Thus, the contextual information should make a belief-consistent interpretation of the recordings come to mind more easily than an alternative interpretation. Consequently, individuals may misattribute the ease with which a belief-consistent interpretation comes to mind as evidence that it is the correct interpretation of the degraded audio.

4.1. The Current Study

The degree to which participants exhibit confirmation bias should depend on the extent to which the audio recordings are degraded as well as the amount of contextual information upon which participants have to base their interpretation. To test whether fluency misattribution is a mechanism for confirmation bias in cases involving degraded audio recordings, I manipulated perceptual fluency (i.e., the degree to which the recordings were degraded) as well as conceptual fluency (i.e., the amount of contextual information participants received). Specifically, I presented either non-degraded, minimally degraded, or moderately degraded audio recordings to test whether the degree to which the recordings were degraded affected participants' perceptual fluency, or the ease with which they interpreted the recordings. Additionally, participants either received no contextual information, learned that the recordings came from wiretapped conversations with criminal suspects that had been used as evidence in criminal trials, or received written transcripts containing incriminating errors. Thus, I investigated whether the amount of contextual information available affected participants' conceptual fluency, or the ease with which they processed the meaning of the audio recordings. I acknowledge that it is impossible to completely separate perceptual from conceptual fluency because there is an interplay between these two aspects of fluency inherent to these stimuli (Jacoby, 1991). However, it is theoretically important to begin to investigate the mechanisms that underlie confirmation bias in these cases, and thus, to test fluency misattribution as a contributing process. Therefore, I manipulated overall processing fluency and measured the effect on both reaction time (i.e., my independent measure of fluency) and participants' tendency to exhibit confirmation bias.

4.2. Methodology

I pre-registered this study on Open Science Framework prior to collecting data (https://doi.org/10.17605/OSF.IO/DNBH2).

4.2.1. Participants

I conducted an a priori power analysis in G*Power 3.1 to determine the sample size required for this study (Faul et al., 2009). Previous research has demonstrated a medium effect of incriminating contextual information on participants' incriminating misinterpretations of recordings (OR = 4.56; Lange et al., 2011, Experiment 1). Given that previous research has not tested the effects of reaction time (i.e., my independent measure of fluency) on confirmation bias for degraded audio recordings, I assumed a more conservative, small-to-medium effect. Thus, with f = .20, α = .05, Power = .95, and number of groups = 3, I required a sample of at least 390 participants. However, because I planned to conduct moderated mediation analyses as well, I wanted to ensure that I would have enough participants to detect an effect. Thus, I used a Monte Carlo power analysis (https://schoemanna.shinyapps.io/mc_power_med/) and determined that I required a sample of 450 participants to detect a small effect (*r* = .2) with 95% power.

I recruited 549 participants through MTurk (*n* = 426) and the SFU undergraduate research pool (*n* = 32). The MTurk sample was restricted to individuals living in Canada or the U.S. Additionally, any individuals who participated in previous studies using these stimuli were not permitted to participate. Finally, MTurk participants must have achieved at least a 95% approval rating and successfully completed at least 500 tasks on MTurk to participate. MTurk participants received $8.00 USD for their participation and SFU students received 2% course credits in exchange for their participation.

Ninety-one participants (16.58%) were excluded from analyses for the following reasons: (1) they did not provide responses to the baseline typing speed questions (5.83%); (2) they failed the manipulation check (6.19%); or (3) they had missing data for more than 25% of trials (4.55%). Thus, a total of 458 participants were included in the final analyses. The average age of participants was 37.10 years (*SD* = 11.58 years; *Range* = 17 – 71 years). While 0.87% of participants did not report their gender, 44.57% of participants identified as female, 54.35% identified as male, and 0.22% identified as non-binary. Additionally, 71.74% of participants were White, 9.13% were Asian (including 1.09% East Asian, 1.52% South Asian, and 1.30% Southeast Asian), 8.04% were Black, 4.56% were Hispanic, 0.87% were Latinx, 0.43% were Middle Eastern, 0.43% were Indigenous, 4.13% were Mixed Race, and 0.65% did not report their ethnicity.

4.2.2. Design

I conducted a 3 (degradation level: non-degraded; minimally degraded; moderately degraded) x 3 (contextual information: no context; criminal suspect; written transcripts) mixed factorial design with degradation level as the within-subject factor. Participants in the written transcripts condition received more contextual information than participants in the criminal suspect condition, who in turn received more contextual information than participants in the no context condition. Minimally degraded recordings included recordings for which participants were able to identify roughly 60-85% of the words in a given statement as determined in Pilot 2. Moderately degraded recordings included recordings for which participants were able to identify roughly 30-55% of the words in a given statement. There were a few statements from Pilot 2 for which identification percentages were higher than the 30-55% range for all degradation levels, including 600 Hz (see Appendix E). For these statements, I degraded the recordings to 450 Hz and included this as the degradation level for the moderately degraded condition.

The primary dependent variable was the proportion of incriminating misinterpretations that participants made. There would be evidence of confirmation bias if participants in the criminal suspect and written transcripts conditions made more incriminating misinterpretations than participants in the no context condition. I also collected reaction time data as an independent measure of fluency. I measured reaction time as the amount of time that passed in milliseconds from the offset of the audio recording to the time participants hit the enter key once they finished typing a statement. I controlled for varying lengths of statements by dividing the reaction time for each statement by the number of characters in a given statement.

4.2.3. Procedure

I conducted Study 2 online using the E-Prime Go software. E-Prime is a stimulus presentation software for designing experiments and collecting data electronically. It allows for precise collection of reaction time. E-Prime Go is an extension of E-Prime that allows users to host the experiment online through E-Prime's website and share links to the experiment with participants. After signing up for the study, participants were directed to a Qualtrics page where they completed the consent process. Once participants gave their informed consent to participate, they were redirected to an E-Prime Go webpage, where they were able to download the study. They completed the study in the E-Prime Go software, which was downloaded locally to their computer. At the end of the experiment, their data was automatically uploaded to the E-Prime Go data repository specific to the lab's account.

Participants first completed a hearing check, in which they were prompted to type what they heard in an audio clip that was played. This also served as a "bot check"—that is, if people gave nonsense responses to this question, I assumed they were a bot. Then, participants completed three items that measured their baseline typing speed. For each baseline typing speed item, participants saw a passage of text containing three sentences that was between 32-40 words long. They had to type the text as quickly and accurately as possible. Then, participants were instructed that they would be hearing a series of degraded audio recordings that they would need to transcribe as quickly and accurately as possible. One third of participants received no additional contextual information about the recordings (no context condition). One third of participants learned that the recordings came from wiretapped conversations with criminal suspects that have been used as evidence in criminal trials (criminal suspect condition). The remaining participants were given written transcripts of the recordings that contained incriminating errors (written transcripts condition; see Appendix F).

Participants heard 42 audio recordings that varied in degradation level. There were an equal number of non-degraded, minimally degraded, and moderately degraded recordings.[3] The different degradation levels were counterbalanced, such that all recorded

[3] There was a programming error which caused 46 participants in the criminal suspect condition to hear the same statement twice at the minimally degraded level. Thus, they heard 14 non-degraded, 15 minimally degraded, and 13 moderately degraded recordings, with two of the recordings in the minimally degraded

statements had an equal chance of appearing at one of the three degradation levels for each participant.

After hearing each statement, participants transcribed the statement to the best of their ability as quickly and accurately as possible. Participants in the written transcripts condition saw each written transcript immediately before the corresponding recorded statement played without the transcript showing. They also received the following instructions at the beginning of the experiment: "Your job is to verify the transcripts by listening to each of the recordings and typing what you hear in the recordings on the screen." The intention of this instruction was to prevent them from simply typing what they saw in the transcript without first listening to the recorded statement. Finally, participants completed the LexTALE measure of English language fluency and a demographic questionnaire that included a manipulation check. The procedure took approximately 45 minutes in total.

4.2.4. Coding Participants' Transcriptions

Three condition-blind coders (including myself) independently coded participants' transcriptions into the following categories as described above: (a) Invalid responses; (b) accurate interpretations; (c) non-incriminating errors; and (d) incriminating misinterpretations. I instructed coders how to code the responses. Then, for the purposes of training, we each coded the same 10% of participants' responses. We compared our

condition being the same. Additionally, 55 participants in the criminal suspect condition heard the same statement twice at the moderately degraded level. Thus, they heard 13 non-degraded, 14 minimally degraded, and 15 moderately degraded recordings, with two of the recordings in the moderately degraded condition being the same. For these participants who heard the same recording twice, I retained their transcription from the first time they heard the statement and removed their transcription from the second time they heard the statement. This resulted in 46 participants hearing 14 non-degraded, 14 minimally degraded, and 13 moderately degraded recordings and 55 participants hearing 13 non-degraded, 14 minimally degraded, and 14 moderately degraded conditions. The remaining 357 participants heard 14 non-degraded, 14 minimally degraded, and 14 moderately degraded statements.

codes, discussing where we disagreed and settling on agreed-upon codes. We did not track disagreements on this set of codes as the purpose of this training phase was to ensure that all coders understood how to consistently assign codes to participants' responses. After this training phase, all three coders independently coded approximately 25% of the responses (24.78%) and then worked together to compare our codes for the purposes of inter-rater reliability. We noted and discussed disagreements to determine an agreed-upon code. Inter-rater reliability was high (Fleiss' kappa = .93). Finally, each coder individually coded approximately one-third of the remaining responses.

4.3. Hypotheses

I had three sets of hypotheses relating to (1) the effects of audio degradation and contextual information on confirmation bias; (2) the effects of audio degradation and contextual information on fluency; and (3) the mediating effects of fluency on confirmation bias. Because I operationalized fluency as participants' reaction time, I will refer to "fluency" as "reaction time" throughout the remainder of this section. Additionally, as this is the first exploration into the mechanisms for confirmation bias in cases involving degraded audio evidence, I acknowledge that some of my hypotheses are not based on prior findings/theory. However, I am using the term "hypotheses" instead of a term like "expectations" because most people in the literature use the term "hypotheses" in this context.

4.3.1. Hypothesis 1

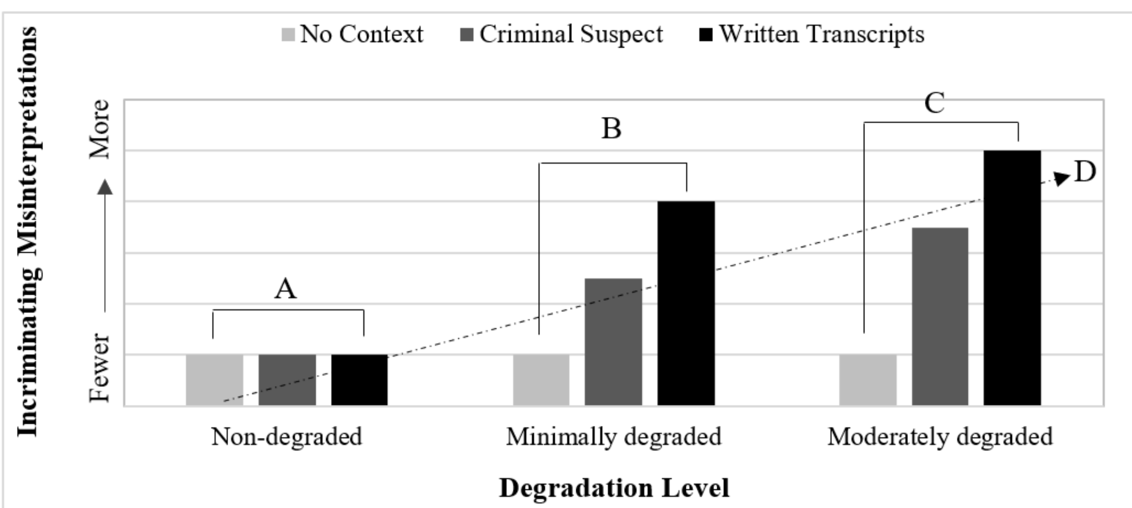
Hypothesis 1a

I expected a main effect of degradation level on participants' tendency to make incriminating misinterpretations. Specifically, I expected that participants who heard non-degraded recordings would make fewer incriminating misinterpretations than participants who heard minimally degraded recordings, who in turn would make fewer incriminating misinterpretations than participants who heard moderately degraded recordings (see D in Figure 3 for hypothetical data pattern).

Hypothesis 1b

I expected an interaction between degradation level and contextual information on participants' tendency to make incriminating misinterpretations, and thus, their tendency to exhibit confirmation bias. Specifically, in the non-degraded condition, I did not expect any differences in incriminating misinterpretations as a function of contextual information (see A in Figure 3). This is because participants were able to clearly hear what was said in the recordings, and thus, should not rely on contextual information to interpret the recordings. In the minimally degraded condition, I expected that participants in the no context condition would make fewer incriminating misinterpretations than participants in the criminal suspect condition, who in turn would make fewer incriminating misinterpretations than participants in the written transcripts condition (see B in Figure 3). In the moderately degraded condition, I expected that participants in the no context condition would make fewer incriminating misinterpretations than participants in the criminal suspect condition, who in turn would make fewer incriminating misinterpretations than participants in the written transcripts condition (see C in Figure 3).

Figure 3 **Hypothetical Data Pattern for the Proportion of Incriminating Misinterpretations that Participants will Make as a Function Degradation Level and Contextual Information in Study 2**



4.3.2. Hypothesis 2

Hypothesis 2a

First, I expected a main effect of degradation level on participants' reaction time. Specifically, I expected that participants' reaction times would increase as recordings became more degraded, with the quickest reaction times for non-degraded recordings and the longest reaction times for moderately degraded recordings (see D in Figure 4 for hypothetical data pattern).

Hypothesis 2b

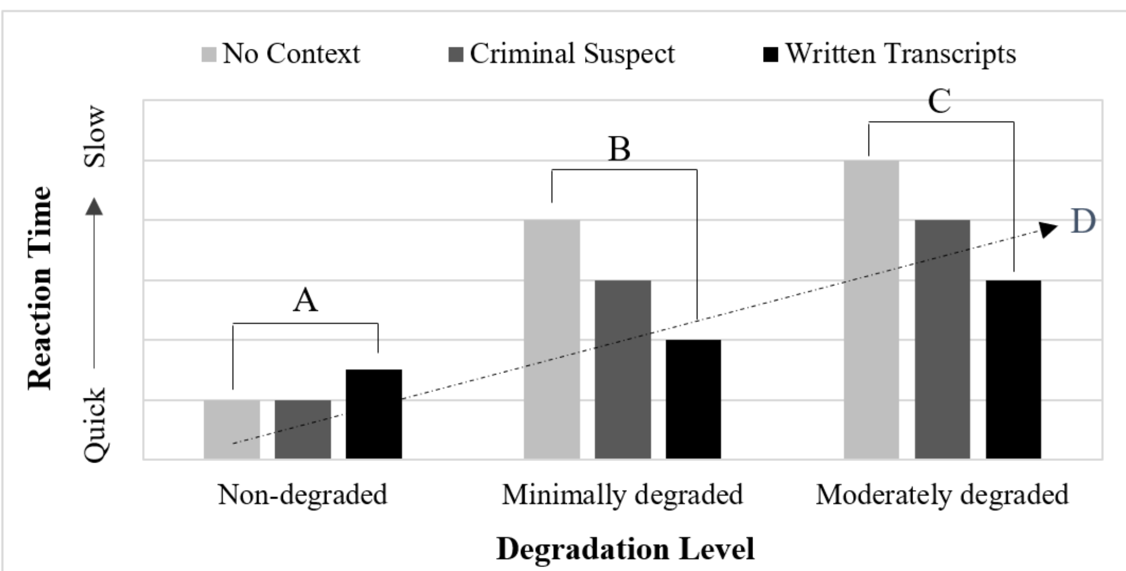
I expected an interaction between degradation level and contextual information on participants' overall reaction time. I will discuss the expected effects for each degradation level. For the non-degraded recordings, I expected that participants would be able to process the recordings fluently given that there should be no disruptions to perceptual fluency. As such, participants should rely on their perceptual experience of the recordings, rather than on contextual information. Thus, participants should have the quickest reaction times for non-degraded recordings, particularly in the no context and criminal suspect conditions. In the written transcripts condition, I expected a slightly longer reaction time given that participants had to reconcile the difference between what they read and what they heard (see A in Figure 4).

For the minimally degraded recordings, I expected that participants would experience some perceptual disfluency. As such, participants should be more likely to utilize contextual information to help them process the audio recordings. Thus, participants in the no context condition should find it most challenging to process the recordings. Therefore, I expected that the no context condition would have the slowest reaction time for minimally degraded recordings. Next, participants in the criminal suspect condition would have some contextual information to rely on to get "the gist" of the recordings, or to help them determine the most likely interpretation given the context from which the recordings were taken. Thus, compared to the no context condition, the criminal suspect condition should have a quicker reaction time. Finally, participants in the written transcripts condition would have the most contextual information to rely on.

Thus, they should process the minimally degraded recordings most fluently, and therefore have the quickest reaction time (see B in Figure 4).

For the moderately degraded recordings, I expected the same data pattern as for the minimally degraded recordings. That is, when participants hear moderately degraded recordings, they should experience perceptual disfluency. Subsequently, they should rely on contextual information to help them process the audio recordings. When participants have no contextual information to rely on, they should have the hardest time processing the recordings. Therefore, the no context condition should have the slowest reaction time. Next, participants in the criminal suspect condition should rely on the contextual information to help them determine the most likely interpretation given the context from which the recordings were taken. Therefore, compared to the no context condition, the criminal suspect condition should have a quicker reaction time because they are able to process the recordings more easily. Finally, I expected participants in the written transcripts condition to have the quickest reaction time because they had the most contextual information to rely on when forming their judgment. Thus, they should process the moderately degraded recordings most fluently (see C in Figure 4).

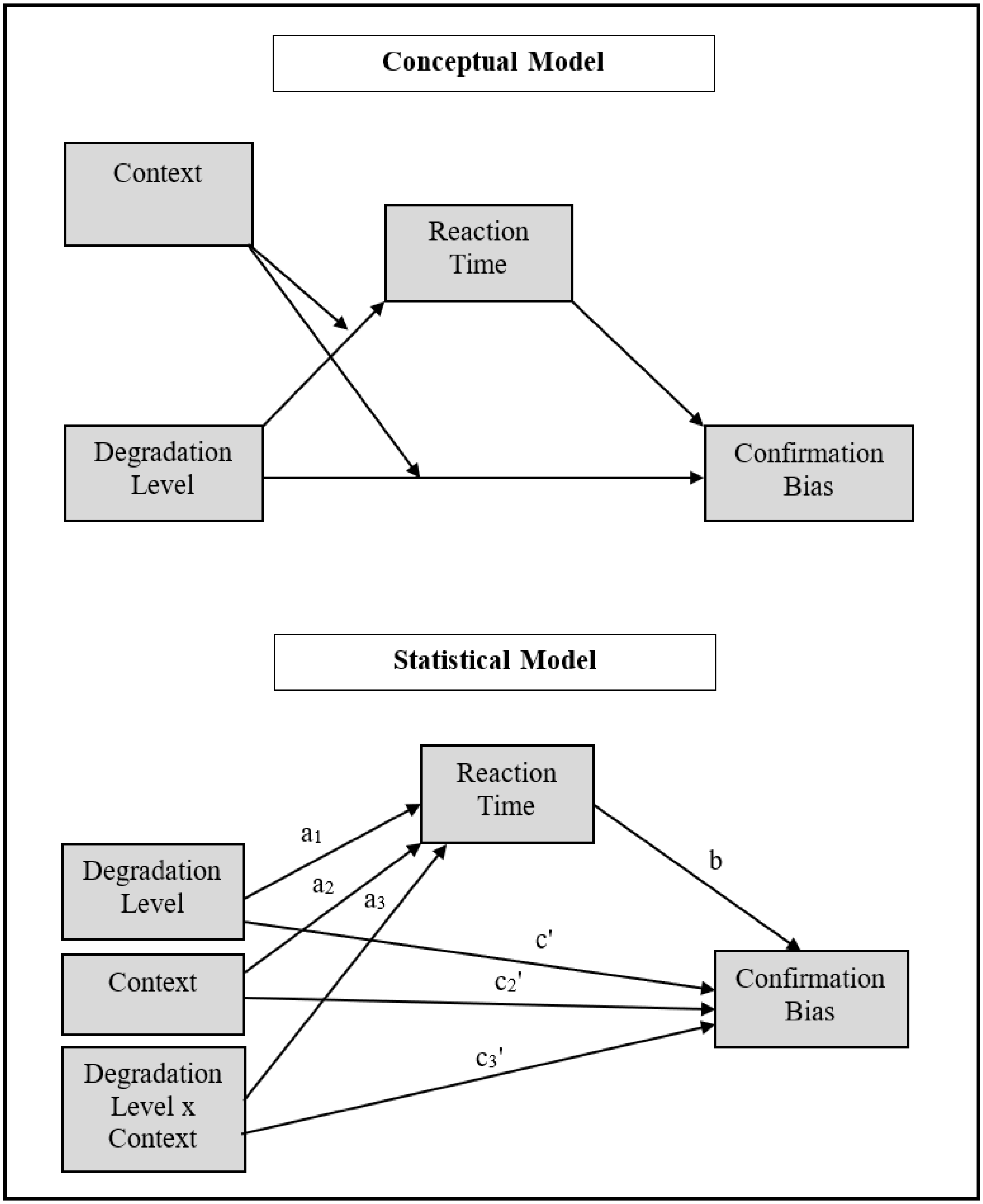
Figure 4 Hypothetical Data Pattern for Participants' Reaction Time as a Function of Degradation Level and Contextual Information in Study 2



4.3.3. Hypothesis 3

I expected reaction time to partially mediate the relationship between degradation level, contextual information, and confirmation bias (see Figure 5 for conceptual and statistical models of this moderated mediation effect). This is because reaction time is a measure of fluency, and I expected that fluency misattribution is one of the mechanisms that underlies confirmation bias. Thus, I expected degradation level, contextual information, and the interaction between degradation level and contextual information to affect the degree to which participants exhibited confirmation bias through differences in participants' reaction time (i.e., fluency). Additionally, I expected that reaction time would only partially mediate the relationship between degradation level, contextual information, and confirmation bias because I suspect that there are other mechanisms that contribute to confirmation bias as well.

Figure 5 The Conceptual and Statistical Models for the Moderated Mediation Analysis in Study 2. Adapted from Hayes, 2015.



