(Dis) continuing the continued influence effect of misinformation

Saoirse Connor Desai

A dissertation submitted for the degree of

Doctor of Philosophy (PhD)

of

City, University of London.

Department of Psychology

September, 2018

Declaration

I grant powers of discretion to the University Librarian to copy this thesis in whole or in part without further reference to the author. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Abstract

Misinformation can often have a lasting impact on the causal inferences people make about events even after it is unambiguously corrected. This is known as the continued influence effect of misinformation. Understanding the underlying cognitive processes involved in correcting misinformation is important for developing effective counter-misinformation strategies. A review of continued influence studies suggests that methodological factors, such as the scenario in which misinformation appears, can affect the strength of a correction independently of experimental manipulations. This thesis' primary aim was to advance understanding the continued influence effect and the conditions that give rise to it, by addressing issues with the methodological approach. Experiments 1A-2B developed and validated a methodology for web-based testing of the continued influence effect in order to target larger and more diverse samples. Two key claims from the continued influence literature were replicated; and the introduction of a novel control condition showed that misinformation is referred to even if it is only mentioned as part of the correction. Experiments 3-5 re-examined the claim that corrections, which explain how misinformation occurred, reduce reliance on misinformation. Findings showed no evidence that explanatory corrections reduce misinformation reliance in multiple scenarios; and continued influence of misinformation was observed in some scenarios but not others. Experiments 6-7 revisited the claim that misinformation, which implies a likely cause of an outcome is more impervious to correction than explicitly stated misinformation. Findings showed no evidence for this claim in three scenarios. I argue that the continued influence effect is a tenuous claim predicated on findings from a small set of scenarios that are unrepresentative of real-world situations in which misinformation is encountered. I propose that the conditions under which continued influence of misinformation occurs are poorly understood and recommend formally modelling the continued influence effect to gain a better understanding of this phenomenon.

“Man's memory shapes

Its own Eden within”

— Jorge Luis Borges, Dreamtigers

Published articles

Experiments 1A, 1B, 2A, and 2B appears in Connor Desai, S. & Reimers, S. (2018). Comparing the use of open and closed questions for Web-based measures of the continued-influence effect, *Behavior Research Methods,* doi: https://doi.org/10.3758/s13428-018-1066-z.

Experiment 6 appears in Connor Desai, S., & Reimers, S. (2018). Some misinformation is more easily countered: An experiment on the continued influence effect. In Proceedings of the 40th Annual Meeting of the Cognitive Science Society (pp. 1542-1547). Austin, TX: Cognitive Science Society.

Experiment 3 appears in Connor Desai, S., & Reimers, S. (2017). But where's the evidence? The effect of explanatory corrections on inferences about false information. In Proceedings of the 39th Annual Meeting of the Cognitive Science Society (pp. 1824-1829). Austin, TX: Cognitive Science Society.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First, and foremost, I would like to thank my supervisor Stian Reimers for his excellent supervision. Stian has taught me a lot about how good experimental psychology should be done. His continued patience, guidance, and support have given me the motivation and confidence to develop as a researcher in my own right. Stian's passion and enthusiasm for research have been a constant source of inspiration, from the days I was sat at the back of his 'Psychology of Time' lectures conjuring up images of dinosaurs, to the final stages my PhD.

I also owe a great deal of thanks to my second supervisor Peter Ayton. Peter has always been available to lend an ear, offering guidance and support. Peter's unique perspective on the academic world has made it a far more interesting place. I could not have imagined a better set of supervisors. Thank you also to the wonderful academic and administrative staff in the Department of Psychology at City, who have been a pleasure to work with.

A lot of the ideas in thesis were shaped by attending the CoDes MSc programme at UCL. For this, I would like to thank the academic staff in Experimental Psychology at UCL. In particular, Dave Lagnado who has encouraged, supported, and guided me over the last 6 years - even though I'm an Arsenal fan! I would also like to thank Adam Harris and Maarten Speekenbrink who have always happily offered advice and listened my musings on 'Love Island'.

I would like to acknowledge the friends and colleagues who helped and encouraged me throughout this journey. In no particular order: Phil Newell for a mind-altering workshop, Leo Cohen for his witty reflections on science and dating, Neil Bramley for always being down for an adventure, Eric Schulz for our Pacific coast road trip, Paula Parpart for giving me the confidence to make a change, Sonia Abad Hernando for always being there for me, Abby Ipser for showing me the ropes, and Will Crichton and Chloe Bukata for being great pals (and always doing my pilot studies!). I also want to thank all of the Psychology PhD group at City – you're an awesome bunch. Thank you to my colleague and friend Toby Pilditch - who gave some invaluable comments on an earlier draft of this thesis - and Tom Hardwicke who advised me during the early stages of this project.

Finally, none of this would have been possible without the support of my incredible family. Dad - for getting me to read Sophie's World, Kahil for always encouraging and motivating me when my confidence was lacking. Of course, I can't forget Granny, Shelly, Amala, Brian, Pappy, and Grandad. But most of all, thank you to Mum and Ray – I am eternally grateful.

“You Can't Unring a Bell.”

— David Foster Wallace, Infinite Jest

Definition of Key Terms and Abbreviations

Term	Definition
Target (Mis)information	The piece of information which for conditions featuring a correction is misinformation but for conditions not featuring a correction is just a piece of causal information
References to Target (Mis)information	The number of references to target (mis)information that a participant makes in response to inference questions
CIE	Continued influence effect of misinformation

1 You can’t unring a bell

1.1. Introduction

Causal thinking is central to the ways in which the human cognitive system represents the external world and is critical for a broad range of cognitive operations (Oaksford & Chater, 2001; Sloman & Lagnado, 2014). Knowledge of causal relations plays a key role in how we represent unfolding events, construct coherent memory representations, create explanations, and reason in everyday life (Johnson-Laird, 1983; Van Dijk & Kintsch, 1983). The causal inferences we make about events also underpin how memory representations are updated when previously encountered information is corrected as erroneous. For example, we might initially learn that a train crashed because the driver was asleep but then later discover that this information was false and the true cause is unknown. As this Chapter’s title suggests, erroneous information that provides a causal explanation for an event or outcome is particularly difficult to correct (Johnson & Seifert, 1994).

Misinformation – defined here as information that is initially thought to be true but which later turns out to be erroneous - can have serious and widespread repercussions for society. As such, misinformation has become an important issue for governments, media organisations, and citizens over recent years (see Lewandowsky, Ecker, & Cook, 2017). Concerns about the prevalence and impact of misinformation have been heightened by the increase in use of social media and user-driven content online (Del Vicario et al., 2016; World Economic Forum, 2018). Moreover, large networks that are facilitated by social media may serve to maintain and strengthen mistaken beliefs rather than improving and correcting them (Madsen, Bailey, & Pilditch, 2018).

Given the increase in the availability of misinformation, understanding the underlying cognitive processes involved in correcting misinformation is

important for developing effective counter-misinformation strategies. An obvious strategy is to issue clear and incontrovertible corrections to the misinformation. However, cognitive psychology research suggests that this strategy may not be as effective as assumed. An accumulating body of experimental evidence indicates that misinformation often continues to influence thinking despite clear and credible corrections (see Chan, Jones, Hall Jamieson, & Albarracín, 2017; Cook et al., 2015; Lewandowsky, Ecker, Schwarz, & Cook, 2012; Walter & Murphy, 2018, for reviews).

One real-world example of the continuing influence of misinformation is the widespread belief that there is a causal link between the measles mumps and rubella (MMR) vaccination and autism, despite scientific evidence refuting the myth (Horne, Powell, Hummel, & Holyoak, 2015). Decreased acceptance of the MMR vaccination has contributed to a 7% drop in vaccination rates in the UK and a 1.7-fold increase in refusal to vaccinate in the US (Smith, Ellenberg, Bell, & Rubin, 2008), and consequently, an increase in a vaccine-preventable disease. Continued reliance on misinformation can also have consequences in the political domain. For instance, Vote Leave's erroneous referendum pledge that the UK would recover £350 million after leaving the European Union was widely considered to have swayed the decision to leave (Ipsos MORI, 2016), despite being discredited by the UK statistics authority as a 'clear misuse of statistics' (Full Fact, 2017).