4.4. Results

4.4.1. Confirmation Bias

Given the vastly unequal sample sizes of MTurk and SFU participants, it was impossible to conduct a valid test for differences across the samples. However, I did not expect sample-based differences given that I found no evidence for these differences in Study 1. I conducted a 3 (degradation level: non-degraded; minimally degraded; moderately degraded) x 3 (contextual information: no context; criminal suspect; written transcripts) mixed ANCOVA with degradation level as the within-subject factor, proportion of incriminating misinterpretations as the dependent variable, and English language fluency (as measured by the LexTALE scale) as a covariate. I centered participants' LexTALE scores by subtracting the mean LexTALE score from each LexTALE score. This procedure is recommended when conducting ANCOVAs including within-subject factors (Schneider et al., 2015). I also conducted Bayesian ANCOVAs to supplement the null hypothesis significance testing. Bayesian ANCOVAs compare each possible combination of covariates, main effects, and interactions to the null model. However, I typically only want to investigate the overall effect of certain factors or interactions. Thus, when reporting the Bayes factors for these analyses, I report the BF_{INCL} across matched models. This compares models that contain the effect of interest to models that do not to determine the overall likelihood of the data pattern occurring under the alternative distribution versus the null distribution for the effect of interest specifically.

Consistent with Hypothesis 1a (Figure 3), I found a main effect of degradation level, $F(1.96, 888.86) = 420.58$, $p < .001$, $\eta_p^2 = .48$ (BF_{INCL} = 1.01e+109). [4] I also found a main effect of contextual information, $F(2, 454) = 182.05$, $p < .001$, $\eta_p^2 = .45$ (BF_{INCL} = 2.53e+55), which I did not hypothesize. Finally, consistent with Hypothesis 1b (Figure 3), I found a significant interaction between degradation level and contextual information, $F(3.92, 888.86) = 55.57$, $p < .001$, $\eta_p^2 = .20$ (BF_{INCL} = 4.69e+37; see Figure

[4] The assumption of sphericity was violated so I applied a Huynh-Feldt correction.

6).[5] To probe this interaction, I conducted one-way ANCOVAs for each degradation level (i.e., non-degraded; minimally degraded; moderately degraded) with contextual information as the independent variable, proportion of incriminating misinterpretations as the dependent variable, and LexTALE score as the covariate.

For non-degraded recordings, there was a significant effect of contextual information, $F(2, 454) = 58.43$, $p < .001$, $\eta_p^2 = .21$ ($BF_{INCL} = 1.81e+20$). Post-hoc comparisons with a Bonferroni correction revealed that participants in the no context condition ($M = .05$) made significantly fewer incriminating misinterpretations than participants in the criminal suspect condition ($M = .12$), $p < .001$, $d = 0.48$, 95% CI [.03, .10] ($BF_{INCL} = 7.97e+7$) and written transcripts condition ($M = .22$), $p < .001$, $d = 1.21$, 95% CI [.13, .21] ($BF_{INCL} = 1.45e+16$). Furthermore, participants in the criminal suspect condition made significantly fewer incriminating misinterpretations than the written transcripts condition, $p < .001$, $d = 0.69$, 95% CI [.07, .15] ($BF_{INCL} = 248{,}592.66$).

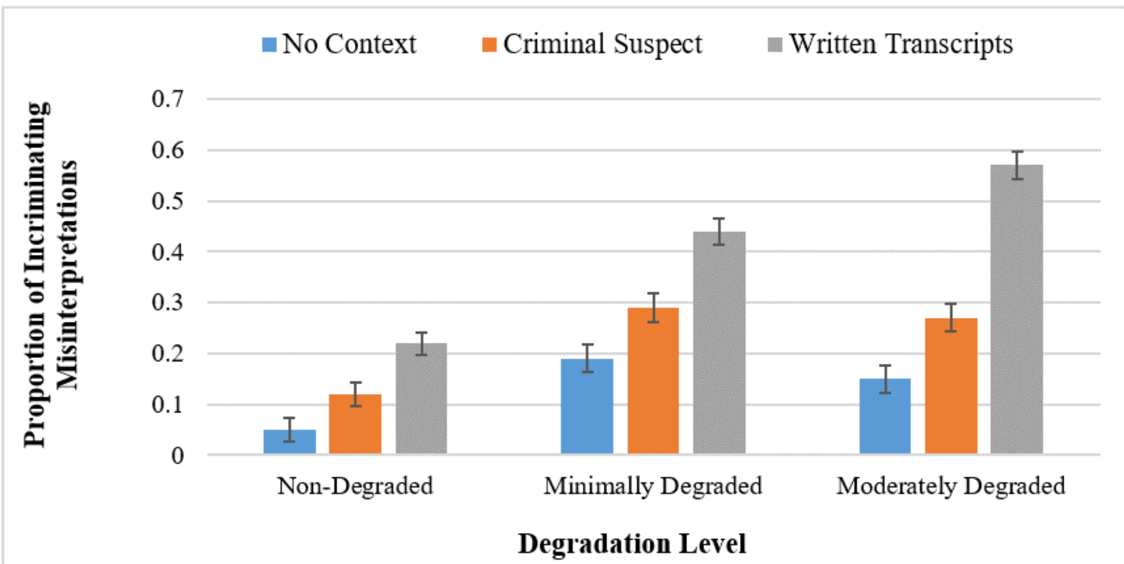
For minimally degraded recordings, there was a significant effect of contextual information, $F(2, 454) = 80.15$, $p < .001$, $\eta_p^2 = .26$ ($BF_{INCL} = 2.86e+27$). Post-hoc comparisons with a Bonferroni correction revealed that participants in the no context condition made significantly fewer incriminating misinterpretations ($M = .19$) than participants in both the criminal suspect condition ($M = .29$), $p < .001$, $d = 0.59$, 95% CI [.05, .15] ($BF_{INCL} = 1.92e+7$), and the written transcripts condition ($M = .44$), $p < .001$, $d = 1.43$, 95% CI [.20, .29] ($BF_{INCL} = 1.58e+23$). Furthermore, participants in the criminal suspect condition made significantly fewer incriminating misinterpretations than participants in the written transcripts condition, $p < .001$, $d = 0.86$, 95% CI [.10, .19] ($BF_{INCL} = 9.23e+7$).

For moderately degraded recordings, there was a significant effect of contextual information, $F(2, 454) = 263.53$, $p < .001$, $\eta_p^2 = .54$ ($BF_{INCL} = 1.57e+73$). Post-hoc comparisons with a Bonferroni correction revealed that participants in the no context condition made significantly fewer incriminating misinterpretations ($M = .15$) than

[5] The assumption of sphericity was violated so I applied a Huynh-Feldt correction.

participants in both the criminal suspect condition ($M = .27$), $p < .001$, $d = 0.71$, 95% CI [.07, .17] ($BF_{INCL} = 5.73e+8$) and the written transcripts condition ($M = .57$), $p < .001$, $d = 2.54$, 95% CI [.37, .46] ($BF_{INCL} = 1.02e+60$). Furthermore, participants in the criminal suspect condition made significantly fewer incriminating misinterpretations than participants in the written transcripts condition, $p < .001$, $d = 1.82$, 95% CI [.25, .35] ($BF_{INCL} = 1.44e+34$).

Figure 6 **Proportion of Incriminating Misinterpretations Participants Made as a Function of Degradation Level and Contextual Information in Study 2**



Note. Error bars are 95% Confidence Intervals.

Given that these analyses did not reveal the source of the interaction, I further probed the interaction by examining whether there were differences in the proportion of incriminating misinterpretations participants made across the degradation levels for each contextual information condition. The interaction appeared to stem from participants in the no context and criminal suspect conditions making the most incriminating misinterpretations for minimally degraded recordings and participants in the written transcripts condition making the most incriminating misinterpretations for moderately degraded recordings. Overall, there is clear evidence for confirmation bias given that across degradation levels, participants in the criminal suspect and written transcripts conditions made more incriminating misinterpretations than participants in the no context condition.

4.4.2. Reaction Time

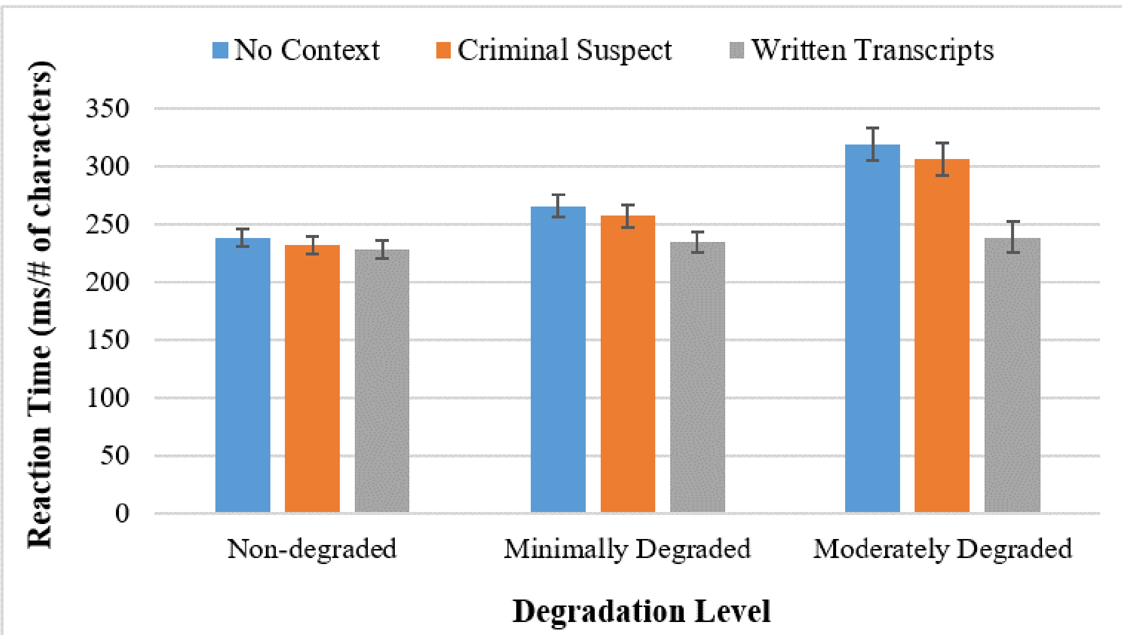
To determine whether fluency is a mechanism for confirmation bias in cases involving degraded audio recordings, I included reaction time as an independent measure of fluency. I measured reaction time as the amount of time that passed in milliseconds from the offset of the audio recording to the time participants hit the enter key once they finished typing a statement. I controlled for varying lengths of statements by dividing the reaction time for each statement by the number of characters in a given statement. I took participants' median reaction times for non-degraded, minimally degraded, and moderately degraded trials to account for outliers in participants' reaction times. This allowed me to determine whether there were differences in reaction time as a function of degradation level (e.g., test perceptual fluency). I also tested for differences in reaction time as a function of contextual information (e.g., test conceptual fluency). Furthermore, I controlled for individual participants' typing speed by taking a baseline measure of typing speed that I co-varied with this reaction time measure. To account for errors in the covariate baseline typing speed measure, I used the following formula: Time in milliseconds/(number of characters – number of uncorrected errors). I then centered participants' baseline typing speed by subtracting the mean baseline typing speed from each baseline typing speed (Schneider et al., 2015).

I conducted a 3 (degradation level: non-degraded; minimally degraded; moderately degraded) x 3 (contextual information: no context; criminal suspect; written transcripts) mixed ANCOVA with degradation level as the within-subject factor, LexTALE score and baseline typing speed as covariates, and median reaction time for transcribing statements as the dependent variable. Consistent with Hypothesis 2a (Figure 4), there was a main effect of degradation level, $F(1.47, 667.68) = 208.23$, $p < .001$, $\eta_p^2 = .32$ ($BF_{INCL} = 8.08e+58$; see Figure 7). Follow-up post hoc comparisons with a Bonferroni correction revealed that when baseline typing speed and LexTALE score were controlled, median reaction times were significantly shorter for non-degraded audio recordings (*M* = 232.95 ms) than for both minimally degraded audio recordings (*M* = 252.61 ms), $p < .001$, 95% CI [15.25, 24.08] (BF_{INCL} = 4.99e+18) and moderately degraded audio recordings (*M* = 288.09 ms) $p < .001$, 95% CI [47.07, 63.21] (BF_{INCL} = 2.92e+35). Additionally, reaction times were significantly shorter for minimally degraded audio recordings than for moderately degraded audio recordings, $p < .001$, 95% CI [28.74, 42.21] (BF_{INCL} = 3.41e+23). These findings suggest that degrading the recordings reduced perceptual fluency.

There was also a main effect of contextual information, $F(2, 453) = 21.65$, $p < .001$, $\eta_p^2 = .09$ (BF_{INCL} = 5.88+6), which was qualified by a significant interaction between degradation level and contextual information, $F(2.95, 667.68) = 36.97$, $p < .001$, $\eta_p^2 = .14$ (BF_{INCL} = 5.21e+25). This interaction supported Hypothesis 2b (Figure 4). Thus, I conducted follow-up one-way ANCOVAs for each of the degradation levels. When the recordings were non-degraded, there were no differences in median reaction time across the contextual information conditions, $F(2, 453) = 1.61$, $p = .202$, $\eta_p^2 = .01$ (BF_{INCL} = .11). When the recordings were minimally degraded, there was a significant effect of contextual information, $F(2, 453) = 11.65$, $p < .001$, $\eta_p^2 = .05$ (BF_{INCL} = 773.25). Follow-up post hoc comparisons with a Bonferroni correction revealed that participants in the no context condition (*M* = 265.81 ms) had significantly longer reaction times than participants who received written transcripts (*M* = 235.04 ms), $p < .001$, $d = 0.51$, 95% CI [-46.60, -14.95] (BF_{INCL} = 12,987.00). Additionally, participants in the criminal suspect condition (*M* = 256.99 ms) had significantly longer reaction times than participants in the written transcripts condition, $p = .003$, $d = 0.36$, 95% CI [-37.85, -6.04] (BF_{INCL} = 17.23). However, there was no difference between the no context condition and the criminal suspect condition, $p = .569$, $d = 0.15$, 95% CI [-24.97, 7.32] (BF_{INCL} = 0.23).

Finally, when the recordings were moderately degraded, there was once again a significant effect of contextual information, $F(2, 453) = 39.71$, $p < .001$, $\eta_p^2 = .15$ (BF_{INCL} = 4.30e+13). Follow-up post hoc comparisons with a Bonferroni correction revealed that participants in the no context condition had significantly longer reaction times (*M* = 319.19 ms) than participants who received written transcripts (*M* = 238.52 ms), $p < .001$, $d = 0.94$, 95% CI [-104.15, -57.18] (BF_{INCL} = 4.01e+13). Additionally, participants in the criminal suspect condition (*M* = 306.55 ms) had significantly longer reaction times than participants in the written transcripts condition, $p < .001$, $d = 0.79$, 95% CI [-91.63, -44.42] ($BF_{INCL} = 1.15e+11$). However, there was no difference between the no context and criminal suspect conditions, $p = .616$, $d = 0.15$, 95% CI [-36.61, 11.32] (BF_{INCL} = 0.21). Overall, the interaction stemmed from there being differences in reaction time between contextual information conditions for minimally and moderately degraded recordings but not for non-degraded recordings. Thus, there is some evidence that increasing the amount of contextual information available increased conceptual fluency.

Figure 7 **Participants' Reaction Time as a Function of Degradation Level and Contextual Information in Study 2**



Note. Error bars are 95% Confidence Intervals.

4.4.3. Fluency as a Mechanism for Confirmation Bias

The reaction time analyses suggest differences across degradation level and contextual information conditions consistent with what I hypothesized. Therefore, I conducted a multilevel moderated mediation analysis to test whether reaction time mediates the relationship between the degradation level and contextual information interaction and participants' tendency to make incriminating misinterpretations. Moderated mediation analyses test whether the effects of an independent variable X on an outcome variable Y via a mediator variable M vary across levels of a moderator variable W (Edwards & Lambert, 2007; Hayes, 2009; 2015; Muller et al., 2005; Preacher et al., 2007). This is calculated using an *index of moderated mediation,* which provides a Monte Carlo confidence interval for the test of whether W moderates the indirect effect of X on Y (Hayes & Rockwood, 2020). If there is a relationship between W and the size of the indirect effect of X on Y through M, the Monte Carlo confidence interval will not contain 0.

Throughout this section, I will interchangeably use the terms "participants' interpretations" and "confirmation bias." I acknowledge that there is only evidence of confirmation bias if participants with contextual information make more incriminating misinterpretations than participants without contextual information. However, I established in previous analyses that participants in the criminal suspect and written transcripts conditions do, in fact, make more incriminating misinterpretations than participants in the no context condition. In this set of analyses, I aimed to demonstrate the process through which this occurs (i.e., through decreased perceptual fluency accompanied by increased conceptual fluency resulting in quicker overall reaction times). While the outcome variable I entered in my model measures whether participants' interpretations were incriminating, I will sometimes simply refer to this as "confirmation bias" given that (a) I have already established that confirmation bias is operating and (b) I want to know whether increased fluency (as measured by decreased reaction time) explains increases in participants' tendency to make incriminating misinterpretations when they have incriminating contextual information compared to when they do not (i.e., confirmation bias).

I used the MLMED Macro in SPSS to conduct this multilevel moderated mediation analysis (Hayes & Rockwood, 2020; Hu et al., 2020). I conduced a 1-1-1 moderated mediation with degradation level as my independent variable, reaction time as my mediator, participants' interpretations (dummy coded as 0 = non-incriminating; 1 = incriminating) as my outcome variable, contextual information as my moderator, and baseline typing speed as a Level 2 covariate (see Figure 8). The index of moderated mediation suggested that the indirect effect of degradation level on participants' interpretations made through reaction time is positively moderated by contextual information, MCLL = .0004, MCUL = .0014.[6]

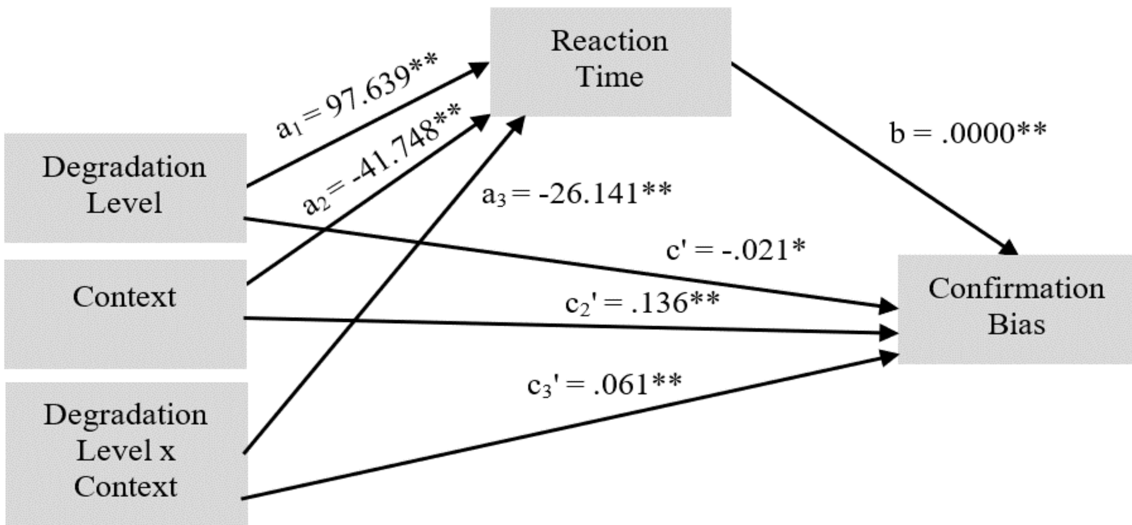
Next, I examined the effects of degradation level and contextual information on participants' reaction time. Both degradation level (B = 97.639) and contextual information (B = -41.748), as well as the interaction between degradation level and contextual information (B = -26.141) significantly predicted reaction time, ps < .0001. Then, I examined the effects of degradation level, contextual information, and reaction time on participants' interpretations. Contextual information (B = .136), the interaction between degradation level and contextual information (B = .061) and reaction time (B = .0000) all significantly predicted participants' interpretations, ps <.0001.

Finally, I examined the direct and indirect effects of degradation level on participants' interpretations. The direct effect of degradation level on participants' interpretations was significant (B = -.021; t(18345.76) = 2.21, p = .027, 95% CI [-.039, -.002]) as was the indirect effect of degradation level on participants' interpretations through reaction time (B = -.003; z = 3.89, p < .001, 95% MCCI[7] [-.005, -.002]). Because the moderated mediation analysis revealed that the indirect effect of degradation level on proportion of incriminating misinterpretations through reaction time varied based on the type of contextual information participants received, I followed up this analysis with simple mediation models for each degradation level.

[6] MCLL = Monte Carlo Lower Limit; MCUL = Monte Carlo Upper Limit

[7] MCCI = Monte Carlo Confidence Interval

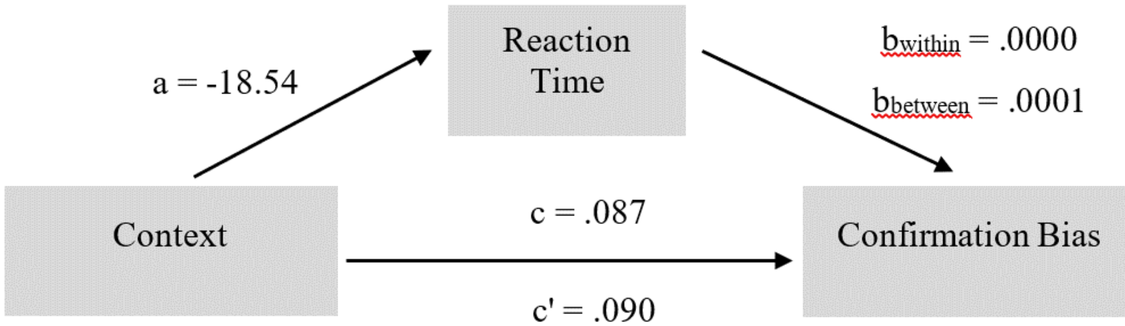
Figure 8 Results of the Multilevel Moderated Mediation Analyses in Study 2



Note. Unstandardized beta coefficients are reported for the a, b, and c’ pathways.
**p* < .05, ***p* < .0001

First, I examined whether, for non-degraded recordings, reaction time mediated the differences in confirmation bias between the contextual information conditions. I conducted a 2-1-1 mediation analysis with contextual information as my independent variable, reaction time as my mediator, participants’ interpretations (dummy coded as 0 = non-incriminating and 1 = incriminating) as my dependent variable, and baseline typing speed as my covariate (see Figure 9). Contextual information significantly predicted reaction time, $t(452.58) = 2.89$, $p = .004$, 95% CI [-31.14, -5.94]. Reaction time did not significantly predict participants’ interpretations at the trial level, $t(5878.31) = .02$, $p = .981$, 95% CI [.0000, .0000]. However, reaction time did significantly predict participants’ interpretations at the participant level, $t(453.79) = 2.12$, $p = .034$, 95% CI [.0000, .0002]. Finally, the indirect effect of contextual information on participants’ interpretations mediated by reaction time was not significant, $z = 1.65$, $p = .099$, 95% MCCI [-.0054, -.0001]. Thus, reaction time does not appear to mediate the effect of contextual information on confirmation bias for non-degraded recordings.