There are several reasons people might continue to believe corrected misinformation. First, they may remember the misinformation (e.g. there is a causal link between the MMR vaccination and autism) but not its correction (e.g. no causal link between autism and MMR vaccination). Second, they may have a strong motivation to ignore the correction – either because it is inconsistent with their world-view or they distrust the correction's source. Third, the correction may not carry as much weight or be as convincing as the original information.

The issue appears to be more widespread than this, however. Many laboratory-based studies have shown that when people are asked, they often remember that a piece of information was corrected, but still use it to reason about scenarios describing unfolding events (Ecker, Lewandowsky, & Tang,

2010; Ecker, Lewandowsky, Swire, & Chang, 2011; Ecker, Lewandowsky, & Apai, 2011; Guillory & Geraci, 2013; Johnson & Seifert, 1994; Rich & Zaragoza, 2016). People also refer to misinformation in neutral scenarios in which there is no inherent motivation to believe the misinformation over the correction, and even when the correction is explicit and incontrovertible. This phenomenon is referred to as the continued influence effect (CIE) of misinformation (Johnson & Seifert, 1994).

As discussed above, the CIE has potentially very serious implications for societal decisions in domains such as the media, law, and healthcare. For instance, members of a jury might incorporate invalid or inadmissible information into their verdict decisions despite instructions to disregard it (Fein, McCloskey, & Tomlinson, 1997). People might continue to believe scientific claims are true even though the scientific article, in which those claims were made, has been retracted (Greitemeyer, 2014). Belief of social stereotypes may persist even after evidence disconfirms it (Kunda & Oleson, 1995), and false political beliefs about politics may remain after they have been explicitly corrected (Nyhan & Reifler, 2010).

Continued influence studies typically involve examining how people process corrections to constructed misinformation in fictional scenarios (e.g. a warehouse fire) rather than misinformation that has featured in real-world situations (e.g. weapons of mass destruction in Iraq). Studying corrections to constructed misinformation in fictional scenarios has methodological value. It allows for testing the effectiveness of corrections to misinformation in neutral scenarios in which participants have no reason to prefer one version of events over another, and it is easier to control for prior exposure, defensive processing, and guards against potential floor/ceiling effects (Thorson, 2016). The methodological approach used in CIE studies therefore allows for exploration of the individual-level cognitive and memory mechanisms involved in the continuing influence of misinformation, whilst providing the means to control for attitudinal or ideological factors.

Although the CIE paradigm has several advantages for studying the ways in which people process corrections to misinformation, issues with the methodological approach could affect the ecological validity and

generalisability of findings. First, studies on the CIE have investigated the corrections to misinformation in a small number of scenarios (e.g. warehouse fire), that may not be representative of different types of situations in which people encounter misinformation in the real-world. Second, most studies have been laboratory-based which entails small samples of mainly university students, who may not behave like a more diverse group of participants. Third, although the CIE has been replicated across a range of studies, it is still not well understood why it occurs, and under which circumstances. This can pose significant challenges to understanding the prevalence of the CIE as a phenomenon in the world.

In order to gain a better understanding the circumstances under which the CIE occurs, it is important to address methodological issues that could affect the validity and reliability of findings obtained using the CIE paradigm. The thesis' primary aim was to advance our understanding of the continued influence effect, and the conditions under which it occurs, and to overcome existing methodological problems of validity and reliability allowing for more systematic testing of the CIE in the future. This was achieved through a series of methodological steps. The first step was to develop and validate a methodology that allows for web-based testing of the CIE in order to be able to collect data from larger and more representative samples. This methodology was then used to examine the prevalence of the CIE across different scenarios in more realistic settings (e.g. road accident), and with larger samples. This methodology was further used to investigate the robustness of two important claims from the CIE literature, and also examine whether continued reliance on misinformation occurs when misinformation is only mentioned in the correction.