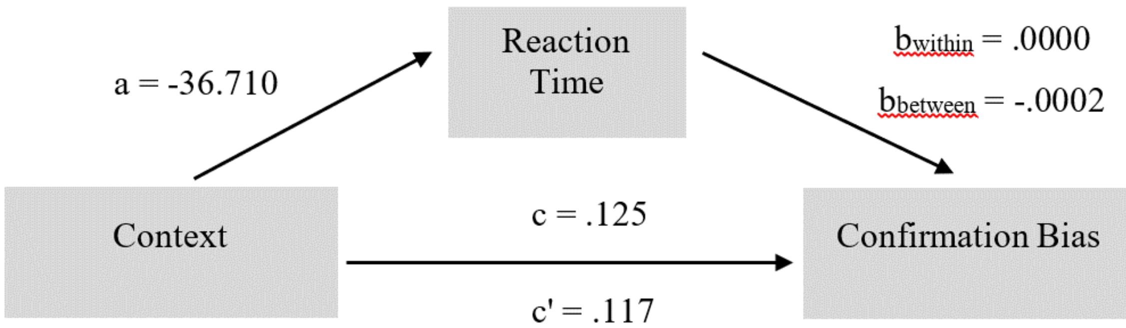
Figure 9　　Mediation Model for Non-Degraded Recordings in Study 2



Note. Unstandardized beta coefficients are reported for the a, b, c, and c' pathways where c is the total effect of context on confirmation bias when reaction time is not controlled for and c' is the direct effect of context on confirmation bias when reaction time is controlled for.

Next, I examined whether, for minimally degraded recordings, reaction time mediated the differences in confirmation bias between the contextual information conditions. Once again, I conducted a 2-1-1 mediation analysis with contextual information as my independent variable, reaction time as my mediator, participants' interpretations (dummy coded as 0 = non-incriminating and 1 = incriminating) as my dependent variable, and baseline typing speed as my covariate (see Figure 10). Contextual information significantly predicted reaction time, $t(436.76) = 4.68$, $p < .0001$, 95% CI [-52.13, -21.29]. Additionally, reaction time significantly predicted participants' interpretations at both the trial level, $t(6073.75) = 2.52$, $p = .012$, 95% CI [-.0001, .0000] and the participant level, $t(485.01) = 3.61$, $p < .001$, 95% CI [-.0003, -.0001]. Finally, the indirect effect of context on participants' interpretations mediated by reaction time was significant, $z = 2.82$, $p = .005$, 95% MCCI [.003, .014]. Thus, reaction time appears to partially mediate the effect of contextual information on confirmation bias for minimally degraded recordings.

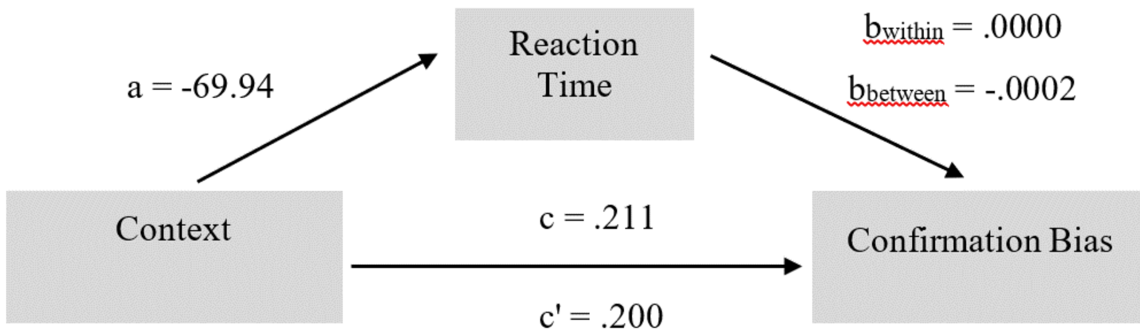
Figure 10 Mediation Model for Minimally Degraded Recordings in Study 2



Note. Unstandardized beta coefficients are reported for the a, b, c, and c' pathways where c is the total effect of context on confirmation bias when reaction time is not controlled for and c' is the direct effect of context on confirmation bias when reaction time is controlled for.

Finally, I examined whether, for moderately degraded recordings, reaction time mediated the differences in confirmation bias between the contextual information conditions. I conducted a 2-1-1 mediation analysis with contextual information as my independent variable, reaction time as my mediator, participants' interpretations (dummy coded as 0 = non-incriminating and 1 = incriminating) as my dependent variable, and baseline typing speed as my covariate (see Figure 11). Contextual information significantly predicted reaction time $t(447.96) = 5.91$, $p < .0001$, 95% CI [-93.19, -46.68]. Additionally, reaction time significantly predicted participants' interpretations at both the trial level, $t(5783.75) = 2.96$, $p = .003$, 95% CI [-.0001, .0000] and at the participant level, $t(482.23) = 4.35$, $p < .0001$, 95% CI [-.0002, -.0001]. Finally, the indirect effect of context on participants' interpretations mediated by reaction time was significant, $z = 3.47$, $p = .001$, 95% MCCI [.006, .019]. Thus, reaction time appears to partially mediate the effect of contextual information on confirmation bias for moderately degraded recordings.

Figure 11 Mediation Model for Moderately Degraded Recordings in Study 2



Note. Unstandardized beta coefficients are reported for the a, b, c, and c' pathways where c is the total effect of context on confirmation bias when reaction time is not controlled for and c' is the direct effect of context on confirmation bias when reaction time is controlled for.

4.5. Discussion

The purpose of Study 2 was to determine whether fluency misattribution is a mechanism for confirmation bias in cases involving degraded audio recordings. I manipulated perceptual fluency across trials by presenting audio recordings that varied in degradation level. Additionally, I manipulated conceptual fluency by varying the amount of contextual information participants learned about the recordings. Overall, and consistent with previous research, I found evidence of confirmation bias such that participants with contextual information made more incriminating misinterpretations than participants without contextual information. Also, the more contextual information participants had, the more confirmation bias they demonstrated. Specifically, participants in the written transcripts condition made more incriminating misinterpretations than those in the criminal suspect condition, who in turn made more incriminating misinterpretations than those in the no context condition.

I also found differences in reaction time as a function of degradation level and context. Specifically, as recordings became more degraded, they became perceptually less fluent as evidenced by longer reaction times. Thus, there is evidence that manipulating the degradation of the recordings affected perceptual fluency. Furthermore, I found that for degraded recordings, participants with written transcripts interpreted the recordings more fluently, as evidenced by quicker reaction times. However, the criminal suspect condition did not have significantly quicker reaction times than the no context condition. Thus, there is mixed support for the hypothesis that manipulating the amount of contextual information participants received affected conceptual fluency.

Decreasing perceptual fluency and increasing conceptual fluency increased confirmation bias to a certain extent. Specifically, participants with contextual information exhibited more confirmation bias (i.e., made more incriminating misinterpretations than the no context condition) when the recordings were degraded compared to when they were non-degraded. Additionally, participants in the written transcripts condition made the most incriminating misinterpretations for moderately degraded recordings. Alternatively, participants in the criminal suspect condition made the most incriminating misinterpretations for minimally degraded recordings. This suggests that perceptual ambiguity increases the likelihood that participants will rely on contextual information when evaluating the audio recordings. However, when contextual information is limited and the stimuli become too perceptually disfluent, participants may be less likely to exhibit confirmation bias.

It is also important to note that there are both objective and subjective measures of fluency (Schwarz, 2004). Objective fluency, as the name implies, refers to the actual ease with which people process information. Objective fluency can be inferred from measures such as processing speed (i.e., reaction time) and accuracy (Reber et al., 2004). Subjective fluency, conversely, refers to people's perceptions of the ease with which they were able to process stimuli (Forster et al., 2013). Subjective fluency can be inferred from measures such as people's impressions of how effortful it was to generate judgments, how quickly they were able to process information, or their own accuracy in their judgments. Both objective and subjective fluency can influence decision-making (Oppenheimer, 2008; Reber et al., 2004; Schwarz, 2004).

Thus, it is possible that learning the criminal suspect context increased subjective feelings of fluency that weren't captured in participants' reaction time (i.e., objective fluency), but that nonetheless affected their tendency to make incriminating misinterpretations. That is, participants may have found it subjectively easier to generate interpretations that were consistent with the criminal suspect context (compared to non-incriminating alternatives), though this decision was not reached objectively quicker. Consequently, they attributed the relative ease with which they generated incriminating misinterpretations and the overall coherence of these incriminating statements as evidence that these interpretations were correct.

Indeed, research has explored how objective manipulations of fluency affect subjective experiences of fluency. For example, in one study, researchers subliminally primed participants with masked images that were either perceptually congruent or incongruent with the target images participants saw. In two counterbalanced blocks, participants independently rated (1) how much they liked the target images and (2) how easily they perceived the target images (which were all somewhat perceptually disfluent because Gaussian noise had been added) on a scale from 1 (not at all) to 7 (very much). Compared to incongruently-primed images, participants rated congruently-primed images as easier to perceive and also provided higher liking ratings. Interestingly, the researchers also found that participants' subjective ratings of how easily they perceived the images influenced their liking ratings to a greater extent than the manipulations to perceptual fluency (Forster et al., 2013). Perhaps self-reported feelings of fluency measured on an ease of processing scale (e.g., from "it was easy to process this recording" to "it was difficult to process this recording") or an ease of generation scale (e.g., from "it was easy to generate this interpretation" to "it was difficult to generate this interpretation") would similarly reveal differences across context conditions in the current paradigm.

Other research has shown that there are different perceptual stages that contribute to the overall experience of subjective fluency, namely (1) detection and (2) identification. In one study, Reber et al. (2004) manipulated the perceptual fluency of words by varying the figure-ground contrast and font. Participants either had to (1) detect whether a word was presented between two masks; (2) identify the word presented on screen; or (3) rate the readability of the words on a 9-point scale (i.e., provide subjective fluency ratings). Participants' subjective fluency ratings revealed that the high contrast words were easier to perceive than the low contrast words and words presented in Times (i.e., more regular typeface) were easier to perceive than words presented in Tremor (i.e., less regular typeface). Thus, manipulating the perceptual fluency of words influenced participants' subjective experiences of fluency. However, the researchers found that neither detection performance nor identification performance alone reflected participants' subjective experience of fluency. While contrast had a large impact on detection accuracy and speed, font had only a marginal influence. Alternatively, font, but not contrast, had a significant impact on identification accuracy and speed. This suggests that detection and identification both contribute to the subjective experience of fluency (Reber et al., 2004).

Perhaps in the current study there were similar stages of processing that jointly impacted subjective fluency. Thus, despite the criminal suspect context not significantly contributing to differences in identification reaction time, there may be other stages of processing that occurred that would jointly impact overall experiences of fluency. While the current study was an important first step in exploring the contribution of fluency in confirmation bias for degraded audio recordings, perhaps other measures of processing fluency will reveal more minute differences across conditions.

In any case, according to the partial mediation analyses, fluency, as measured by reaction time, appears to be a mechanism for confirmation bias for minimally and moderately degraded recordings. However, as I previously noted, reaction time does not fully mediate this effect and does not explain differences in confirmation bias across the context conditions for non-degraded recordings. Perhaps other measures of fluency such as subjective fluency ratings partially mediate the interactive effects of degradation level and context on confirmation bias for degraded audio recordings as well. Indeed, one study found that subjective processing fluency measured by subjective ratings of the ease with which participants comprehended a scenario mediated confirmation bias (Zaleskiewicz & Gasiorowska, 2021). However, it is likely that other mechanisms underlie confirmation bias as well. For example, belief persistence or selectively attending to information that supports one's pre-existing beliefs may underlie confirmation bias (Prat-Ortega & de la Rocha, 2018; Ross et al., 1975; Talluri et al., 2018). Future research should continue to explore the mechanisms that underlie confirmation bias generally, and for degraded audio recordings specifically.

4.6. Limitations

There were several limitations of this work. First, I collected the reaction time data online. While I originally intended to conduct this study in-person, I was forced to transition this study online due to the COVID-19 pandemic. There has traditionally been concern over the reliability and validity of reaction time measures collected online. However, research comparing reaction time studies conducted in the lab versus online reveals that the web-based data patterns replicate the lab-based findings (Hilbig, 2016; Kim et al., 2019). Regardless, future research should measure reaction time in a lab-based study to replicate the current findings.

Furthermore, I did not separate perceptual from conceptual fluency. Again, it would be impossible to completely distinguish perceptual from conceptual fluency because even when the recordings are degraded, individuals experience some conceptual fluency from the words they are able to interpret in the recorded statements. Additionally, if I were to completely isolate perceptual fluency, I would have to use stimuli such as non-words which would be irrelevant to this study. While this study was an important first step in exploring how fluency misattribution contributes to confirmation bias, future research should investigate how each form of fluency individually contributes to confirmation bias. Furthermore, future research should explore methods for mitigating this bias. Thus, Study 3 explores whether jury instructions can effectively reduce confirmation bias for degraded audio recordings.

Chapter 5.

Study 3

Confirmation bias is pervasive in decision-making and oftentimes causes people to draw flawed conclusions. This bias can have particularly negative implications within the legal system, where biased evaluations of evidence, such as degraded audio evidence, can contribute to erroneous legal decisions. Given this, it is not only essential to understand the circumstances in which confirmation bias occurs and the mechanisms behind this bias, but also to explore methods for mitigating this bias. Overall, few studies have investigated how to mitigate confirmation bias.

Those studies that have investigated how to mitigate confirmation bias have yielded mixed findings. Some research has found that priming participants with a counterfactual mindset can lead them to consider disconfirming information more thoroughly (Slovic & Fischhoff, 1977). Specifically, before completing a target task, participants considered how a hypothetical scenario might have turned out differently. This hypothetical scenario was unrelated to their target task, and instead, was meant to provide participants with the opportunity to consider counterfactuals. The researchers found that participants who were primed with a counterfactual mindset were more likely to consider information that contradicted their own initial beliefs when completing the target task (Kray & Galinsky, 2003; Galinsky & Moskowitz, 2000). Subsequently, these participants showed less confirmation bias.

Other research has demonstrated that disfluency can mitigate confirmation bias (Hernandez & Preston, 2013). After learning positive or negative information about an accused person, participants read information about an alleged crime in which the guilt of the accused was ambiguous. The researchers found that when participants read this information in a fluent font, they exhibited confirmation bias. Alternatively, participants who read this information in a disfluent font did not exhibit confirmation bias. However, if participants read the information in a disfluent font while under cognitive load (i.e., time constraint or memorization task), they exhibited confirmation bias. The authors concluded that when participants had cognitive resources available, they processed the ambiguous information more deeply when it was presented in a disfluent font (Hernandez & Preston, 2013).

Given the life-changing implications of erroneous decision-making in the legal realm, it is particularly important to explore methods for mitigating confirmation bias in this context. Some research has examined methods for reducing confirmation bias in the investigative stage. For example, O'Brien (2009; see also O'Brien & Ellsworth, 2006) found that when participants who evaluated a mock police file considered why their initial hypothesis about the target suspect might be incorrect, they exhibited less confirmation bias. However, when participants attempted to generate additional hypotheses about alternative suspects, there was no reduction in confirmation bias. This may have occurred because it was relatively challenging to generate alternative hypotheses. Thus, as research on processing fluency would suggest, participants may have misattributed the relative difficulty they had in generating alternative hypotheses to mean that their initial hypothesis must be correct (Jacoby & Whitehouse, 1989; Schwarz et al., 1991; Winkielman et al., 1998, 2003).

A similar line of research explored whether presenting an alternative story by the defense can alter charging decisions (Schmittat et al., 2021). A prosecutor's decision to charge a suspect is largely based on information presented in the police report, which may be biased due to tunnel vision (i.e., confirmation bias) that occurred during the investigative process. The prosecutor's decision may be further impacted by irrelevant contextual information, such as prior conviction evidence. Given the consequences of charging individuals with crimes they did not commit, the researchers explored whether presenting participants with a written statement from the defense that suggests an alternative story reduced the likelihood of proceeding with charges. Across three studies, they found that including an alternative story from the defense in the investigative materials reduced the likelihood of the accused being charged (Schmittat et al., 2021).

Researchers have also recommended procedures that reduce bias in forensic science. For example, some researchers have suggested that forensic examiners should adhere to a “linear sequential unmasking” strategy whereby relevant information is released to the examiner only when it is absolutely necessary (Dror, 2016; Dror et al., 2015; Krane et al., 2008). Experts should observe the evidence and document their initial observations in isolation from biasing information such as police’s beliefs about the target suspect’s identity. While there is no empirical research on the effectiveness of this de-biasing procedure, the potential for limiting biased evaluations of forensic evidence is promising. Despite these limited examinations into mitigating confirmation bias in forensic contexts, to my knowledge, no research has explored methods for mitigating confirmation bias in cases involving degraded audio evidence. Thus, the goal of Study 3 was to test the effectiveness of the Canadian Model Jury Instructions for audio recordings and written transcripts in de-biasing individuals’ evaluations of degraded audio evidence.

5.1. Canadian Model Jury Instructions for Audio Recordings and Written Transcripts

Currently, there are Canadian Model Jury Instructions that may be administered in cases involving audio evidence and written transcripts (Canadian National Judicial Institute Model Jury Instructions, 2011). One section of these instructions states, “If what you read on the transcript differs from what you hear on the tape, you are to go by what you hear for yourself, and not what you read in the transcript.” However, if the transcripts conflict with the recording due to transcription errors, individuals may be unable to identify the conflict and disregard the transcripts. Additionally, individuals may have a particularly difficult time reconciling differences between written transcripts and audio recordings when the recordings are degraded. This makes it especially important to test the tenets of these instructions in cases involving degraded audio evidence. Therefore, the first goal of Study 3 was to test whether mock jurors can identify errors in transcripts, and if so, if they can disregard the transcripts when evaluating the degraded audio evidence.

According to the Canadian Model Jury Instructions, if the transcripts are given to jurors as aids but are not filed as exhibits, the judge is also supposed to state the following: “Please follow carefully as the tape is played in the courtroom. The tape will be available to you in the jury room, but not the transcript. We will collect the transcripts after the tape has been played." However, it remains unclear how an earlier exposure to transcripts affects subsequent evaluations of the degraded audio recordings without the transcripts. Yet, trials often last several days or weeks, and therefore, it is likely that jurors would be asked to deliberate on audio evidence that they previously evaluated with transcripts in court. Therefore, the second goal of Study 3 was to test whether prior exposure to transcripts influences later interpretations of the audio evidence without transcripts by including a delay manipulation.

5.2. Methodology

I pre-registered this study on Open Science Framework prior to collecting data (https://doi.org/10.17605/OSF.IO/7MS9F).

5.2.1. Participants

I conducted an a priori power analysis in G*Power 3.1 to determine the sample size required for this study (Faul et al., 2009). Previous research has demonstrated a medium effect of incriminating contextual information on participants' incriminating misinterpretations of recordings (e.g., $d = 0.5 – 0.76$, Study 1; OR = 4.56, Lange et al., 2011, Experiment 1). However, given that previous research has not tested the effects of instructions or delay on confirmation bias for degraded audio recordings, I assumed a more conservative small-to-medium effect. Thus, with $f = .20$, $\alpha = .05$, Power = .95, numerator df = 1 and number of groups = 4, I required a sample of $N = 327$. However, given that I also planned to include a control condition of participants who did not receive written transcripts to ensure that the transcripts were in fact eliciting confirmation bias, I aimed to recruit an additional 82 participants (327/4 = ~82 participants per condition) for my fifth condition. Thus, I required a total sample of $N = 409$.

I recruited 607 participants through the SFU undergraduate research pool. Participants received 3% course credit in exchange for their participation. A total of 79 participants (13.01%) did not complete the second part of the study. Of the remaining 528 participants who completed Parts 1 and 2 of the study, 119 participants were excluded from analyses for the following reasons: (1) they withdrew early (4.17%); (2) they failed the manipulation check (14.77%); or (3) they had missing data for more than 25% of trials (3.60%). Thus, a total of 409 participants were included in the final analyses. The average age of participants was 19.57 years (*SD* = 2.95 years; *Range* = 17 – 44 years) with 74.33% of participants identifying as female and 25.67% identifying as male. Additionally, 39.12% of participants were White, 46.45% were Asian (including 13.45% East Asian, 20.05% South Asian, and 5.62% Southeast Asian), 0.98% were Black, 0.24% were Hispanic, 0.49% were Latinx, 2.93% were Middle Eastern, 0.49% were Indigenous, 5.62% were Mixed Race, and 3.67% did not report their ethnicity.

5.2.2. Design

This study was a 2 (instructions: present; absent) x 2 (delay: no delay; one-week delay) between-subject design with a control condition of participants who listened to the recordings without seeing the incriminating written transcripts (i.e., the "no transcripts condition"). This control condition served as a manipulation check to verify that participants with written transcripts were more likely to make incriminating misinterpretations than participants without written transcripts. The primary dependent variable was the proportion of incriminating misinterpretations participants made. Once again, I operationalized confirmation bias as the proportion of incriminating misinterpretations that participants in the written transcripts condition made over and above the no transcripts condition. I also measured participants' perceptions of the likelihood that the speakers were guilty on a scale from 1 (Not at all likely) to 10 (Extremely likely).

5.2.3. Procedure

Participants completed the study online through Qualtrics. They first completed a Captcha Verification followed by the same hearing test described previously. Then, participants in the no transcripts condition (i.e., the control condition) simply learned that they would be listening to a series of audio recordings. No other contextual information about the recordings was presented. All remaining participants were told the following: "Imagine you are a juror in a series of criminal trials. The primary evidence against each of the accused is a series of audio recordings that come from wiretapped conversations with the accused. You will see written transcripts of each statement that are presented at the same time as the statement. Please listen to the statements and read the transcriptions carefully." Additionally, participants in the instructions present condition read the Canadian Model Jury Instructions for audio recordings and written transcripts (see Appendix G). The remaining participants did not receive these instructions.

Next, all participants heard 48 degraded audio recordings. These recordings were presented across two randomly determined fixed orders. Like Study 1, degradation levels ranged between 600 Hz and 1,600 Hz based on the degradation levels for which the proportion of incriminating misinterpretations participants made in the no context condition was less than half of what participants made in the criminal suspect condition in the pilot studies (see items with asterisks in Appendices A and D). Participants in the no transcripts condition listened to the recordings with no additional information, while all remaining participants saw written transcripts containing incriminating errors as the statements played. To ensure that participants were actively listening to the recordings, five attention check questions appeared throughout the recordings which contained simple math questions (e.g., "What is 5 + 6?"). After listening to all 48 recordings, participants in the no delay condition immediately listened to the recordings again without any transcripts and attempted to transcribe each statement. After transcribing each statement, participants also indicated the likelihood that the accused (i.e., the speaker) was guilty on a scale from 1 (Not at all likely) to 10 (Extremely likely). Finally, participants answered a manipulation check question, completed the LexTALE measure of English language fluency, and answered several demographic questions.

Participants in the one-week delay condition finished the first part of the study after listening to the 48 degraded statements and answering demographic questions. Then, after approximately one week (*Range* = 7-9 days), they received the link to the second part of the study. They learned that they would be hearing the same audio recordings they heard in Part 1 of the study without transcripts and that they should transcribe what they heard to the best of their ability. Additionally, participants in the instructions present condition were reminded of the Canadian Model Jury Instructions they received. Thus, if what they heard in the statements differed from what they read in the transcripts, they were to go by what they heard rather than what they read. Then, they listened to and attempted to transcribe the 48 degraded audio statements. After transcribing each statement, they also indicated the likelihood that the speaker of each statement was guilty on a scale from 1-10. Finally, they answered a manipulation check question and completed the LexTALE measure of English language fluency.

5.2.4. Coding Participants' Transcriptions

Three condition-blind coders (including myself) independently coded participants' transcriptions into the following categories as described above: (a) Invalid responses (i.e., blank responses, don't know responses, and transcriptions with less than 3 words); (b) accurate interpretations (i.e., transcriptions in which participants correctly identified every word in a given statement); (c) non-incriminating errors (i.e., transcriptions that contained at least one error, but were not incriminating); and (d) incriminating misinterpretations (i.e., transcriptions which appeared to implicate the accused). I instructed coders how to code the responses. Then, for the purposes of training, we each coded the same 10% of participants' responses. We compared our codes, discussing where we disagreed and settling on agreed-upon codes. We did not track disagreements on this set of codes as the purpose of this training phase was to ensure that all coders understood how to consistently assign codes to participants' responses. After this training phase, all three coders independently coded approximately 30% (29.3%) of the same participant responses and then worked together to compare our codes for the purposes of inter-rater reliability. We noted and discussed disagreements to determine an agreed-upon code. Inter-rater reliability was high (Fleiss' kappa = .96). Finally, each coder individually coded approximately one-third of the remaining responses.

5.3. Hypotheses

5.3.1. Incriminating Misinterpretations

Hypothesis 1/Manipulation Check

I expected that participants who received written transcripts would exhibit confirmation bias by making significantly more incriminating misinterpretations than participants who did not receive written transcripts.

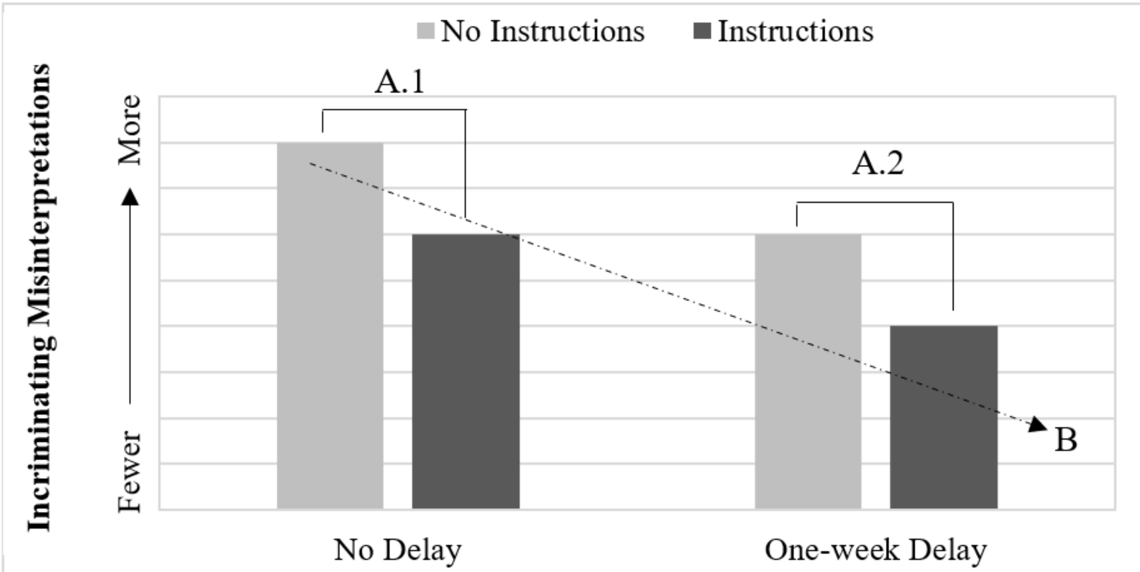
Hypothesis 2a

I expected a main effect of instructions on the degree to which participants made incriminating misinterpretations (see Figure 12 for hypothetical data pattern). Specifically, I expected that if the Canadian Model Jury Instructions are effective in reducing the impact of incriminating written transcripts on participants' evaluations of innocuous degraded audio recordings, participants who received instructions would make fewer incriminating misinterpretations than participants who did not receive these instructions (see A.1 and A.2 in Figure 12).

Hypothesis 2b

Second, I expected a main effect of delay on the degree to which participants made incriminating misinterpretations. Participants who received incriminating transcripts should make fewer incriminating misinterpretations in the one-week delay condition than participants in the no delay condition across instructions conditions (see B in Figure 12). I expected to observe this data pattern because introducing a delay should make participants less likely to remember the incriminating transcripts they previously read. Support for this prediction comes from research on priming which shows that primed constructs are less likely to impact judgments as the delay between priming and stimulus presentation increases, though the persistence of priming effects across delays may vary based on the type of stimuli and task (Cave, 1997; DeCoster & Claypool, 2004; Higgins et al., 1985; Weingarten et al., 2016).

Figure 12 **Hypothetical Data Pattern for the Proportion of Incriminating Misinterpretations that Participants will Make as a Function of Instructions and Delay in Study 3**



5.3.2. Likelihood of Guilt

Overall, I expected that the data pattern for participants' likelihood of guilt ratings would be identical to the data pattern I predicted for participants' incriminating misinterpretations.

Hypothesis 3/Manipulation Check

I expected that participants who received written transcripts would assign higher likelihood of guilt ratings to the speakers in the audio recordings than participants who did not receive written transcripts.

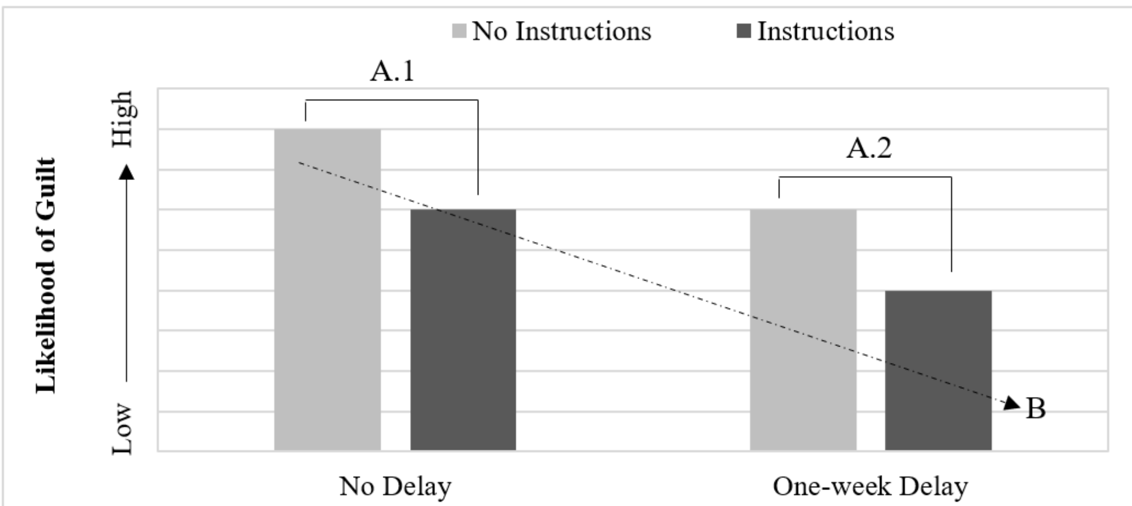
Hypothesis 4a

I expected a main effect of instructions on participants' likelihood of guilt ratings (see Figure 13 for hypothetical data pattern). Specifically, if the Canadian Model Jury Instructions are effective in mitigating confirmation bias in the presence of incriminating written transcripts, I expected that participants who received the instructions would give lower likelihood of guilt ratings to the accused speakers than participants who did not receive these instructions (see A.1 and A.2 in Figure 13).

Hypothesis 4b

Finally, I expected a main effect of delay on participants' likelihood of guilt ratings (see Figure 13 for hypothetical data pattern). I expected that participants would assign lower likelihood of guilt ratings to the accused in the one-week delay condition than in the no delay condition, across instructions conditions (see B in Figure 13).

Figure 13 Hypothetical Data Pattern for Participants' Likelihood of Guilt Ratings as a Function of Instructions and Delay in Study 3



5.4. Results

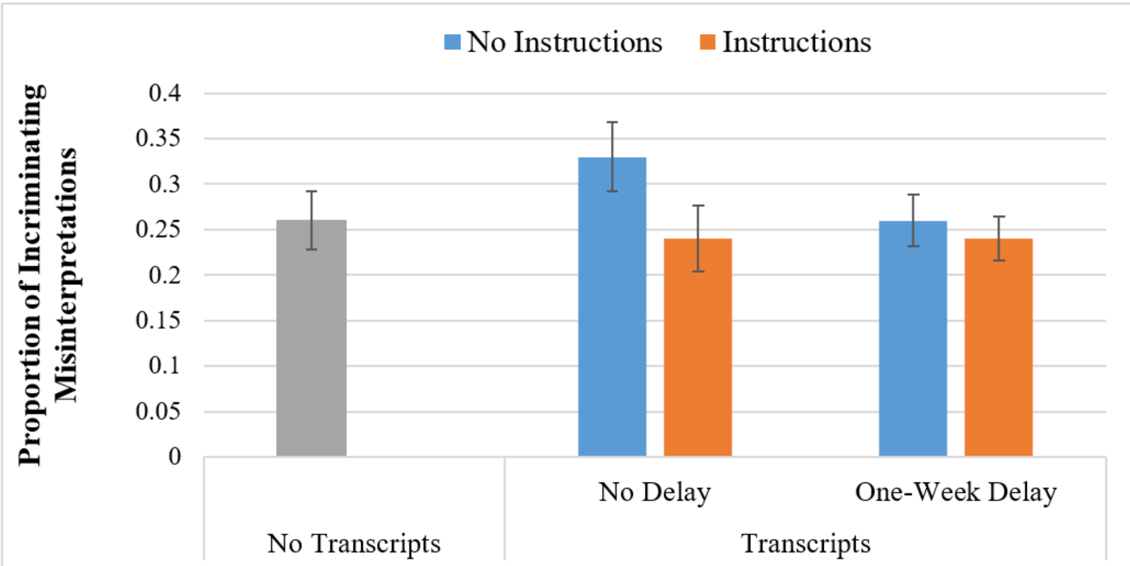
5.4.1. Incriminating Misinterpretations

First, to ensure that written transcripts were eliciting confirmation bias, I conducted an Analysis of Covariance (ANCOVA) with transcripts (no transcripts vs. written transcripts) as the independent variable, proportion of incriminating misinterpretations as the dependent variable, and LexTALE score as the covariate. In the transcripts condition, I only included participants who did not experience a delay or receive instructions given that I hypothesized that both delay and instructions would reduce participants' tendency to make incriminating misinterpretations. Consistent with Hypothesis 1, this analysis revealed that participants in the no transcripts condition (*M* =

.26) made fewer incriminating misinterpretations than participants who received incriminating transcripts (*M* = .33), $F(1, 170) = 8.43$, $p = .004$, $\eta_p^2 = .05$, 95% CI [-.12, -.02] ($BF_{INCL} = 6.06$). Thus, incriminating written transcripts did, in fact, produce the expected confirmation bias effect.

Next, to determine whether the presence of instructions or a delay mitigated this effect, I conducted a 2 (instructions: present; absent) x 2 (delay: no delay; one-week delay) ANCOVA with proportion of incriminating misinterpretations as the dependent variable and LexTALE score as the covariate. Consistent with Hypotheses 2a and 2b (see Figure 12), there was a main effect of instructions, $F(1, 309) = 9.61$, $p = .002$, $\eta_p^2 = .03$ ($BF_{INCL} = 21.73$) as well as a main effect of delay, $F(1, 309) = 4.71$, $p = .031$, $\eta_p^2 = .02$ ($BF_{INCL} = 0.69$). However, these main effects were qualified by an interaction between instructions and delay, $F(1, 309) = 5.18$, $p = .024$, $\eta_p^2 = .02$ ($BF_{INCL} = 1.99$), which I did not hypothesize (see Figure 14). Follow up ANCOVAs revealed that when there was no delay, participants in the instructions absent condition (*M* = .33) made significantly more incriminating misinterpretations than participants in the instructions present condition (*M* = .24), $F(1, 161) = 11.18$, $p = .001$, $\eta_p^2 = .07$, 95% CI [.04, .14] ($BF_{INCL} = 27.62$). However, when there was a delay, there was no difference in the proportion of incriminating misinterpretations that participants made between the instructions absent (*M* = .26) and instructions present (*M* = .24) conditions, $F(1, 147) = .54$, $p = .463$, $\eta_p^2 = .004$, 95% CI [-.02, .05] ($BF_{INCL} = .23$).

Figure 14 Proportion of Incriminating Misinterpretations that Participants Made as a Function of Instructions and Delay in Study 3

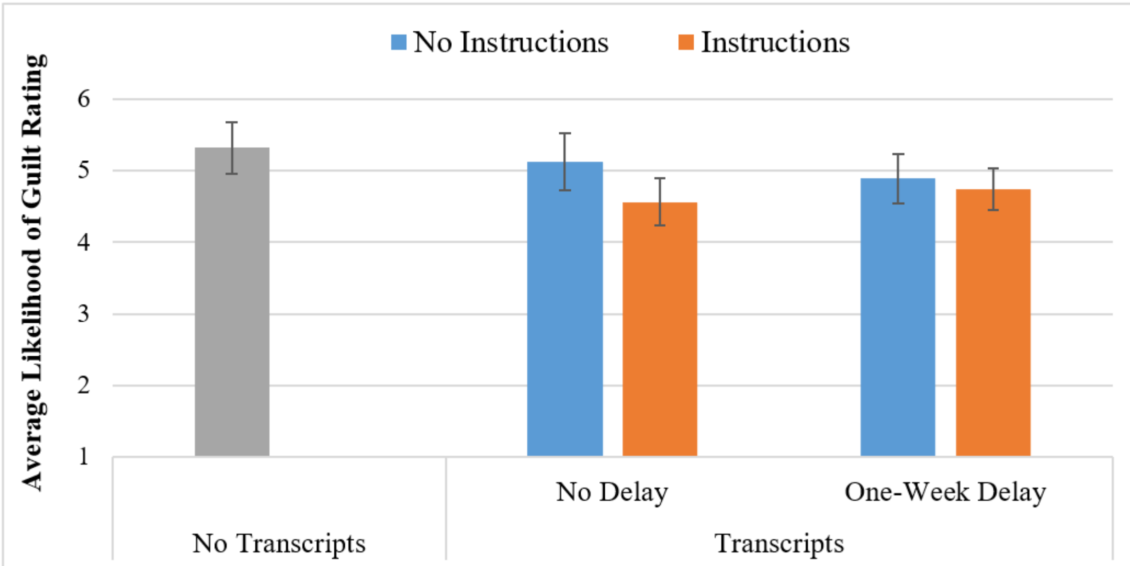


Note. Error bars are 95% Confidence Intervals.

5.4.2. Likelihood of Guilt

First, to ensure that written transcripts were eliciting higher likelihood of guilt ratings, I conducted an ANCOVA with transcripts (no transcripts vs. written transcripts) as the independent variable, average likelihood of guilt ratings as the dependent variable, and LexTALE score as the covariate. Once again, I only included participants who did not experience a delay or receive instructions in the written transcripts condition given that I hypothesized that both delay and instructions would reduce confirmation bias. Additionally, I only included likelihood of guilt ratings on trials for which participants gave a valid response. In contradiction to Hypothesis 3, this analysis revealed no significant differences between the transcript absent (M = 5.32) and transcript present (M = 5.12) conditions, $F(1, 170) = 0.56$, $p = .456$, $\eta_p^2 = .003$, 95% CI [-0.33, 0.74] (BF_{INCL} = 0.22; See Figure 15). Given that written transcripts did not lead to significantly higher likelihood of guilt ratings, I did not conduct any additional analyses to determine whether instructions or delay would reduce perceptions of guilt.

Figure 15 **Average Likelihood of Guilt Ratings as a Function of Instructions and Delay in Study 3**



Note. Error bars are 95% Confidence Intervals.

5.5. Discussion

The purpose of Study 3 was to determine whether (a) transcripts containing incriminating errors elicited more incriminating misinterpretations and higher likelihood of guilt ratings (i.e., confirmation bias); (b) the Canadian Model Jury Instructions for audio recordings and written transcripts could mitigate this confirmation bias effect; and (c) a delay between participants' exposure to the incriminating transcripts and when they transcribed the statements decreased participants' tendency to make incriminating misinterpretations. With respect to (a), I once again found that incriminating contextual information in the form of written transcripts biased participants' interpretations of degraded audio recordings. Compared to participants with no contextual information about the recordings, those who received written transcripts containing incriminating errors were more likely to perceive the statements as incriminating.

With respect to (b), I found that the Canadian Model Jury Instructions for audio recordings and written transcripts effectively mitigated this confirmation bias effect when there was no delay between participants' exposure to the incriminating transcripts and when they transcribed their interpretations of the audio recordings. Participants who received these jury instructions made significantly fewer incriminating misinterpretations than participants who did not receive these instructions. This is an important and encouraging finding given that administering instructions is one of the simplest and most cost-effective ways to address biases in legal cases.

In general, previous research has demonstrated that jury instructions have minimal impact on mock jurors' judgments (Alvarez et al., 2016; Lieberman, 2009). For example, jury instructions do not influence mock jurors' perceptions of eyewitness accuracy (Perez & Miller, 2015), reduce mock jurors' perceptions of guilt in recanted confession cases (Gomes et al., 2016; Kassin & Sukel, 1997), or decrease mock jurors' reliance on inadmissible evidence learned through pretrial publicity (Fein et al., 1997). However, other research shows that in some cases, jury instructions can limit the use of inadmissible evidence when judges indicate the reason why inadmissible evidence should be excluded and the instructions do not contain complicated legal jargon and convoluted legal issues (Alvarez et al., 2016; Nietzel et al., 1999; Steblay et al., 2006). Additionally, some research shows that presenting jury instructions before the evidence in the case, rather than after all the evidence has been presented, can increase comprehension and produce more favorable outcomes for the accused (Bourgeois et al., 1995; Conklin, in press; ForsterLee et al., 1993; Kassin & Wrightsman, 1979). Given that the Canadian Model Jury Instructions for audio recordings and written transcripts are written in simple language, are presented before triers of fact evaluate the evidence, and clearly indicate that the transcripts are only an aid to help jurors follow the recordings rather than being direct evidence, participants might be able to more easily and effectively follow these instructions.

Furthermore, this provides initial evidence that these instructions can lead to more critical evaluation of audio recordings and written transcripts. This is important given that these instructions are currently administered in cases involving this type of evidence, but their effectiveness has not, until now, been formally tested through research. This is also an important finding given that written transcripts are often produced by individuals with information about the context from which the recordings were taken. For example, in some jurisdictions, police officers are responsible for producing written transcripts to accompany degraded audio recordings (French & Fraser, 2018). Police officers clearly have a suspicion of guilt (otherwise, they wouldn't suspect the individual committed an offense), and this belief can lead them to perceive the recordings as incriminating (Charman et al., 2017; Lange et al., 2011). Even when police officers are not responsible for producing transcripts, we know that other people's general knowledge that the recordings came from a criminal investigation can similarly bias their interpretations. When incriminating transcripts make their way into the courtroom, they can bias triers' interpretations of degraded audio. Thus, it is important that there is a safeguard to decrease the prejudicial effects of such transcripts.

While it is encouraging that these instructions seem to be effective in the face of incriminating transcripts, future research should consider whether this remains true when there is additional incriminating evidence presented in the case. Oftentimes, wrongful conviction cases involve multiple lines of incriminating evidence (Kassin et al., 2012). Thus, even if an accused person is factually innocent, there can be multiple forms of evidence which seem to incriminate him. It is important to determine whether, in the face of multiple lines of incriminating evidence, these instructions are still effective in mitigating participants' tendency to make incriminating misinterpretations of degraded audio evidence. This will provide important information about the utility of these instructions across various contexts.

Although instructions were effective in the no delay condition, there were no differences between the instructions absent and instructions present conditions when there was a one-week delay between participants' exposure to the incriminating transcripts and when they transcribed the statements. I did not hypothesize this data pattern, but rather, expected a main effect of instructions and a main effect of delay. This data pattern seemed to be due, in part, to the fact that participants in the one-week delay condition made fewer incriminating misinterpretations overall than participants in the no delay condition. Thus, participants in the one-week delay condition may not have been able to remember the transcripts they saw one week prior. Indeed, when I compared participants in the delay condition to those in the no transcripts condition, there was no difference in the proportion of incriminating misinterpretations made between participants who received transcripts (M = 0.25) and those who did not (M = 0.26), $F(1, 242) = 0.63$, $p = .429$, $\eta_p^2 = .003$, 95% CI [-.02, .04] ($BF_{INCL} = 0.21$). This finding lends support to the notion that participants in the one-week delay condition could not remember the transcripts. This is an important finding given that, in some cases, written transcripts are only shown to triers of fact in the courtroom when they initially listen to degraded audio evidence but are not filed as exhibits. Thus, the transcripts cannot be used during the deliberation process. If a delay truly reduces participants' tendency to make incriminating misinterpretations because participants can no longer remember the transcripts when making their final evaluations of the audio, perhaps there is some support for allowing triers of fact to consider written transcripts in their initial evaluations of degraded audio evidence but not in their deliberation.