The first claim from the CIE literature examined, was that corrections which explain how misinformation occurred (e.g. from unintentional error or intentional lie) can help people understand the contradiction between misinformation and the correction, thereby reducing continued reliance on misinformation (as suggested by Bush, Johnson, & Seifert, 1994; Johnson & Seifert, 1994; Seifert, 2002). The second claim investigated was that misinformation, which implies a likely cause of an adverse outcome, is more

resistant to correction than misinformation explicitly stating a likely cause (Rich & Zaragoza, 2016). The robustness of these claims was examined to establish the circumstances under which the CIE is likely to occur. The key development that my research has made to this field is to show that the CIE does not occur under all circumstances and is likely moderated by the specifics of the scenario in which misinformation appears.

The rest of this introductory chapter examines the current status of CIE research. First, I give a brief overview of approaches in cognitive psychology and memory research that have been used to study how people disregard prior information. In doing so, I make the case for why the CIE approach is ideal for studying the cognitive mechanisms involved in processing corrections to misinformation. I then describe the CIE and summarise its key findings, explain the experimental paradigm used to study the CIE, and discuss the potential advantages and limitations of this approach. Following this, I discuss the two cognitive mechanisms that have been proposed to explain the CIE, and the relative evidence for these two positions. Next, factors that have been shown to moderate the CIE are considered. Finally, I address the methodological approach used in CIE studies and examine whether methodological factors can potentially moderate the CIE.

1.2. Disregarding Prior Information

The CIE is perhaps the most researched paradigm for examining how people disregard prior information. However, there are several other areas of research that have similarly looked at how people disregard prior information that has been deemed irrelevant. In this section, I first describe three alternative approaches that have been used to examine how people disregard prior information with respect to social (belief perseverance), legal (disregarding invalid or inaccurate testimony), and memory (direct forgetting) judgments.

1.2.1. Belief perseverance

Belief perseverance describes the human tendency to cling onto initial beliefs to an unwarranted extent (Anderson, New, & Speer, 1985; Ross & Anderson, 1982), particularly with regard to complex social judgments. Studies on belief perseverance use variants of a 'debriefing' paradigm attributed to Ross, Lepper, and Hubbard (1975). In the debriefing task, participants are misled to create a belief, which is subsequently discredited by explaining that the created belief was all part of the experimental manipulation. For example, Ross et al., (1975) showed participants cards containing real and fictional suicide notes; and then asked them to categorise the real from fictional notes, receiving feedback about their accuracy. After being 'debriefed' about the manipulation and told that their score was randomly allocated, participants' post-debriefing ratings of their performance and abilities revealed a continuing impact of the discredited success-failure manipulation.

The belief perseverance effect has been replicated with a range of belief systems (e.g. self-impressions, social theories) and subject matters (e.g. risk preference, firefighting ability, mathematical ability, political beliefs; Anderson, 1982; Anderson, 1983; Anderson & Kellam, 1992; Anderson, Lepper, & Ross, 1980; Anderson, 1983; Anderson & Barrios, 1961; Davies, 1997; Wyer & Budesheim, 1987). These findings have led to the view that people are conceptually inflexible which may explain why people cling to mistaken beliefs in the face of clear counter-evidence.

One possible mediator of belief perseverance is the availability of causal arguments (i.e. in favour of a social theory or impression). For instance, Anderson et al., (1985) gave participants data describing either a positive or negative relationship between risk-taking and firefighting ability, which was later explained as fictitious during debriefing. They found a classic perseverance effect in that debriefing participants clung to the data-induced theories despite knowing that they were fictitious. More interestingly, analysis of participants' written explanations of why a positive or negative relationship

might exist, revealed that causal argument availability accounted for approximately half the observed belief perseverance effect.[1] Anderson et al. argued that people start to think in causal terms as soon as they experience surprising or interesting events. The idea is that subsequently people engage in hypothesis-confirming information searches and biased processing of new information. If no new data disconfirming their initial belief are encountered, then the initial belief will be maintained by the relative availability of arguments supporting that belief.

1.2.2. Disregarding invalid testimony

A separate applied strand of research has looked at presentation of evidence and its correction, using courtroom scenarios to examine juror decision making. This approach requires participants to process complex stimuli about unfolding events. Unlike belief perseverance research, studies on discounting invalid information in courtroom settings have shown successful discounting of invalid information under specific conditions. The basic courtroom paradigm involves presenting a mock jury with several testimonies and targeting one testimony for later discounting. For example, target testimony may be ruled invalid because an eyewitness had poor vision and was not wearing glasses or because a piece of wiretap evidence was obtained illegally.