While this study suggests that a delay on its own might sufficiently suppress the biasing effects of presenting incriminating transcripts, future research should replicate this interaction and explore the bounds of this delay effect. For example, it is unclear whether we could expect to observe this same effect if the delay were shorter. Additionally, if there were fewer target statements to evaluate, such as what might be expected in a real criminal trial involving one accused person, participants might be more likely to recall, and subsequently rely on, the incriminating transcripts after a delay. Furthermore, it is unclear whether multiple exposures to the written transcripts would impact this effect. In this study, participants only saw each of the transcribed statements once. Perhaps listening to the statements with the transcripts multiple times would not only increase participants' memory of the transcripts, but also increase their tendency to hear in the degraded recordings what they read in the written transcripts. Indeed, similar research on auditory hindsight bias found that the more times participants were primed with the clear version of a word before hearing a degraded version of the same word in a later test phase, the more likely they were to identify that degraded word in the test phase (Bernstein et al., 2018).

While incriminating transcripts led to significantly more incriminating misinterpretations, they did not significantly increase participants' perceptions of the speakers' guilt. This finding contradicts what a confirmation bias effect would predict. However, one important factor that might have impacted these results is that participants with no contextual information were asked to rate the likelihood that the individuals speaking in the recordings were guilty. Because they had no previous reason to believe the recordings were nefarious, being asked to indicate the likelihood that the accused was guilty might have surprised them. Therefore, they may have been more skeptical about the incriminating nature of all the statements, including statements that they ultimately transcribed as non-incriminating. Alternatively, participants in the written transcript condition expected to hear incriminating statements. Consequently, when they heard statements that they perceived as non-incriminating, they may have assigned lower likelihood of guilt ratings than participants in the no transcripts condition.

To determine whether this was in fact the case, I conducted an ANCOVA with transcripts (no transcripts vs. written transcripts) as the independent variable, average likelihood of guilt ratings for non-incriminating interpretations only (including accurate interpretations and non-incriminating errors) as the dependent variable, and LexTALE score as the covariate. I found that, indeed, participants in the no transcripts condition assigned higher likelihood of guilt ratings to non-incriminating statements (M = 4.49) than participants in the written transcripts condition (M = 3.93), $F(1, 170) = 5.04$, p = .026, η_p^2 = .03, 95% CI [.07, 1.07] (BF_{INCL} = 13.27). Thus, while participants in the written transcripts condition assigned lower likelihood of guilt ratings to speakers when they interpreted their statements as non-incriminating, participants in the no transcripts condition appeared to be more skeptical of the speakers' guilt (see Table 2).

Table 2. Participants' Mean Likelihood of Guilt Ratings [and Confidence Intervals] in Study 3

	No Transcripts	Transcripts			
		No Delay		One-Week Delay	
		No Instructions	Instructions	No Instructions	Instructions
Overall Likelihood of Guilt	5.32 [4.96, 5.68]	5.12 [4.72, 5.52]	4.56 [4.23, 4.88]	4.89 [4.54, 5.24]	4.74 [4.44, 5.03]
Likelihood of Guilt for Incriminating Interpretations	7.62 [7.25, 7.98]	7.89 [7.48, 8.29]	7.79 [7.49, 8.10]	7.36 [7.00, 7.72]	7.49 [7.18, 7.79]
Likelihood of Guilt for Non-Incriminating Interpretations	4.49 [4.16, 4.83]	3.93 [3.56, 4.29]	3.56 [3.28, 3.85]	4.05 [3.72, 4.39]	3.90 [3.62, 4.18]

5.6. Limitations

Unfortunately, it is challenging to assess the no transcripts condition's likelihood of guilt ratings without inadvertently providing them with some context about the case. The very nature of asking them to rate the speakers' likelihood of guilt may have led them to believe the recordings were incriminating. Thus, in much the same way that asking participants to rate the speakers' likelihood of guilt appeared to affect their overall perceptions of guilt in the no transcripts condition, this may have similarly affected participants' interpretations of the audio recordings. While I found that participants in the written transcripts condition made significantly more incriminating misinterpretations than participants in the no transcripts condition, it is likely that this effect was underestimated in the current data due to there being more incriminating misinterpretations in the "no context" condition than what I would typically expect to observe. Indeed, the proportion of incriminating misinterpretations that participants in the no context (i.e., no transcripts) condition made in Study 3 (M = .26) was notably higher than the proportion made in Study 1 (*M* = .22) and Study 2 (*M* = .13; collapsed across degradation level).

Another limitation of this work is that participants made their decisions alone rather than in groups. Given that juries deliberate in groups, it is possible that these results do not fully capture the processes that would operate in a jury deliberation after a delay. For example, if some participants remember the incriminating transcripts after the delay, they could remind other mock jury members of what they read in the transcripts. This could subsequently lead the group as a whole to perceive the recordings as more incriminating. Previous work on confirmation bias has shown that this bias can operate in group decision-making (Schulz-Hardt et al., 2000). Thus, it is possible that asking participants to deliberate in groups would lead to confirmation bias in the delay condition as well. Future work should test the effects of delay and instructions on mock jury's decisions.

Chapter 6.

General Discussion

Audio recordings are increasingly being used as evidence in criminal trials (Fishman, 2006; Fraser, 2018; Lange et al., 2011). In some circumstances, these recordings may be hard to hear due to factors such as wiretapping, poor connections in phone calls, or background noise. Despite the increased prevalence of degraded forensic audio evidence, there is a lack of research exploring how cognitive biases such as confirmation bias may affect individuals' evaluations of such evidence. Thus, the goal of the current research was to examine how confirmation bias influences individuals' evaluations of degraded forensic audio evidence.

Across three studies, I demonstrated the pervasiveness of confirmation bias in cases involving degraded audio recordings. In Study 1, I replicated and extended Lange et al.'s (2011) findings by exploring the effects of contextual information on participants' interpretations of degraded audio recordings. I investigated whether participants' knowledge that the recordings came from wiretapped conversations with criminal suspects influenced their interpretations of degraded audio. Furthermore, I explored whether there is an additive effect of contextual information on confirmation bias by informing some participants that there was also eyewitness evidence implicating the accused. I found that compared to participants with no contextual information, participants with contextual information (i.e., in the criminal suspect and criminal suspect + eyewitness evidence conditions) made more incriminating misinterpretations (see Figure 2).

This finding provides clear evidence that confirmation bias can operate in these types of judgments. However, learning additional contextual information did not increase this confirmation bias effect. Thus, I found no support for additivity. This may have been due to the nature of the additional contextual information presented in this research. For example, the eyewitness evidence may not have been detailed enough to further impact participants' judgments. Alternatively, perhaps other types of evidence (e.g., confession evidence) would have been more suggestive of the accused's guilt, and thus, would have increased participants' tendency to make incriminating misinterpretations. Future work should further explore this question. Regardless, this work has important implications for understanding the circumstances under which people's pre-existing beliefs influence their evaluations of degraded forensic audio evidence. Simply learning that the recordings came from a criminal context led to systematic misinterpretations of degraded audio. This can have serious consequences for legal cases involving degraded audio evidence.

In Study 2, I examined whether fluency misattribution is a mechanism that underlies confirmation bias in cases involving degraded audio evidence. This is an important contribution to the field given that research has not yet begun to explore the mechanisms that underlie confirmation bias in these cases. To test whether fluency is a mechanism for confirmation bias in cases involving degraded audio recordings, I manipulated the degree to which the audio recordings were degraded and the amount of contextual information participants learned. I found that as the recordings became more degraded, it took longer for participants to transcribe what they heard (see Figure 7). This suggests that degrading the recordings reduced perceptual fluency.

I also found that across degradation levels, participants with more contextual information made more incriminating misinterpretations than participants with less contextual information (see Figure 6). Additionally, compared to participants in the no context and criminal suspect conditions, participants in the written transcripts condition more quickly reached their decisions about what they heard in the recordings as evidenced by quicker reaction times (see Figure 7). This suggests that participants with more contextual information experienced increased conceptual fluency. However, it is important to note that the criminal suspect condition's reaction times did not significantly differ from the no context condition's reaction times.

Furthermore, participants' reaction times partially mediated participants' tendency to make incriminating misinterpretations for minimally and moderately degraded recordings, providing preliminary evidence that fluency misattribution contributes to confirmation bias for degraded audio recordings. Given the increased prevalence of recorded conversations as evidence and the potential for erroneous decision-making in legal contexts, it is essential to understand when and why confirmation bias influences people's interpretations of degraded audio recordings. Future work should continue to explore objective and subjective measures of fluency as well as other potential mechanisms.

Finally, given the risks that written transcripts containing incriminating errors pose, I tested whether the current Canadian Model Jury Instructions for audio recordings and written transcripts effectively mitigate biased evaluations of degraded audio evidence in Study 3. A major assumption of the current Canadian Model Jury Instructions is that individuals can identify discrepancies between what they read in the written transcripts and what they hear in the degraded audio recordings. However, this assumption has not been formally tested. Additionally, in some cases, transcripts may only be administered to triers in the courtroom. That is, triers may not be permitted to utilize the transcripts when they deliberate. Thus, it is important to understand how previous exposure to transcripts influences individuals' later evaluations of degraded audio recordings without transcripts.

I found an interaction between the presence of instructions and the presence of a delay in Study 3 (see Figure 14). Specifically, when participants transcribed the recordings immediately after an exposure phase in which they saw the transcripts while listening to the recordings, the instructions effectively reduced participants' tendency to make incriminating misinterpretations. Alternatively, when participants transcribed the recordings one week after their exposure to the incriminating transcripts, there were no differences between the instruction present and instruction absent conditions. Thus, the instructions appeared to reduce incriminating misinterpretations, but only when participants made their final decisions about what they heard in the recordings shortly after exposure to incriminating transcripts (i.e., in the no delay condition). It is possible that there were no effects of instructions in the one-week delay condition given that participants did not remember the incriminating transcripts after the delay. In any case, these results suggest that it may be possible to reduce confirmation bias in cases involving degraded audio evidence. This is an important contribution to the literature because it can help to reduce erroneous decision-making in legal settings. Future work should explore the effectiveness of these instructions and impacts of delay in the context of individual cases involving additional forms of evidence and multiple exposures to the incriminating transcripts.

Overall, participants' exposure to both the general legal context (Studies 1 and 2), and incriminating transcripts specifically (Studies 2 and 3) systematically biased their interpretations of degraded audio recordings. While this finding is concerning in its own right, it is important to note that in real criminal cases, triers of fact will almost certainly encounter more contextual information than what participants learned in the current set of studies (Lange et al., 2011). Thus, we might expect to see even stronger bias in real criminal cases. Furthermore, people's incriminating misinterpretations of degraded audio evidence could subsequently influence their evaluations of other evidence in a case. Once they hear an accused incriminate himself, this can further strengthen their belief in the accused's guilt and lead them to evaluate other evidence in a more incriminating manner (Dror et al. 2017; Kassin et al., 2013).

These findings also suggest the need for more standardization in the processes of both producing written transcripts and determining their admissibility. The current research establishes that people's knowledge that audio recordings may contain incriminating content can systematically influence their interpretations. Thus, if written transcripts must be produced, they should ideally be developed by individuals with no contextual information, including the fact that the recordings come from a legal investigation. To further limit the biasing effects of context, transcriptionists may even consider identifying words from the recordings in isolation rather than in the context of complete sentences. We naturally use the other words in a sentence to provide context to our interpretations of ambiguous words (Miller et al., 1951; Whittlesea, 1993). It is much harder to identify single words in isolation than it is to identify words in the context of a sentence. Therefore, perhaps transcriptionists should try to interpret words, and particularly ambiguous words, in isolation rather than in the context of sentences. This approach may also work for triers without transcripts, although the context of the criminal case may be enough to bias their interpretations. However, I acknowledge that this approach may be challenging to implement practically given that the recordings are collected as full sentences rather than single words. Regardless, limiting as much contextual information as possible can help to mitigate some of this bias (Dror et al., 2015).

Additionally, when possible, multiple individuals should produce their own transcripts and their interpretations should be compared and verified. If there is no clear consensus on what is said in the recordings, this may be grounds for determining that transcripts should not be permitted to accompany the audio. Alternatively, triers could be made aware of the reliability of the transcripts. As Lange et al. (2011) suggest, reporting precise reliability information about the proportion of transcriptionists who agreed with a particular interpretation (e.g., 10 out of 20) may help to adjust triers' perceptions of the evidence, though this assertion should be tested. Another option is to omit inaudible words from transcripts. That is, if a word is so unclear that a reasonable decision cannot be reached about what the speaker said, it should be noted as such in the transcripts (e.g., by putting "[inaudible]" in place of the word(s) in the transcripts).

Judges ultimately determine the admissibility of audio recordings and written transcripts, and they have broad discretion in making this determination (State v. Lavers, 1991; U.S. v. Harrell, 1986; U.S. v. Stephens, 2002). However, previous research shows that judges are not immune to bias (Oeberst & Goeckenjan, 2016; Parks, 2015; Rachlinski et al., 2008). Judges' knowledge of other evidence in the case or their examination of the written transcripts are likely to influence their beliefs about the quality of audio recordings and the reliability of associated transcripts. Furthermore, how judges determine the quality and reliability of such evidence is likely to vary across judges. At the very least, judges should be made aware of the potential issues associated with such evidence. Ideally, there should be standardized procedures for determining the admissibility of degraded audio evidence and written transcripts.

6.1. Conclusion

In the current set of studies, I consistently found that incriminating contextual information biased participants' interpretations of degraded audio recordings. Whether participants learned general information about the recordings coming from a criminal context or received written transcripts containing incriminating errors, they were more likely than participants without contextual information to make incriminating misinterpretations. This work not only contributes to the growing literature on confirmation bias in forensic settings (Kassin et al., 2013), but also has methodological, theoretical, and applied significance.

I contributed to methodology by creating a large set of degraded audio recordings that elicit confirmation bias given various types of incriminating contextual information. I contributed to theory by providing preliminary evidence that fluency misattribution, as measured through reaction time, plays a role in confirmation bias for degraded audio recordings. This may also inform future research exploring the mechanisms that underlie confirmation bias across different modalities. Finally, I contributed to practice by (1) demonstrating the impacts of different types of forensically-relevant contextual information on participants' evaluations of degraded audio recordings and (2) producing the first empirical evidence that the Canadian Model Jury Instructions for audio recordings and written transcripts can mitigate the biasing effects of context in these cases. As audio recordings become increasingly prevalent as evidence in legal cases, it is imperative that we continue to study the factors that affect people's interpretations of this type of evidence and understand the implications of using audio recordings in civil and criminal cases.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219-235. https://doi.org/10.1177/1088868309341564

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

Alvarez, M. J., Miller, M. K., & Bornstein, B. H. (2016). "It will be your duty…:" The psychology of criminal jury instructions. In *Advances in psychology and law* (pp. 119-158). Springer, Cham.

Anderson, N. H., & Jacobson, A. (1965). Effect of stimulus inconsistency and discounting instructions in personality impression formation. *Journal of Personality and Social Psychology, 2,* 531-539. https://doi.org/10.1037/h0022484

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*(3), 258-290. https://doi.org/10.1037/h0055756

Ask, K., & Granhag, P. A. (2005). Motivational sources of confirmation bias in criminal investigations: The need for cognitive closure. *Journal of Investigative Psychology and Offender Profiling*, *2*(1), 43-63. https://doi.org/10.1002/jip.19

Ask, K., Rebelius, A., & Granhag, P. A. (2008). The 'elasticity' of criminal evidence: A moderator of investigator bias. *Applied Cognitive Psychology*, *22*(9), 1245-1259. https://doi.org/10.1002/acp.1432

Bacon, F. (1939). Novum Organum. In Burtt, E. A. (Ed.), *The English Philosophers from Bacon to Mill* (pp. 24-123). New York: Random House. (Original work published in 1620)

Bell, B. E., & Loftus, E. F. (1988). Degree of detail of eyewitness testimony and mock juror judgments. *Journal of Applied Social Psychology*, *18*(14), 1171-1192. https://doi.org/10.1111/j.1559-1816.1988.tb01200.x

Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2014). Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American Journal of Political Science*, *58*(3), 739-753. https://doi.org/10.1111/ajps.12081

Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2016). Can we turn shirkers into workers? *Journal of Experimental Social Psychology*, *66*, 20-28. https://doi.org/10.1016/j.jesp.2015.09.010

Bernstein, D. M., Kumar, R., Masson, M. E., & Levitin, D. J. (2018). Fluency misattribution and auditory hindsight bias. *Memory & Cognition*, *46*, 1331–1343. https://doi.org/10.3758/s13421-018-0840-6

Bernstein, D. M., Wilson, A. M., Pernat, N. L., & Meilleur, L. R. (2012). Auditory hindsight bias. *Psychonomic Bulletin & Review*, *19*(4), 588-593. https://doi.org/10.3758/s13423-012-0268-0

Bild, E., Redman, A., Newman, E. J., Muir, B. R., Tait, D., & Schwarz, N. (2021). Sound and credibility in the virtual court: Low audio quality leads to less favorable evaluations of witnesses and lower weighting of evidence. *Law and Human Behavior*, *45*(5), 481-495. https://doi.org/10.1037/lhb0000466

Bourgeois, M. J., Horowitz, I. A., ForsterLee, L., & Grahe, J. (1995). Nominal and interactive groups: Effects of preinstruction and deliberations on decisions and evidence recall in complex trials. *Journal of Applied Psychology*, *80*(1), 58-67. https://doi.org/10.1037/0021-9010.80.1.58

Bressan, P., & Dal Martello, M. F. (2002). Talis pater, talis filius: Perceived resemblance and the belief in genetic relatedness. *Psychological Science*, *13*(3), 213-218. https://doi.org/10.1111/1467-9280.00440

Bruner, J. S., & Potter, M. C. (1964). Interference in visual recognition. *Science*, *144*(3617), 424-425. https://doi.org/10.1126/science.144.3617.424

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3-5. https://doi.org/10.1177/1745691610393980

Cacciarelli v. Boniface, 737 A.2d 1170, 1174-76, N.J. Super. Ct. Ch. Div. (1999)

Canadian National Judicial Institute Model Jury Instructions (2011, March). Audio tape recordings and transcripts. Retrieved from: https://www.nji-inm.ca/index.cfm/publications/model-jury-instructions/mid-trial-instructions/audio-tape-recordings-and-transcripts/

Cave, C. B. (1997). Very long-lasting priming in picture naming. *Psychological Science*, *8*(4), 322-325. https://doi.org/10.1111/j.1467-9280.1997.tb00446.x

Charman, S. D., Gregory, A. H., & Carlucci, M. (2009). Exploring the diagnostic utility of facial composites: Beliefs of guilt can bias perceived similarity between composite and suspect. *Journal of Experimental Psychology: Applied*, *15*(1), 76-90. https://doi.org/10.1037/a0014682

Charman, S. D., Kavetski, M., & Mueller, D. H. (2017). Cognitive bias in the legal system: Police officers evaluate ambiguous evidence in a belief-consistent manner. *Journal of Applied Research in Memory and Cognition, 6*(2), 193-202. https://doi.org/10.1016/j.jarmac.2017.02.001

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, *11*(4), 464-473. https://doi.org/10.1177/1948550619875149

Clifford, B. R. (1980). Voice identification by human listeners: On earwitness reliability. *Law and Human Behavior, 4*(4)*,* 373-394. https://doi.org/10.1007/BF01040628

Conklin, M. (in press). I knew it all along: The promising effectiveness of a pre-jury instruction at mitigating hindsight bias. *Baylor Law Review.* http://dx.doi.org/10.2139/ssrn.3962419

Criminal Code, R.S.C. 1985, c. C-46 as amended.

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Development*, *82*(3), 766-779. https://doi.org/10.1111/j.1467-8624.2010.01529.x

Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, *44*(1), 20-33. https://doi.org/10.1037/0022-3514.44.1.20

DeCoster, J., & Claypool, H. M. (2004). A meta-analysis of priming effects on impression formation supporting a general model of informational biases. *Personality and Social Psychology Review*, *8*(1), 2-27. https://doi.org/10.1207/S15327957PSPR0801_1

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5,* https://doi.org/10.3389/fpsyg.2014.00781

Droit de la famille – 152607 (2015), Q.J. No. 10930

Dror, I. E. (2016). A hierarchy of expert performance. *Journal of Applied Research in Memory and Cognition*, *5*(2), 121-127. https://doi.org/10.1016/j.jarmac.2016.03.001

Dror, I. E., Charlton, D., & Péron, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, *156*(1), 74-78. https://doi.org/10.1016/j.forsciint.2005.10.017

Dror, I. E., & Cole, S. A. (2010). The vision in “blind” justice: Expert perception, judgment, and visual cognition in forensic pattern recognition. *Psychonomic Bulletin & Review*, *17*(2), 161-167. https://doi.org/10.3758/PBR.17.2.161

Dror, I. E., & Hampikian, G. (2011). Subjectivity and bias in forensic DNA mixture interpretation. *Science and Justice*, *51*(4), 204-208. https://doi.org/10.1016/j.scijus.2011.08.004

Dror, I. E., Morgan, R., Rando, C., & Nakhaeizadeh, S. (2017). The bias snowball and the bias cascade effects: Two distinct biases that may impact forensic decision making. *Journal of Forensic Sciences*, *62*(3), 832-833. https://doi.org/10.1111/1556-4029.13496

Dror, I. E., Thompson, W. C., Meissner, C. A., Kornfield, I., Krane, D., Saks, M., & Risinger, M. (2015). Context management toolbox: A linear sequential unmasking (LSU) approach for minimizing cognitive bias in forensic decision making. *Journal of Forensic Sciences, 60*(4), 1111-1112. https://doi.org/10.1111/1556-4029.12805

Duncan, B. L. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of blacks. *Journal of Personality and Social Psychology*, *34*(4), 590-598. https://doi.org/10.1037/0022-3514.34.4.590

Elaad, E., Ginton, A., & Ben-Shakhar, G. (1994). The effects of prior expectations and outcome knowledge on polygraph examiners' decisions. *Journal of Behavioral Decision Making*, *7*(4), 279-292. https://doi.org/10.1002/bdm.3960070405

Elliot, A. J., & Devine, P. G. (1994). On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of Personality and Social Psychology*, *67*(3), 382-394.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Fein, S., McCloskey, A. L., & Tomlinson, T. M. (1997). Can the jury disregard that information? The use of suspicion to reduce the prejudicial effects of pretrial publicity and inadmissible testimony. *Personality and Social Psychology Bulletin*, 23, 1215–1226. https://doi.org/10.1177/01461672972311008

Festinger, L. (1957). *A cognitive theory of dissonance*. Stanford, CA: Stanford University Press.

Findley, K. A., & Scott, M. S. (2006). Multiple dimensions of tunnel vision in criminal cases. *Wisconsin Law Review*, 291-398.

Fishman, C. S. (2006). Recordings, transcripts, and translations as evidence. *Washington Law Review*, *81*, 473-524.

Forster, M., Leder, H., & Ansorge, U. (2013). It felt fluent, and I liked it: Subjective feeling of fluency rather than objective fluency determines liking. *Emotion*, *13*(2), 280-289. https://doi.org/10.1037/a0030115

ForsterLee, L., Horowitz, I. A., & Bourgeois, M. J. (1993). Juror competence in civil trials: Effects of preinstruction and evidence technicality. *Journal of Applied Psychology*, *78*(1), 14-21. https://doi.org/10.1037/0021-9010.78.1.14

Fraser, H. (2018). 'Assisting' listeners to hear words that aren't there: Dangers in using police transcripts of indistinct covert recordings. *Australian Journal of Forensic Sciences*, *50*(2), 129-139. https://doi.org/10.1080/00450618.2017.1340522

French, P., & Fraser, H. (2018). Why "Ad Hoc Experts" should not provide transcripts of indistinct audio, and a better approach. *Criminal Law Journal*, 298-302.

Gale, M., & Ball, L. J. (2002). Does positivity bias explain patterns of performance on Wason's 2-4-6 task? In W. D. Gray, & C. D. Schunn (Eds.) *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 340-344). Mahwah, NJ: Lawrence Erlbaum Associates.

Galinsky, A. D., & Moskowitz, G. B. (2000). Counterfactuals as behavioral primes: Priming the simulation heuristic and consideration of alternatives. *Journal of Experimental Social Psychology*, *36*(4), 384-409. https://doi.org/10.1006/jesp.1999.1409

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review, 19*(5), 847–857. https://doi.org/10.3758/s13423-012-0296-9

Golden v. State, 429 So.2d 45 (Fla. App. 1 Dist, 1983).

Gomes, D. M., Stenstrom, D. M., & Calvillo, D. P. (2016). Examining the judicial decision to substitute credibility instructions for expert testimony on confessions. *Legal and Criminological Psychology*, *21*(2), 319-331. https://doi.org/10.1111/lcrp.12068

Gregory, R.L. (1970). *The intelligent eye.* New York: McGraw-Hill.

Gregory, R.L. (1974). *Concepts and mechanisms of perception*. New York: Scribners.

Hasel, L. E., & Kassin, S. M. (2009). On the presumption of evidentiary independence: Can confessions corrupt eyewitness identifications? *Psychological Science*, *20*(1), 122-126. https://doi.org/10.1111/j.1467-9280.2008.02262.x

Haselton, M. G., Nettle, D., & Murray, D.R. (2015). The evolution of cognitive bias. In D. M. Buss (Ed.), *The Handbook of Evolutionary Psychology* (pp. 724–746). New York: Wiley.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400-407. https://doi.org/10.3758/s13428-015-0578-z

Hayden, T., & Mischel, W. (1976). Maintaining trait consistency in the resolution of behavioral inconsistency: The wolf in sheep's clothing? *Journal of Personality*, *44*(1), 109-132. https://doi.org/10.1111/j.1467-6494.1976.tb00587.x

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*(4), 408-420. https://doi.org/10.1080/03637750903310360

Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research*, *50*(1), 1-22. https://doi.org/10.1080/00273171.2014.962683

Hayes, A. F., & Rockwood, N. J. (2020). Conditional process analysis: Concepts, computation, and advances in the modeling of the contingencies of mechanisms. *American Behavioral Scientist*, *64*(1), 19-54. https://doi.org/10.1177/0002764219859633

Hernandez, I., & Preston, J. L. (2013). Disfluency disrupts the confirmation bias. *Journal of Experimental Social Psychology*, *49*(1), 178-182. https://doi.org/10.1016/j.jesp.2012.08.010

Higgins, E. T., Bargh, J. A., & Lombardi, W. J. (1985). Nature of priming effects on categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(1), 59-69. https://doi.org/10.1037/0278-7393.11.1.59

Higham, P. A., Neil, G. J., & Bernstein, D.M. (2017). Auditory hindsight bias: Fluency misattribution versus memory reconstruction. *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 1143-1159. https://doi.org/10.1037/xhp0000405

Hilbig, B. E. (2016). Reaction time effects in lab-versus Web-based research: Experimental evidence. *Behavior Research Methods*, *48*(4), 1718-1724. https://doi.org/10.3758/s13428-015-0678-9

Hodgins, H. S., & Zuckerman, M. (1993). Beyond selecting information: Biases in spontaneous questions and resultant conclusions. *Journal of Experimental Social Psychology*, *29*(5), 387-407. https://doi.org/10.1006/jesp.1993.1018

Hu, X., Zheng, J., Fan, T., Su, N., Yang, C., & Luo, L. (2020). Using multilevel mediation model to measure the contribution of beliefs to judgments of learning. *Frontiers in Psychology*, *11*, 637. https://doi.org/10.3389/fpsyg.2020.00637

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513-541. https://doi.org/10.1016/0749-596X(91)90025-F

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/10.1037/0096-3445.110.3.306

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger L F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving*, (pp. 391-422). Hillsdale, NJ: Erlbaum.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, *118*(2), 126-135. https://doi.org/10.1037/0096-3445.118.2.126

Johnston, W. A., Dark, V. J., & Jacoby, L. L. (1985). Perceptual fluency and recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(1), 3-11. https://doi.org/10.1037/0278-7393.11.1.3

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, *80*(4), 557-571. https://doi.org/10.103700022-3514.80.4.557

Kappes, A., Harvey, A. H., Lohrenz, T., Montague, P. R., & Sharot, T. (2020). Confirmation bias in the utilization of others' opinion strength. *Nature Neuroscience*, *23*(1), 130-137. https://doi.org/10.1038/s41593-019-0549-2

Kassin, S. M. (2012). Why confessions trump innocence. *American Psychologist, 67*, 431–445. https://doi.org/10.1037/a0028212

Kassin, S. M., Bogart, D., & Kerner, J. (2012). Confessions that corrupt: Evidence from the DNA exoneration case files. *Psychological Science*, *23*(1), 41-45. https://doi.org/10.1177/0956797611422918

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, *2*(1), 42-52. https://doi.org/10.1016/j.jarmac.2013.01.001

Kassin, S. M., Goldstein, C. C., & Savitsky, K. (2003). Behavioral confirmation in the interrogation room: On the dangers of presuming guilt. *Law and Human Behavior*, *27*(2), 187-203. https://doi.org/10.1023/A:1022599230598

Kassin, S. M., & Wrightsman, L. S. (1979). On the requirements of proof: The timing of judicial instruction and mock juror verdicts. *Journal of Personality and Social Psychology*, *37*(10), 1877-1887. https://doi.org/10.1037/0022-3514.37.10.1877

Kerstholt, J. H., Jansen, N. J., Van Amelsvoort, A. G., & Broeders, A. P. A. (2006). Earwitnesses: Effects of accent, retention and telephone. *Applied Cognitive Psychology*, *20*(2), 187-197. https://doi.org/10.1002/acp.1175

Kim, J., Gabriel, U., & Gygax, P. (2019). Testing the effectiveness of the Internet-based instrument PsyToolkit: A comparison between web-based (PsyToolkit) and lab-based (E-Prime 3.0) measurements of response choice and response time in a complex psycholinguistic task. *PLoS One*, *14*(9), e0221802. https://doi.org/10.1371/journal.pone.0221802

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*(2), 211-228. https://doi.org/10.1037/0033-295X.94.2.211

Knobloch-Westerwick, S., Mothes, C., & Polavin, N. (2020). Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information. *Communication Research*, *47*(1), 104-124. https://doi.org/10.1177/0093650217719596

Krane, D. E., Ford, S., Gilder, J. R., Inman, K., Jamieson, A., Koppl, R., ... & Thompson, W. C. (2008). Sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation. *Journal of Forensic Sciences*, *53*(4), 1006-1007. https://doi.org/10.1111/j.1556-4029.2008.00787.x

Kray, L. J., & Galinsky, A. D. (2003). The debiasing effect of counterfactual mind-sets: Increasing the search for disconfirmatory information in group decisions. *Organizational Behavior and Human Decision Processes*, *91*(1), 69-81. https://doi.org/10.1016/S0749-5978(02)00534-4

Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind: "Seizing" and "freezing". *Psychological Review*, *103*(2), 263-283. https://doi.org/10.1037/0033-295X.103.2.263

Kruglanski, A. W., Webster, D. M., & Klem, A. (1993). Motivated resistance and openness to persuasion in the presence or absence of prior information. *Journal of Personality and Social Psychology*, *65*(5), 861-876. https://doi.org/10.1037/0022-3514.65.5.861

Kukucka, J., & Kassin, S. M. (2014). Do confessions taint perceptions of handwriting evidence? An empirical test of the forensic confirmation bias. *Law and Human Behavior*, *38*(3), 256-270. https://doi.org/10.1037/lhb0000066

Lange, N. D., Thomas, R. P., Dana, J., & Dawes, R. M. (2011). Contextual biases in the interpretation of auditory evidence. *Law and Human Behavior*, *35*(3), 178-187. https://doi.org/10.1007/s10979-010-9226-4

Langer, E. J., & Abelson, R. P. (1974). A patient by any other name...: Clinician group difference in labeling bias. *Journal of Consulting and Clinical Psychology*, *42*(1), 4-9. https://doi.org/10.1037/h0036054

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325-343. https://doi.org/10.3758/s13428-011-0146-0

Lidén, M., Gräns, M., & Juslin, P. (2018). The presumption of guilt in suspect interrogations: Apprehension as a trigger of confirmation bias and debiasing techniques. *Law and Human Behavior*, *42*(4), 336-354. http://dx.doi.org/10.1037/lhb0000287

Lieberman, J. D. (2009). The psychology of the jury instruction process. In J. D. Lieberman & D. A. Krauss (Eds.), *Jury Psychology: Social Aspects of Trial Processes: Psychology in the Courtroom* (pp. 129–155). Ashgate Publishing Co.

Lingle, J. H., & Ostrom, T. M. (1981). Principles of memory and cognition in attitude formation. In R. E. Petty, T. M. Ostrom, & T. C. Brock (Eds.), *Cognitive Responses in Persuasive Communications: A text in Attitude Change* (pp. 399-420). Hillsdale, NJ: Erlbaum.

Lopez v. United States, 373 US 427 (1963)

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098-2109. https://doi.org/10.1037/0022-3514.37.11.2098

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, *58*, 593-614. https://doi.org/10.1146/annurev.psych.58.110405.085542

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252-271. https://doi.org/10.1037/0033-295X.87.3.252

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1-23. https://doi.org/10.3758/s13428-011-0124-6

Matthiesen, B.W., Wickert, G.L., & Lehrer, D.W. (2018, August 6). *Laws on recording conversations in all 50 states.* Retrieved from https://www.mwl-law.com/wp-content/uploads/2013/03/LAWS-ON-RECORDING-CONVERSATIONS-CHART.pdf

McLaughlin, E.C. (2017, April 26). *Suspect OKs Amazon to hand over echo recordings in murder case.* Retrieved from: https://www.cnn.com/2017/03/07/tech/amazon-echo-alexa-bentonville-arkansas-murder-case/index.html

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, *41*(5), 329-335. https://doi.org/10.1037/h0062491

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, *89*(6), 852-863. https://doi.org/10.1037/0022-3514.89.6.852

Munro, G. D., & Stansbury, J. A. (2009). The dark side of self-affirmation: Confirmation bias and illusory correlation in response to threatening information. *Personality and Social Psychology Bulletin*, *35*(9), 1143-1153. https://doi.org/10.1177/0146167209337163

Murrie, D. C., Boccaccini, M. T., Guarnera, L. A., & Rufino, K. A. (2013). Are forensic experts biased by the side that retained them? *Psychological Science*, *24*(10), 1889-1897. https://doi.org/10.1177/0956797613481812

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, *29*(1), 85-95. https://doi.org/10.1080/00335557743000053

National Academy of Sciences (2009). Strengthening Forensic Science in the United States: A Path Forward. Washington, DC: National Academies Press.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175-220. https://doi.org/10.1037/1089-2680.2.2.175

Nietzel, M. T., McCarthy, D. M., & Kern, M. J. (1999). Juries: The current state of the empirical literature. In R. Roesch, S. D. Hart, & J. R. P. Ogloff (Eds.), *Psychology and Law: The State of the Discipline* (pp. 23–42). New York: Kluwer/Plenum.

O'Brien, B. (2009). Prime suspect: An examination of factors that aggravate and counteract confirmation bias in criminal investigations. *Psychology, Public Policy, and Law*, *15*(4), 315-334. https://doi.org/10.1037/a0017881

O'Brien, B., & Ellsworth, P. C. (2006). Confirmation bias in criminal investigations. Retrieved from: https://ssrn.com/abstract=913357

Oeberst, A., & Goeckenjan, I. (2016). When being wise after the event results in injustice: Evidence for hindsight bias in judges' negligence assessments. *Psychology, Public Policy, and Law*, *22*(3), 271-279. https://doi.org/10.1037/law0000091

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237-241. https://doi.org/10.1016/j.tics.2008.02.014

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867-872. https://doi.org/10.1016/j.jesp.2009.03.009

Parks, G. S. (2015). Judicial recusal: Cognitive biases and racial stereotyping. *New York University Journal of Legislation and Public Policy*, 18(3), 681-698.

Prat-Ortega, G., & de la Rocha, J. (2018). Selective attention: A plausible mechanism underlying confirmation bias. *Current Biology*, *28*(19), R1151-R1154. https://doi.org/10.1016/j.cub.2018.08.024

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, *42*(1), 185-227. https://doi.org/10.1080/00273170701341316

R. v. Bennett (2011), N.J. No. 302

R. v. Clark [No 1] (2008), NSWSC 820

R. v. Clark [No 2] (2008), NSWSC 821

R. v. Clark [No 3] (2008), NSWSC 795

R. v. Demeter (1975), 6 O.R. (2d) 83

R. v. Goldman (1980), 1 S.C.R. 976

R. v. Fegan (1993), 62 O.A.C. 146 (C.A.)

R. v. Madrid (B.C.C.A.) (1994), B.C.J. No. 1786

R. v. Randall and Weir (1983), N.S.J. No. 438

R. v. Strano (2001), 80 C.R.R. (2d) 93 (Ont. H.C.J.)

Rachlinski, J. J., Johnson, S. L., Wistrich, A. J., & Guthrie, C. (2008). Does unconscious racial bias affect trial judges? *Notre Dame Law Review*, *84*, 1195-1246.

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, *8*(3), 338-342. https://doi.org/10.1006/ccog.1999.0386

Reber, R., Wurtz, P., & Zimmermann, T. D. (2004). Exploring "fringe" consciousness: The subjective experience of perceptual fluency and its objective bases. *Consciousness and Cognition*, *13*(1), 47-60. https://doi.org/10.1016/S1053-8100(03)00049-7

Rhodes, G., Halberstadt, J., & Brajkovich, G. (2001). Generalization of mere exposure effects to averaged composite faces. *Social Cognition*, *19*(1), 57-70. https://doi.org/10.1521/soco.19.1.57.18961

Risinger, D. M., Saks, M. J., Thompson, W. C., & Rosenthal, R. (2002). The Daubert/Kumho implications of observer effects in forensic science: Hidden problems of expectation and suggestion. *California Law Review*, *90*, 1-56. https://doi.org/10.15779/Z38GM7K

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. New York: Holt, Rinehart & Winston

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32,* 880-892. https://doi.org/10.1037/0022-3514.32.5.880

Russo, J. E., Medvec, V. H., & Meloy, M. G. (1996). The distortion of information during decisions. *Organizational Behavior and Human Decision Processes*, *66*(1), 102-110. https://doi.org/10.1006/obhd.1996.0041

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110*(4), 474-494. https://doi.org/10.1037/0096-3445.110.4.474

Samuel, A.G. (1996). Phoneme restoration. *Language and Cognitive Processes, 11*, 647–653. https://doi.org/10.1080/016909696387051

Schmittat, S. M., Englich, B., Sautner, L., & Velten, P. (2021). Alternative stories and the decision to prosecute: An applied approach against confirmation bias in criminal prosecution. *Psychology, Crime & Law*, *28*(6), 608-635. https://doi.org/10.1080/1068316X.2021.1941013

Schneider, B. A., Avivi-Reich, M., & Mozuraitis, M. (2015). A cautionary note on the use of the Analysis of Covariance (ANCOVA) in classification designs with and without within-subject factors. *Frontiers in Psychology*, *6*, 474. https://doi.org/10.3389/fpsyg.2015.00474

Schulz-Hardt, S., Frey, D., Lüthgens, C., & Moscovici, S. (2000). Biased information search in group decision making. *Journal of Personality and Social Psychology*, *78*(4), 655-669. https://doi.org/10.1037//0022-3514.78.4.655

Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, *14*(4), 332-348. https://doi.org/10.1207/s15327663jcp1404_2

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology*, *61*(2), 195-202. https://doi.org/10.1037/0022-3514.61.2.195

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, *134*(2), 207-222. https://doi.org/10.1037/0033-2909.134.2.207

Sher, A. E., & Owens, E. (1974). Consonant confusions associated with hearing loss above 2000 Hz. *Journal of Speech, Language, and Hearing Research*, *17*(4), 669-681. https://doi.org/10.1044/jshr.1704.669

Sherman, S. J., Zehner, K. S., Johnson, J., & Hirt, E. R. (1983). Social explanation: The role of timing, set, and recall on subjective likelihood estimates. *Journal of Personality and Social Psychology, 44,* 1127-1143. https://doi.org/10.1037/0022-3514.44.6.1127

Sherrin, C. (2015). Earwitness evidence: The reliability of voice identifications. *Osgoode Hall Law Journal, 52,* 819-862.

Skov, R. B., & Sherman, S. J. (1986). Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. *Journal of Experimental Social Psychology*, *22*(2), 93-121. https://doi.org/10.1016/0022-1031(86)90031-4

Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(4), 544-551. https://doi.org/10.1037/0096-1523.3.4.544

Snyder, M., & Campbell, B. (1980). Testing hypotheses about other people: The role of the hypothesis. *Personality and Social Psychology Bulletin*, *6*(3), 421-426. https://doi.org/10.1177/014616728063015

Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology*, *36*(11), 1202-1212. https://doi.org/10.1037/0022-3514.36.11.1202

State v. Lavers, 814 P.2d 333 (Ariz., 1991).

State v. Trask, Case No. CF-2007-184, Okla. Dist. Ct. (2007–2009)

Statistics Canada (2021). *Table 35-10-0031-01 Adult criminal courts, guilty cases by most serious sentence* [Data table]. https://doi.org/10.25318/3510003101-eng

Steblay, N., Hosch, H. M., Culhane, S. E., & McWethy, A. (2006). The impact on juror verdicts of judicial instruction to disregard inadmissible evidence: A meta-analysis. *Law and Human Behavior*, *30*(4), 469-492. https://doi.org/10.1007/s10979-006-9039-7

Talluri, B. C., Urai, A. E., Tsetsos, K., Usher, M., & Donner, T. H. (2018). Confirmation bias through selective overweighting of choice-consistent evidence. *Current Biology*, *28*(19), 3128-3135. https://doi.org/10.1016/j.cub.2018.07.052

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207-232. https://doi.org/10.1016/0010-0285(73)90033-9

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*(4940), 301-306. https://doi.org/10.1126/science.2296719

Undorf, M., Zimdahl, M. F., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects of stimulus size on judgments of learning. *Journal of Memory and Language*, *92*, 293-304. https://doi.org/10.1016/j.jml.2016.07.003

U.S. v. Harrell, 788 F.2d 1524 (C.A. Fla., 1986).

U.S. v. Stephens, 202 F.Supp. 2d 1361 (N.D. Ga., 2002).

Vaidya, C. J., Gabrieli, J. D., Keane, M. M., Monti, L. A., Gutierrez-Rivas, H., & Zarella, M. M. (1997). Evidence for multiple mechanisms of conceptual priming on implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(6), 1324-1343. https://doi.org/10.1037/0278-7393.23.6.1324

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*, 392–393. https://doi.org/10.1126/science.167.3917

Warren, R. M. (1984). Perceptual restoration of obliterated sounds. *Psychological Bulletin*, *96*(2), 371-383. https://doi.org/10.1037/0033-2909.96.2.371

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*(3), 129-140. https://doi.org/10.1080/17470216008416717

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*(3), 273-281. https://doi.org/10.1080/14640746808400161

Webster, D. M., Richter, L., & Kruglanski, A. W. (1996). On leaping to conclusions when feeling tired: Mental fatigue effects on impressional primacy. *Journal of Experimental Social Psychology*, *32*(2), 181-195. https://doi.org/10.1006/jesp.1996.0009

Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., & Albarracín, D. (2016). From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words. *Psychological Bulletin, 142*(5), 472–497. https://doi.org/10.1037/bul0000030

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, *6*(3), 291-298. https://doi.org/10.1177/1745691611406923

Whittlesea, B. W. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(6), 1235-1253. https://doi.org/10.1037/0278-7393.19.6.1235

Winkielman, P., Schwarz, N., & Belli, R. F. (1998). The role of ease of retrieval and attribution in memory judgments: Judging your memory as worse despite recalling more events. *Psychological Science, 9*(2), 124-126. https://doi.org/10.1111/1467-9280.00022

Winkielman, P., Schwarz, N., Fazendeiro, T., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K.C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 189-217). Mahwah, NJ: Erlbaum.

Yarmey, A. D. (1995). Earwitness speaker identification. *Psychology, Public Policy, and Law*, *1*(4), 792-816. https://doi.org/10.1037/1076-8971.1.4.792

Zadny, J., & Gerard, H. B. (1974). Attributed intentions and informational selectivity. *Journal of Experimental Social Psychology*, *10*(1), 34-52. https://doi.org/10.1016/0022-1031(74)90055-9

Zaleskiewicz, T., & Gasiorowska, A. (2021). Evaluating experts may serve psychological needs: Self-esteem, bias blind spot, and processing fluency explain confirmation effect in assessing financial advisors' authority. *Journal of Experimental Psychology: Applied*, *27*(1), 27-45. https://doi.org/10.1037/xap0000308

Appendix A.

Pilot 1 Stimuli

Note: Italicized items were included in Studies 1-3; Items with an asterisk were included in Studies 1 and 3 only.

Target Statements [expected incriminating misinterpretation in brackets]:

1. *I kept my wife [knife] hidden from him.*
2. *Next thing I knew, there was mud [blood] everywhere.*
3. *She didn't know it, but I was charmed [armed].*
4. *I knew I needed to call [kill] him.*
5. *She was scared of what it [I] was going to do to her.*
6. *I had to get rid of that hobby [body].*
7. *I packed [snapped] because I was so pissed at him.*
8. *After everything he'd done to me, it was time for a pact [payback].**
9. *I didn't feel bad when I saw what it'd [I'd] done to him.*
10. *I looked at him and told him he was red [dead].**
11. *He was becoming a real boredom [problem] to have around.*
12. I looked in his eyes and I could tell he was lying [dying].
13. He asked me to stop working [hurting] him.
14. *He stopped leaving [breathing], and I knew it was over.*
15. *I never mean to hinder [injure] him.*
16. *I bet he never expected me to hug [mug] him.*
17. *I borrowed [buried] all of his stuff, even his clothes.*
18. He never even heard me humming [coming].
19. *I passed [bashed] him while he wasn't looking.*
20. I left him there to dry [die] off.
21. *I shocked [shot] him over and over again.**

22. I went over and grabbed [stabbed] him.

23. He wouldn't listen to me so I started poking [choking] him.

24. I picked her up and placed [raped] her on the bed.

25. He saw me pull out my gum [gun].

26. I was drunk and thriving [driving] when they found me.

27. I remember the day that I sold [stole] her car.

28. I wrangled [strangled] them late that night.

29. I never meant to skirt [hurt] her.

30. I woke [broke] in their house in the middle of the night.

31. I took all her honey [money] while she wasn't looking.

32. I kissed [slit] her neck when she turned away from me.

33. I was eating [speeding] in my car when I got in the accident.

34. I wound [drowned] them behind my house.

35. I chilled [killed] them in my house.

36. She didn't know that I locked [robbed] the bank that night.

37. I darted [started] the fire in the house right in time.

38. I told him I could give him a deal on some trucks [drugs].

39. I'll admit that I loved [mugged] her.

40. I chugged [drugged] her drink while she wasn't looking.

41. I checked to make sure he wasn't grieving [breathing].

*42. I soaked [choked] them in the water.**

43. He begged me to stop nagging [stabbing] him.

44. I dunked [dumped] both of them in the lake.

45. As she turned to walk away, I grabbed her note [throat].

46. She wouldn't stop crying so I stroked [broke] her arm.

47. I calmed [bombed] down everyone in that place.

48. I went to go do the climb [crime] with him.

Filler statements:

1. I met him when we were in high school.
2. I never thought he was as smart as me.
3. I had to pay whenever we went out for drinks.
4. I would always ask him to help me with chores.
5. We spent a lot of time together.
6. We had long talks about our lives.
7. He came over to help me with some yard work.
8. He knew a lot about me.
9. Sometimes he ditched me to hang out with other people.
10. We argued every now and then.
11. I asked him if he was seeing anyone at the moment.
12. I asked him to water the grass and the flowers.
13. We were laying mulch together outside.
14. I was annoyed that he kept talking about his other friends.
15. I would always give him money when he asked.
16. I went inside to cool off.
17. Sometimes, I felt like he took advantage of me.
18. It started to get dark out and I was tired.
19. He went to leave and I started yelling at him.
20. He told me he was going to give me the money he owed me.
21. We liked to go out to bars and go dancing.
22. I really don't have that many good friends.
23. We used to stay up all night long talking about our lives.
24. We liked to play sports together, like baseball and basketball.
25. We always helped each other out.
26. We both liked to work on cars together.

27. I remember I went to the grocery store earlier that day.
28. I had just gotten over a bad cold earlier that week.
29. I told him I would go inside and make some coffee.
30. My phone kept ringing and I couldn't find it.
31. It was a really hot day outside.
32. I told him I would make some burgers for dinner.
33. I own a furniture moving company.
34. I knew I needed to mow the lawn that day.
35. He brought me a new sprinkler for my yard.
36. I was watching tv when he called me.

Appendix B.

Demographic Questionnaire

1. Please indicate your gender:
    - Female
    - Male
    - Other: __________

2. How old are you? ______________

3. What is your ethnicity? __________________

4. What is your first language (e.g., English)? ________________________

5. Approximately how long have you been speaking English (in years)? ________________

6. How well (fluently) do you speak English?
    - Very
    - Moderately
    - Poorly

7. How well (fluently) do you write English?
    - Very
    - Moderately
    - Poorly

8. How often do you use English? (circle one)
    - Everyday
    - Every few days
    - Once a week
    - Once a month
    - Once a year

9. What is the highest level of education you have completed?
    - None
    - GED
    - High School Diploma
    - Associates
    - Bachelors
    - Masters
    - Doctorate (including Ph.D., M.D., and J.D.)

- Other: ________________

10. Do you have any hearing impairments?
    - Yes
    - No

Appendix C.

The Proportion of Participants who Made Incriminating Misinterpretations and Accurate Interpretations Across Degradation Levels and Contextual Information Conditions in Pilot 1

Statement	Incriminating Misinterpretations		Accurate Interpretations	
Target 1	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.13	0.60	0.63
800 hz	0.00	0.00	0.88	0.89
1000 hz	0.00	0.00	0.88	1.00
1200 hz	0.00	0.00	0.92	1.00
1400 hz	0.00	0.00	0.82	1.00
1600 hz	0.00	0.00	1.00	0.92
Target 2	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.27	0.40	0.18
800 hz	0.13	0.22	0.75	0.44
1000 hz	0.00	0.25	0.88	0.75
1200 hz	0.00	0.11	0.77	0.67
1400 hz	0.00	0.00	0.55	0.80
1600 hz	0.00	0.08	0.92	0.92
Target 3	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.00	0.00
800 hz	0.00	0.00	0.00	0.11
1000 hz	0.00	0.00	0.57	0.43
1200 hz	0.00	0.00	0.57	0.67
1400 hz	0.00	0.11	0.91	0.78
1600 hz	0.00	0.00	0.89	1.00

Target 4	No Context	Criminal	No Context	Criminal
600 hz	0.00	0.00	0.67	0.78
800 hz	0.00	0.00	1.00	1.00
1000 hz	0.00	0.00	1.00	1.00
1200 hz	0.00	0.00	1.00	1.00
1400 hz	0.00	0.00	1.00	1.00
1600 hz	0.00	0.00	1.00	1.00
Target 5	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.45	0.56
800 hz	0.00	0.00	0.86	0.78
1000 hz	0.00	0.13	1.00	0.75
1200 hz	0.00	0.00	0.92	0.78
1400 hz	0.00	0.00	0.91	1.00
1600 hz	0.00	0.00	0.92	1.00
Target 6	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.08	0.17	0.33
800 hz	0.11	0.11	0.89	0.56
1000 hz	0.00	0.00	0.86	0.71
1200 hz	0.00	0.00	0.75	0.83
1400 hz	0.00	0.00	0.83	0.89
1600 hz	0.00	0.00	0.89	0.78
Target 7	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.13	0.00	0.25	0.14
800 hz	0.00	0.00	0.62	0.60
1000 hz	0.00	0.00	0.91	0.70
1200 hz	0.00	0.08	0.77	0.54
1400 hz	0.00	0.00	1.00	0.73
1600 hz	0.00	0.00	0.75	1.00
Target 8	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.20	0.00
800 hz	0.08	0.08	0.08	0.15
1000 hz	0.00	0.00	0.50	0.56
1200 hz	0.00	0.00	0.71	0.78
1400 hz	0.00	0.00	1.00	1.00
1600 hz	0.00	0.00	0.83	0.90

Target 9	No Context	Criminal	No Context	Criminal
600 hz	0.11	0.22	0.00	0.11
800 hz	0.29	0.50	0.11	0.13
1000 hz	0.63	0.50	0.25	0.00
1200 hz	0.27	0.67	0.55	0.22
1400 hz	0.22	0.56	0.67	0.11
1600 hz	0.31	0.25	0.62	0.75
Target 10	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.00	0.17
800 hz	0.00	0.00	0.15	0.00
1000 hz	0.00	0.00	0.40	0.20
1200 hz	0.00	0.00	0.50	0.75
1400 hz	0.00	0.00	0.60	0.67
1600 hz	0.00	0.00	0.63	0.56
Target 11	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.13	0.17	0.00	0.00
800 hz	0.08	0.00	0.08	0.00
1000 hz	0.00	0.11	0.67	0.56
1200 hz	0.00	0.00	0.92	1.00
1400 hz	0.00	0.10	0.89	0.60
1600 hz	0.00	0.00	0.88	0.88
Target 12	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.56	0.33
800 hz	0.00	0.00	0.83	0.86
1000 hz	0.00	0.00	1.00	1.00
1200 hz	0.00	0.00	0.92	1.00
1400 hz	0.00	0.00	0.75	0.78
1600 hz	0.00	0.00	0.83	1.00
Target 13	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.08	0.27
800 hz	0.09	0.11	0.36	0.33
1000 hz	0.00	0.00	0.88	0.78
1200 hz	0.00	0.13	0.78	0.75
1400 hz	0.00	0.00	0.92	0.89
1600 hz	0.00	0.00	0.82	0.90

Target 14	No Context	Criminal	No Context	Criminal
600 hz	0.00	0.00	0.29	0.13
800 hz	0.38	0.17	0.13	0.50
1000 hz	0.38	0.50	0.15	0.00
1200 hz	0.09	0.44	0.27	0.11
1400 hz	0.08	0.54	0.08	0.15
1600 hz	0.30	0.36	0.30	0.27
Target 15	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.00	0.00
800 hz	0.00	0.00	0.11	0.00
1000 hz	0.00	0.00	0.08	0.09
1200 hz	0.11	0.11	0.11	0.22
1400 hz	0.00	0.44	0.50	0.22
1600 hz	0.25	0.00	0.75	0.57
Target 16	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.30	0.17	0.10
800 hz	0.00	0.00	0.67	0.63
1000 hz	0.13	0.00	0.75	1.00
1200 hz	0.00	0.11	1.00	0.78
1400 hz	0.00	0.00	1.00	1.00
1600 hz	0.00	0.00	1.00	1.00
Target 17	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.15	0.20	0.46	0.10
800 hz	0.00	0.11	0.78	0.89
1000 hz	0.08	0.08	0.92	0.92
1200 hz	0.00	0.00	1.00	1.00
1400 hz	0.00	0.00	1.00	1.00
1600 hz	0.00	0.00	1.00	1.00
Target 18	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.11	0.22	0.22	0.11
800 hz	0.18	0.58	0.55	0.08
1000 hz	0.11	0.75	0.44	0.00
1200 hz	0.71	0.43	0.29	0.43
1400 hz	0.25	0.67	0.75	0.33
1600 hz	0.18	0.33	0.82	0.56

Target 19	No Context	Criminal	No Context	Criminal
600 hz	0.09	0.22	0.18	0.36
800 hz	0.18	0.08	0.64	0.85
1000 hz	0.09	0.09	0.64	0.64
1200 hz	0.00	0.00	0.88	1.00
1400 hz	0.00	0.00	0.89	1.00
1600 hz	0.00	0.11	0.92	0.78
Target 20	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.56	0.56
800 hz	0.00	0.00	1.00	1.00
1000 hz	0.00	0.00	1.00	0.78
1200 hz	0.00	0.00	1.00	1.00
1400 hz	0.00	0.00	1.00	1.00
1600 hz	0.00	0.00	1.00	1.00
Target 21	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.15	0.38	0.00	0.00
800 hz	0.30	0.67	0.00	0.00
1000 hz	0.63	0.56	0.00	0.00
1200 hz	0.75	0.33	0.00	0.17
1400 hz	0.00	0.22	0.75	0.67
1600 hz	0.18	0.10	0.55	0.80
Target 22	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.83	0.83
800 hz	0.00	0.00	1.00	0.00
1000 hz	0.00	0.00	0.85	0.90
1200 hz	0.00	0.00	0.89	1.00
1400 hz	0.00	0.00	1.00	1.00
1600 hz	0.00	0.00	1.00	0.53
Target 23	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.44	0.44
800 hz	0.08	0.09	0.92	0.91
1000 hz	0.00	0.00	1.00	0.88
1200 hz	0.00	0.00	1.00	1.00
1400 hz	0.13	0.00	0.88	1.00
1600 hz	0.00	0.00	1.00	1.00

Target 24	No Context	Criminal	No Context	Criminal
600 hz	0.00	0.13	0.44	0.00
800 hz	0.00	0.00	0.63	1.00
1000 hz	0.00	0.00	0.77	0.80
1200 hz	0.00	0.00	0.89	0.88
1400 hz	0.00	0.00	0.92	1.00
1600 hz	0.00	0.00	1.00	0.90
Target 25	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.50	0.33	0.25	0.00
800 hz	0.46	0.27	0.00	0.00
1000 hz	0.45	0.70	0.27	0.10
1200 hz	0.38	0.69	0.54	0.15
1400 hz	0.10	0.60	0.90	0.40
1600 hz	0.25	0.56	0.63	0.44
Target 26	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.23	0.00	0.00	0.09
800 hz	0.22	0.44	0.22	0.00
1000 hz	0.29	0.42	0.36	0.33
1200 hz	0.67	0.30	0.22	0.60
1400 hz	0.29	0.25	0.71	0.33
1600 hz	0.25	0.57	0.75	0.29
Target 27	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.18	0.11	0.09	0.33
800 hz	0.23	0.25	0.69	0.58
1000 hz	0.00	0.36	1.00	0.36
1200 hz	0.00	0.22	0.63	0.67
1400 hz	0.25	0.00	0.75	0.83
1600 hz	0.08	0.11	0.92	0.78
Target 28	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.00	0.08
800 hz	0.00	0.27	0.70	0.27
1000 hz	0.00	0.00	0.70	1.00
1200 hz	0.00	0.00	0.75	1.00
1400 hz	0.00	0.00	0.92	0.89
1600 hz	0.00	0.00	0.78	0.78

Target 29	No Context	Criminal	No Context	Criminal
600 hz	0.10	0.11	0.00	0.00
800 hz	0.00	0.00	0.00	0.00
1000 hz	0.40	0.20	0.20	0.10
1200 hz	0.50	0.11	0.25	0.56
1400 hz	0.25	0.33	0.13	0.17
1600 hz	0.08	0.22	0.42	0.33
Target 30	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.14	0.13	0.14
800 hz	0.00	0.10	0.62	0.50
1000 hz	0.00	0.00	0.45	0.90
1200 hz	0.00	0.08	0.85	0.62
1400 hz	0.00	0.00	0.70	0.55
1600 hz	0.00	0.11	0.88	0.56
Target 31	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.29	0.00	0.00	0.00
800 hz	0.13	0.33	0.38	0.00
1000 hz	0.15	0.20	0.69	0.50
1200 hz	0.10	0.33	0.70	0.56
1400 hz	0.08	0.17	0.85	0.75
1600 hz	0.10	0.00	0.90	1.00
Target 32	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.14	0.13	0.29
800 hz	0.00	0.20	0.54	0.10
1000 hz	0.09	0.11	0.55	0.44
1200 hz	0.15	0.08	0.62	0.69
1400 hz	0.00	0.00	0.55	0.55
1600 hz	0.00	0.11	1.00	0.78
Target 33	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.00	0.25
800 hz	0.00	0.00	0.29	0.38
1000 hz	0.00	0.00	0.50	0.17
1200 hz	0.00	0.00	0.50	0.33
1400 hz	0.00	0.00	0.56	0.89
1600 hz	0.00	0.00	1.00	1.00

Target 34	No Context	Criminal	No Context	Criminal
600 hz	0.00	0.20	0.00	0.00
800 hz	0.00	0.13	0.33	0.25
1000 hz	0.00	0.00	0.29	0.29
1200 hz	0.00	0.00	0.71	0.33
1400 hz	0.00	0.00	0.33	0.22
1600 hz	0.00	0.00	0.44	0.11
Target 35	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.14	0.44	0.00	0.00
800 hz	0.63	0.33	0.00	0.00
1000 hz	0.77	0.40	0.08	0.00
1200 hz	0.64	0.40	0.09	0.20
1400 hz	0.15	0.31	0.38	0.46
1600 hz	0.20	0.27	0.60	0.36
Target 36	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.18	0.00
800 hz	0.00	0.00	0.54	0.46
1000 hz	0.00	0.00	0.50	0.55
1200 hz	0.00	0.00	0.63	0.89
1400 hz	0.00	0.33	0.75	0.17
1600 hz	0.08	0.00	0.67	0.78
Target 37	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.00	0.00
800 hz	0.29	0.11	0.00	0.00
1000 hz	0.23	0.38	0.00	0.00
1200 hz	0.56	0.45	0.22	0.00
1400 hz	0.43	0.44	0.00	0.00
1600 hz	0.75	0.50	0.00	0.17
Target 38	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.20	0.00	0.00
800 hz	0.00	0.00	0.27	0.33
1000 hz	0.00	0.07	0.86	0.64
1200 hz	0.10	0.00	0.80	0.67
1400 hz	0.00	0.11	0.63	0.78
1600 hz	0.00	0.00	1.00	1.00

Target 39	No Context	Criminal	No Context	Criminal
600 hz	0.00	0.00	0.13	0.17
800 hz	0.00	0.00	0.38	0.20
1000 hz	0.00	0.00	0.67	0.78
1200 hz	0.00	0.00	0.92	1.00
1400 hz	0.00	0.00	1.00	0.44
1600 hz	0.00	0.00	1.00	0.86
Target 40	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.00	0.00
800 hz	0.00	0.00	0.00	0.00
1000 hz	0.17	0.14	0.17	0.14
1200 hz	0.14	0.00	0.43	0.33
1400 hz	0.09	0.44	0.73	0.22
1600 hz	0.11	0.00	0.67	0.67
Target 41	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.10	0.11	0.00
800 hz	0.00	0.13	0.00	0.00
1000 hz	0.44	0.50	0.22	0.13
1200 hz	0.55	0.50	0.18	0.10
1400 hz	0.36	0.50	0.36	0.30
1600 hz	0.54	0.79	0.38	0.21
Target 42	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.22	0.00	0.00
800 hz	0.00	0.00	0.13	0.00
1000 hz	0.00	0.00	0.38	0.30
1200 hz	0.00	0.00	0.00	0.54
1400 hz	0.00	0.08	0.54	0.92
1600 hz	0.00	0.00	0.90	0.64
Target 43	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.08	0.10	0.00	0.00
800 hz	0.11	0.11	0.00	0.00
1000 hz	0.50	0.75	0.00	0.08
1200 hz	0.33	0.50	0.33	0.38
1400 hz	0.43	0.71	0.43	0.14
1600 hz	0.71	0.50	0.29	0.50

Target 44	No Context	Criminal	No Context	Criminal
600 hz	0.08	0.33	0.00	0.00
800 hz	0.33	0.44	0.11	0.00
1000 hz	0.46	0.67	0.08	0.08
1200 hz	0.78	0.44	0.11	0.00
1400 hz	0.71	0.50	0.00	0.00
1600 hz	0.50	0.43	0.00	0.00
Target 45	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.11	0.20
800 hz	0.00	0.00	0.56	0.33
1000 hz	0.00	0.17	0.88	0.50
1200 hz	0.00	0.00	0.92	0.89
1400 hz	0.00	0.00	0.78	0.89
1600 hz	0.00	0.00	0.85	1.00
Target 46	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.10	0.00	0.00
800 hz	0.17	0.25	0.00	0.00
1000 hz	0.00	0.14	0.14	0.29
1200 hz	0.00	0.00	0.57	0.33
1400 hz	0.00	0.00	0.91	0.56
1600 hz	0.00	0.00	0.67	0.78
Target 47	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.18	0.33
800 hz	0.00	0.00	0.85	0.77
1000 hz	0.00	0.00	1.00	0.70
1200 hz	0.00	0.00	1.00	1.00
1400 hz	0.00	0.00	1.00	1.00
1600 hz	0.00	0.00	1.00	1.00
Target 48	**No Context**	**Criminal**	**No Context**	**Criminal**
600 hz	0.00	0.00	0.00	0.17
800 hz	0.08	0.00	0.15	0.00
1000 hz	0.00	0.30	0.27	0.00
1200 hz	0.23	0.31	0.54	0.46
1400 hz	0.18	0.11	0.36	0.33
1600 hz	0.38	0.44	0.25	0.44

Appendix D.

Additional Target Statements Generated for Pilot 2

Note: Italicized items were included in Studies 1-3; Items with an asterisk were included in Studies 1 and 3 only.

49. I stopped [stabbed] him over and over again.

*50. I had to snap [stab] them before it was too late.**

51. I made sure he was buying [dying].

52. He begged me not to work [hurt] him.

53. I left them there to dry [die].

54. I ranked [raped] her body in front of all those guys.

55. I hugged [mugged] her so I could get some money.

56. I told him we needed to fill [kill] them.

57. I shut [shot] all of those guys up.

*58. I got scared when I saw what it'd [I'd] done to him.**

59. I dropped my gum [gun] on the floor.

60. I couldn't take it anymore, so I stopped [stabbed] him.

61. I carried [buried] her outside my house.

Appendix E.

The Proportion of Participants who Made Incriminating Misinterpretations and Accurate Interpretations Across Degradation Levels and Contextual Information Conditions as well as the Average Percentage of the Statements Identified for Each Degradation Level in Pilot 2

Statement	Incriminating Misinterpretations		Accurate Interpretations		% of statement identified
Target 1	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.00	0.40	0.11	30.16
800 hz	0.09	0.10	0.18	0.70	37.34
1000 hz	0.00	0.13	0.40	0.47	48.58
1200 hz	0.00	0.25	0.45	0.25	51.19
1400 hz	0.00	0.00	0.75	0.67	81.51
1600 hz	0.00	0.00	0.90	0.92	92.09
Target 2	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.09	0.50	0.27	0.25	51.57
800 hz	0.00	0.18	0.76	0.64	72.63
1000 hz	0.00	0.08	0.70	0.85	82.26
1200 hz	0.17	0.10	0.67	0.90	90.42
1400 hz	0.00	0.00	0.93	1.00	94.59
1600 hz	0.00	0.00	0.87	1.00	90.97
Target 3	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.00	0.00	0.00	52.64
800 hz	0.00	0.17	0.00	0.00	57.3
1000 hz	0.00	0.00	0.41	0.45	75.94
1200 hz	0.00	0.00	0.40	0.54	82.7
1400 hz	0.00	0.00	0.75	0.90	90.32
1600 hz	0.00	0.00	0.86	0.93	86.67
Target 4	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.27	0.22	0.07	0.00	24.45
800 hz	0.00	0.17	0.18	0.00	34.04
1000 hz	0.06	0.00	0.35	0.55	48.23
1200 hz	0.00	0.00	0.40	0.92	62.58
1400 hz	0.00	0.00	0.73	0.80	73.9
1600 hz	0.00	0.00	0.87	0.93	84.29

Target 6	No Context	Criminal	No Context	Criminal	
600 hz	0.00	0.11	0.40	0.22	73.33
800 hz	0.00	0.00	0.73	0.83	79.17
1000 hz	0.00	0.00	0.82	0.73	86.06
1200 hz	0.00	0.00	0.80	1.00	90.92
1400 hz	0.00	0.00	0.92	0.90	93.44
1600 hz	0.00	0.00	0.87	0.93	93.34
Target 7	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.08	0.00	0.15	74.1
800 hz	0.00	0.00	0.75	0.70	92.78
1000 hz	0.07	0.00	0.67	0.93	91.49
1200 hz	0.07	0.00	0.60	0.56	92.72
1400 hz	0.09	0.00	0.82	0.92	94.46
1600 hz	0.00	0.00	0.82	1.00	94.21
Target 8*	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.07	0.00	0.00	83.89
800 hz	0.00	0.33	0.07	0.11	68.61
1000 hz	0.00	0.17	0.50	0.50	73.61
1200 hz	0.06	0.00	0.35	0.73	74.18
1400 hz	0.10	0.00	0.50	0.85	84.17
1600 hz	0.08	0.20	0.42	0.60	83.75
Target 9	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.18	0.42	0.00	0.08	51.74
800 hz	0.29	0.36	0.00	0.00	67.52
1000 hz	0.20	0.69	0.20	0.15	73.98
1200 hz	0.50	0.50	0.17	0.20	84.10
1400 hz	0.27	0.67	0.47	0.33	82.78
1600 hz	0.53	0.22	0.33	0.78	88.98
Target 10*	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.08	0.00	0.08	23.00
800 hz	0.08	0.10	0.08	0.50	26.67
1000 hz	0.00	0.00	0.47	0.60	61.34
1200 hz	0.00	0.00	0.33	0.44	66.00
1400 hz	0.00	0.00	0.73	0.82	68.75
1600 hz	0.00	0.09	0.76	0.73	82.36
Target 11	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.20	0.08	0.00	0.00	58.04
800 hz	0.00	0.10	0.00	0.00	71.06
1000 hz	0.00	0.00	0.60	0.80	84.82
1200 hz	0.00	0.00	0.80	0.89	90.87
1400 hz	0.00	0.00	0.91	0.82	87.50
1600 hz	0.00	0.00	0.82	1.00	92.90

Target 14	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.13	0.09	0.25	0.09	38.41
800 hz	0.20	0.23	0.00	0.31	65.52
1000 hz	0.33	0.60	0.17	0.10	69.91
1200 hz	0.71	0.53	0.14	0.13	80.38
1400 hz	0.33	0.67	0.13	0.11	86.17
1600 hz	0.27	0.58	0.18	0.17	82.89
Target 15	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.00	0.00	0.10	20.69
800 hz	0.00	0.07	0.00	0.00	14.45
1000 hz	0.07	0.33	0.00	0.00	29.08
1200 hz	0.09	0.25	0.18	0.08	37.50
1400 hz	0.06	0.18	0.47	0.64	61.13
1600 hz	0.20	0.08	0.60	0.92	72.89
Target 16	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.42	0.45	0.25	48.15
800 hz	0.18	0.18	0.65	0.82	72.26
1000 hz	0.10	0.15	0.80	0.85	83.46
1200 hz	0.17	0.00	0.58	1.00	84.54
1400 hz	0.08	0.00	0.92	1.00	88.89
1600 hz	0.07	0.11	0.80	0.78	78.89
Target 18	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.36	0.47	0.29	0.20	60.00
800 hz	0.33	0.67	0.13	0.11	60.75
1000 hz	0.18	0.33	0.45	0.17	54.17
1200 hz	0.41	0.36	0.41	0.55	71.93
1400 hz	0.30	0.46	0.50	0.54	85.10
1600 hz	0.33	0.00	0.67	1.00	87.09
Target 19	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.07	0.13	0.33	0.33	54.76
800 hz	0.00	0.11	0.53	0.56	71.90
1000 hz	0.00	0.17	0.92	0.67	91.67
1200 hz	0.06	0.00	0.71	0.82	87.86
1400 hz	0.00	0.00	0.80	0.77	91.16
1600 hz	0.00	0.00	0.92	1.00	92.86
Target 21*	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.44	0.67	0.19	0.11	84.29
800 hz	0.36	0.75	0.36	0.00	77.98
1000 hz	0.47	0.64	0.29	0.27	84.82
1200 hz	0.10	0.46	0.60	0.54	82.86
1400 hz	0.17	0.10	0.75	0.90	82.74
1600 hz	0.07	0.00	0.87	1.00	84.76

Target 23	No Context	Criminal	No Context	Criminal	
600 hz	0.07	0.21	0.43	0.64	67.34
800 hz	0.07	0.00	0.73	1.00	84.67
1000 hz	0.00	0.08	0.82	0.92	79.17
1200 hz	0.00	0.00	0.88	1.00	85.62
1400 hz	0.00	0.00	0.90	1.00	92.85
1600 hz	0.00	0.00	0.92	1.00	93.92
Target 26	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.40	0.40	0.10	0.00	39.24
800 hz	0.50	0.20	0.14	0.27	68.89
1000 hz	0.47	0.67	0.27	0.22	85.68
1200 hz	0.45	0.33	0.45	0.50	87.50
1400 hz	0.18	0.09	0.65	0.73	94.16
1600 hz	0.20	0.31	0.60	0.69	91.07
Target 27	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.36	0.40	0.29	0.20	74.08
800 hz	0.20	0.44	0.60	0.44	75.06
1000 hz	0.09	0.42	0.64	0.50	81.95
1200 hz	0.24	0.18	0.65	0.73	87.32
1400 hz	0.10	0.00	0.70	0.92	91.07
1600 hz	0.08	0.10	0.83	0.90	91.49
Target 28	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.00	0.00	0.00	52.96
800 hz	0.09	0.08	0.45	0.67	67.36
1000 hz	0.00	0.00	0.76	0.91	84.63
1200 hz	0.00	0.00	0.80	0.69	83.97
1400 hz	0.00	0.00	0.83	0.90	86.25
1600 hz	0.00	0.07	0.87	0.80	87.22
Target 29	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.43	0.33	0.00	0.00	49.45
800 hz	0.47	0.67	0.00	0.11	65.56
1000 hz	0.09	0.42	0.36	0.08	70.14
1200 hz	0.12	0.09	0.47	0.55	87.57
1400 hz	0.10	0.38	0.30	0.62	86.38
1600 hz	0.08	0.20	0.42	0.60	86.09
Target 30	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.00	0.10	0.23	54.23
800 hz	0.08	0.10	0.50	0.70	73.03
1000 hz	0.00	0.07	0.67	0.73	76.97
1200 hz	0.00	0.00	0.67	1.00	81.01
1400 hz	0.00	0.17	0.73	0.33	79.93
1600 hz	0.06	0.09	0.65	0.55	80.22

Target 31	No Context	Criminal	No Context	Criminal	
600 hz	0.00	0.27	0.19	0.18	30.90
800 hz	0.00	0.31	0.10	0.00	43.92
1000 hz	0.08	0.40	0.67	0.60	74.45
1200 hz	0.07	0.27	0.93	0.67	77.04
1400 hz	0.20	0.33	0.67	0.67	80.99
1600 hz	0.09	0.58	0.91	0.42	80.56
Target 32	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.31	0.00	0.00	25.50
800 hz	0.17	0.20	0.00	0.00	34.09
1000 hz	0.00	0.20	0.47	0.53	77.34
1200 hz	0.13	0.22	0.40	0.33	71.67
1400 hz	0.00	0.00	0.91	0.92	90.00
1600 hz	0.12	0.09	0.76	0.91	93.67
Target 34	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.00	0.00	0.00	36.49
800 hz	0.00	0.17	0.09	0.25	65.98
1000 hz	0.00	0.00	0.18	0.09	77.85
1200 hz	0.00	0.00	0.40	0.38	79.36
1400 hz	0.08	0.00	0.08	0.40	75.84
1600 hz	0.00	0.00	0.43	0.40	78.89
Target 35	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.19	0.27	0.00	0.00	43.80
800 hz	0.60	0.77	0.00	0.08	59.55
1000 hz	0.42	0.90	0.00	0.10	66.67
1200 hz	0.60	0.53	0.13	0.33	77.22
1400 hz	0.47	0.67	0.27	0.22	84.08
1600 hz	0.27	0.33	0.45	0.25	82.64
Target 36	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.07	0.14	0.40	55.34
800 hz	0.00	0.00	0.47	0.56	64.11
1000 hz	0.09	0.00	0.45	0.82	70.84
1200 hz	0.00	0.09	0.82	0.91	89.15
1400 hz	0.00	0.08	0.80	0.85	91.46
1600 hz	0.00	0.00	0.67	0.90	91.00
Target 37	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.10	0.00	0.00	0.00	27.96
800 hz	0.07	0.13	0.07	0.00	44.67
1000 hz	0.47	0.56	0.07	0.11	75.89
1200 hz	0.18	0.42	0.36	0.17	66.67
1400 hz	0.12	0.09	0.18	0.45	75.83
1600 hz	0.20	0.46	0.10	0.00	57.08

Target 38	No Context	Criminal	No Context	Criminal	
600 hz	0.00	0.10	0.00	0.10	19.13
800 hz	0.07	0.20	0.40	0.07	63.06
1000 hz	0.20	0.11	0.60	0.56	77.32
1200 hz	0.00	0.25	0.82	0.50	73.96
1400 hz	0.00	0.09	0.65	0.73	86.06
1600 hz	0.10	0.08	0.80	0.92	90.03
Target 40	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.00	0.00	0.00	17.92
800 hz	0.09	0.00	0.00	0.00	29.69
1000 hz	0.12	0.09	0.06	0.09	45.59
1200 hz	0.10	0.08	0.10	0.08	53.85
1400 hz	0.17	0.10	0.25	0.70	73.02
1600 hz	0.29	0.13	0.43	0.53	78.33
Target 41	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.09	0.00	0.00	0.00	13.02
800 hz	0.07	0.27	0.00	0.00	27.80
1000 hz	0.30	0.67	0.20	0.08	51.06
1200 hz	0.25	0.80	0.33	0.00	69.90
1400 hz	0.53	0.67	0.20	0.27	79.17
1600 hz	0.33	0.78	0.53	0.22	84.03
Target 42*	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.09	0.06	0.00	60.37
800 hz	0.00	0.08	0.20	0.31	67.18
1000 hz	0.00	0.00	0.33	0.80	84.17
1200 hz	0.00	0.07	0.79	0.40	85.56
1400 hz	0.07	0.11	0.67	0.44	87.04
1600 hz	0.00	0.00	0.91	0.67	89.59
Target 44	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.17	0.70	0.17	0.10	59.07
800 hz	0.50	0.47	0.29	0.40	81.07
1000 hz	0.53	0.67	0.33	0.22	84.59
1200 hz	0.27	0.50	0.64	0.42	84.90
1400 hz	0.24	0.36	0.59	0.55	88.39
1600 hz	0.40	0.08	0.50	0.77	91.06
Target 46	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.00	0.00	0.00	39.51
800 hz	0.09	0.17	0.18	0.08	50.47
1000 hz	0.00	0.00	0.47	0.45	71.51
1200 hz	0.00	0.00	0.70	0.77	89.79
1400 hz	0.00	0.10	0.83	0.70	91.20
1600 hz	0.00	0.00	0.80	1.00	91.11

Target 48	No Context	Criminal	No Context	Criminal	
600 hz	0.00	0.00	0.00	0.00	18.33
800 hz	0.00	0.00	0.00	0.20	32.60
1000 hz	0.07	0.27	0.27	0.27	49.26
1200 hz	0.20	0.67	0.33	0.22	61.85
1400 hz	0.00	0.42	0.45	0.42	60.19
1600 hz	0.24	0.55	0.53	0.36	69.85
Target 49	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.33	0.42	0.08	0.00	49.41
800 hz	0.24	0.55	0.47	0.00	74.37
1000 hz	0.30	0.62	0.40	0.38	85.06
1200 hz	0.42	0.50	0.42	0.40	85.00
1400 hz	0.67	0.33	0.27	0.60	89.05
1600 hz	0.53	0.67	0.40	0.22	89.69
Target 50*	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.25	0.64	0.33	69.17
800 hz	0.12	0.36	0.47	0.55	80.67
1000 hz	0.10	0.08	0.30	0.77	85.43
1200 hz	0.17	0.10	0.58	0.90	91.84
1400 hz	0.07	0.07	0.86	0.80	89.00
1600 hz	0.07	0.11	0.87	0.89	91.00
Target 51	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.11	0.88	0.67	87.39
800 hz	0.00	0.00	0.17	0.00	40.98
1000 hz	0.00	0.00	0.65	0.27	55.92
1200 hz	0.00	0.08	0.60	0.62	75.00
1400 hz					
1600 hz	0.00	0.07	0.93	0.93	96.11
Target 52	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.55	0.30	0.09	0.10	53.34
800 hz	0.50	0.50	0.07	0.36	62.39
1000 hz	0.07	0.22	0.93	0.78	72.86
1200 hz	0.00	0.08	1.00	0.92	80.36
1400 hz	0.00	0.00	1.00	1.00	82.84
1600 hz	0.00	0.00	0.90	1.00	80.88
Target 53	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.50	0.33	0.29	0.53	72.78
800 hz	0.33	0.67	0.60	0.33	84.08
1000 hz	0.09	0.08	0.82	0.67	80.56
1200 hz	0.00	0.09	0.88	0.91	89.08
1400 hz	0.00	0.08	0.90	0.85	89.68
1600 hz	0.00	0.00	0.75	1.00	90.00

Target 54	No Context	Criminal	No Context	Criminal	
600 hz	0.00	0.09	0.06	0.00	45.80
800 hz	0.20	0.31	0.00	0.00	69.58
1000 hz	0.42	0.40	0.00	0.30	84.92
1200 hz	0.36	0.33	0.43	0.60	87.00
1400 hz	0.20	0.33	0.60	0.56	93.78
1600 hz	0.27	0.25	0.45	0.67	85.42
Target 55	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.09	0.18	0.27	50.57
800 hz	0.00	0.08	0.80	0.77	81.07
1000 hz	0.00	0.00	0.83	1.00	88.52
1200 hz	0.00	0.00	0.86	1.00	90.75
1400 hz	0.00	0.00	0.93	1.00	95.19
1600 hz	0.00	0.00	0.91	1.00	92.59
Target 56	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.31	0.00	0.00	37.65
800 hz	0.33	0.36	0.25	0.27	58.34
1000 hz	0.13	0.33	0.60	0.40	75.00
1200 hz	0.40	0.56	0.53	0.33	87.22
1400 hz	0.18	0.33	0.55	0.67	89.06
1600 hz	0.00	0.36	0.71	0.64	90.15
Target 57	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.17	0.09	0.00	50.60
800 hz	0.06	0.00	0.24	0.27	75.41
1000 hz	0.20	0.08	0.20	0.38	80.77
1200 hz	0.08	0.20	0.67	0.70	86.19
1400 hz	0.07	0.20	0.93	0.67	91.43
1600 hz	0.20	0.22	0.73	0.78	93.18
Target 58*	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.00	0.00	0.00	70.18
800 hz	0.00	0.20	0.00	0.40	84.85
1000 hz	0.14	0.07	0.36	0.40	80.61
1200 hz	0.07	0.00	0.33	0.44	87.68
1400 hz	0.00	0.08	0.55	0.08	80.68
1600 hz	0.00	0.09	0.12	0.27	84.42
Target 59	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.42	0.40	0.33	0.60	84.41
800 hz	0.21	0.13	0.71	0.87	89.53
1000 hz	0.07	0.44	0.80	0.56	93.49
1200 hz	0.09	0.17	0.82	0.83	90.48
1400 hz	0.00	0.18	0.94	0.82	95.82
1600 hz	0.11	0.08	0.78	0.85	92.47

Target 60	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.10	0.08	0.80	0.58	91.20
800 hz	0.09	0.08	0.73	0.83	86.58
1000 hz	0.04	0.00	0.85	0.96	94.24
1200 hz	0.00	0.00	0.92	1.00	97.13
1400 hz	0.07	0.00	0.86	0.93	93.15
1600 hz	0.07	0.22	0.93	0.78	95.56
Target 61	No Context	Criminal	No Context	Criminal	
600 hz	0.14	0.13	0.29	0.40	60.56
800 hz	0.00	0.11	0.67	0.67	81.49
1000 hz	0.00	0.08	0.82	0.83	84.03
1200 hz	0.00	0.00	0.88	0.91	92.03
1400 hz	0.00	0.00	1.00	1.00	96.03
1600 hz	0.00	0.00	1.00	1.00	95.00
Target 5**	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.00	0.45	0.56	43.67
800 hz	0.00	0.00	0.86	0.78	37.96
1000 hz	0.00	0.13	1.00	0.75	72.72
1200 hz	0.00	0.00	0.92	0.78	72.92
1400 hz	0.00	0.00	0.91	1.00	81.42
1600 hz	0.00	0.00	0.92	1.00	90.88
Target 17**	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.15	0.20	0.46	0.10	65.52
800 hz	0.00	0.11	0.78	0.89	77.46
1000 hz	0.08	0.08	0.92	0.92	88.30
1200 hz	0.00	0.00	1.00	1.00	88.71
1400 hz	0.00	0.00	1.00	1.00	86.86
1600 hz	0.00	0.00	1.00	1.00	99.08
Target 25**	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.50	0.33	0.25	0.00	35.12
800 hz	0.46	0.27	0.00	0.00	40.41
1000 hz	0.45	0.70	0.27	0.10	55.13
1200 hz	0.38	0.69	0.54	0.15	68.14
1400 hz	0.10	0.60	0.90	0.40	72.47
1600 hz	0.25	0.56	0.63	0.44	80.56
Target 45**	**No Context**	**Criminal**	**No Context**	**Criminal**	
600 hz	0.00	0.00	0.11	0.20	32.91
800 hz	0.00	0.00	0.56	0.33	55.48
1000 hz	0.00	0.17	0.88	0.50	67.86
1200 hz	0.00	0.00	0.92	0.89	92.22
1400 hz	0.00	0.00	0.78	0.89	75.55
1600 hz	0.00	0.00	0.85	1.00	93.57

Note. Degradation levels highlighted in blue represent the degradation levels used for recordings in Studies 1 and 3. Degradation levels highlighted in yellow represent the minimally degraded recordings selected for Study 2. Degradation levels highlighted in green represent the moderately degraded recordings selected for Study 2. Target statements highlighted in red were not included in any of the studies. For Target statement 51 1400 Hz (highlighted in grey), the wrong statement played, and thus, there is not data for this degradation level.

*These recordings were not included in Study 2.

**These recordings were not included in Pilot 2 but were included in Pilot 1 and used in Studies 1-3.

Appendix F.

Incriminating Transcripts Used in Studies 2 and 3

Note: Items with an Asterisk were Used in Study 3 Only.

I kept my knife hidden from him.

Next thing I knew, there was blood everywhere.

She didn’t know it, but I was armed.

I knew I needed to kill him.

She was scared of what I was going to do to her.

I had to get rid of that body.

I snapped because I was so pissed at him.

After everything he’d done to me, it was time for payback.*

I didn’t feel bad when I saw what I’d done to him.

I looked at him and told him he was dead.*

He was becoming a real problem to have around.

He stopped breathing, and I knew it was over.

I never mean to injure him.

I bet he never expected me to mug him.

I buried all of his stuff, even his clothes.

I bashed him while he wasn’t looking.

I shot him over and over again.

He wouldn’t listen to me so I started choking him.

He saw me pull out my gun.

I remember the day that I stole her car.

I strangled them late that night.

I never meant to hurt her.

I broke in their house in the middle of the night.

I took all her money while she wasn’t looking.

I slit her neck when she turned away from me.

I drowned them behind my house.

I killed them in my house.

She didn’t know that I robbed the bank that night.

I started the fire in the house right in time.

I told him I could give him a deal on some drugs.

I checked to make sure he wasn’t breathing.

I choked them in the water.*

I dumped both of them in the lake.

As she turned to walk away, I grabbed her throat.

She wouldn’t stop crying so I broke her arm.

I went to go do the crime with him.

I stabbed him over and over again.*

I had to stab them before it was too late.*

I made sure he was dying.

He begged me not to hurt him.

I left them there to die.

I raped her body in front of all those guys.

I mugged her so I could get some money.

I told him we needed to kill them.

I shot all of those guys up.

I got scared when I saw what I'd done to him.*

I dropped my gun on the floor.

I buried her outside my house.

Appendix G.

Directions Participants Received Across the Different Conditions in Study 3

No Transcripts

Part 1:

Throughout this survey, you will hear a series of audio recordings. On each page, you will hear one recording. Once the recording finishes, you can proceed to the next page. You will do this until you've listened to all of the recordings. Please listen to the recordings carefully.

Part 2:

In the second part of this study, you will hear the same audio recordings that you heard in the first part of the study. You will hear each recording once and you will be asked to transcribe what you hear. Please listen to the recordings carefully and type what you hear into a box on the screen. If you're not sure, give your best guess. Once you have finished typing, please proceed to the next page.

(No) Delay/No Instructions

Part 1:

Imagine you are a juror in a series of criminal trials. The primary evidence against each of the accused is a series of audio recordings that come from wiretapped conversations with the accused. You will see written transcripts of each statement that are presented at the same time as the statement. Please listen to the statements and read the transcriptions carefully.

On each page, you will hear one recording and see the corresponding transcript. Once the recording finishes, you can proceed to the next page. You will do this until you've listened to all of the recordings. Please listen to the recordings carefully.

Part 2:

In the second part of this study, you will hear the same audio recordings that you heard in the first part of the study without the transcripts. You will hear each recording once and you will be asked to transcribe what you hear. Please listen to the recordings carefully and type what you hear into a box on the screen. If you're not sure, give your best guess. Once you have finished typing, please proceed to the next page.

(No) Delay/Instructions

Part 1:

Imagine you are a juror in a series of criminal trials. The primary evidence against each of the accused is a series of audio recordings that come from wiretapped conversations with the accused. You will see written transcripts of each statement that are presented at the same time as the statement. Please listen to the statements and read the transcriptions carefully.

The transcripts are just an aid to help you follow the recordings as they are played. The transcripts are not evidence. Only the recordings themselves are evidence. If what you read on the transcripts differs from what you hear in the recordings, you are to go by what you hear for yourself, and not what you read in the transcript. Please follow carefully as the recordings are played.

On each page, you will hear one recording and see the corresponding transcript. Once the recording finishes, you can proceed to the next page. You will do this until you've listened to all of the recordings. Please listen to the recordings carefully.

Part 2:

You will hear the same audio recordings that you heard in the first part of the study without the transcripts. Remember that the transcripts were just an aid to help you follow the recordings. The transcripts were not evidence. Only the recordings themselves are evidence. If what you read on the transcripts differs from what you hear in the recordings, you are to go by what you hear for yourself.

The recordings will now be played once again, without the transcripts. Please listen to each recording and type the statement as you heard it in the box. If you're not sure, give your best guess. Once you have finished typing, please proceed to the next page.