Studies examining disregarding prior information that adopt a courtroom paradigm broadly fall into two categories. Target testimony can either be followed by an explicit instruction from the judge to disregard the invalid information (Thompson, Fong, & Rosenhan, 1981), or discredited under cross-examination (e.g. by getting the witness to acknowledge that they made a mistake; Hatvany & Strack, 1980), or else via the introduction of contradictory evidence (e.g. alibi is given but CCTV evidence is introduced placing the suspect at the crime scene; Lagnado et al., 2011). In the case of discrediting via cross-examination or introduction of contradictory evidence,

[1] This was computed via a covariance analysis. The mean difference between debriefing conditions was subtracted from the original mean difference to partial out the unique relative effect of argument availability

the instruction to discount invalid information is implied, and individuals come to their own conclusion that target information is invalid and should therefore be discounted.

Another factor that distinguishes between these two situations is that discrediting usually provides some reason to disbelieve or discount earlier information (e.g. the witness was drunk at the time of the incident and therefore her testimony is unreliable). Instructions to disregard inadmissible information (e.g. because information is hearsay or evidence violates due process) on the other hand, do not necessarily rule out the relevance or truthfulness of the target information. Instead, they ask the receiver to disregard the information because it may have a biasing effect on their judgment even though it may still be true. In fact, the instruction to disregard could be interpreted as an indication that the information is in fact true but should be discounted because of its potentially prejudicial effects. For example, a juror may infer that information ruled inadmissible – because it violates due process or is hearsay – would not have been mentioned if it was not relevant or true (cf. Grice, 1975).

Studies on discounting invalid information in courtroom settings support the distinction that appeals to discount that offer a reason to disregard prior information (e.g. it is unreliable information) are more successful than requests to discount because of potential bias. For instance, Elliott, Farrington and Manheimer (1988; see also Weinberg & Baron, 1982 for a similar finding) conducted a study in which participants read about two armed-robbery cases. The cases included both direct and circumstantial evidence that were sufficient to evoke judgements of guilt. An eyewitness identifying the defendant was later discredited (the witness conceded under cross-examination that he was near-sighted and his vision had been blurry). Discrediting the eyewitness testimony was fully effective at lowering judgments of guilt, irrespective of the strength of direct and circumstantial evidence, the standard of proof used, and whether participants made serial judgments after reading successive increments in the summaries.

In contrast, people often do not disregard prior information when they are explicitly asked to ignore it because it is inadmissible (see Steblay, Hosch,

Culhane, & McWethy, 2006 for review of studies on disregarding inadmissible information). In fact, the main factor that has been shown to reduce the influence of inadmissible information is whether a rationale for the disregard instruction was offered. For example, Kassin and Sommers (1997) compared verdicts and guilt ratings in a case where evidence (wire-tap evidence from an unrelated trial) was ruled inadmissible because of due-process (it was illegally obtained) to one where it was ruled inadmissible because it was unreliable (the tape was inaudible). The proportion of guilty verdicts following the 'unreliable' correction was equivalent to a control group who were not presented with the evidence. In fact, the 'unreliable' instruction halved guilty verdicts relative to the due-process instruction. This finding suggests that people might continue to rely on information they have been told to disregard, if they think it is still relevant to their judgment. It appears that people are unable to resist the urge to use information they have been asked to disregard in their judgment whether they are aware of its biasing effects or not.

1.2.3. Directed forgetting

Directed forgetting is another branch of research on disregarding prior information. Directed forgetting studies focus on the specific memory processes involved in goal-directed forgetting of simple stimuli, such as word lists. In a directed forgetting paradigm, participants learn a list of items, some of which are cued for later recall (R-cued) and others are cued to be forgotten (F-cued). Performance can either be assessed via explicit memory tests such as recall and recognition (Bjork & Woodward, 1973; Woodward & Bjork, 1971), or by implicit tests (Basden, Basden, & Gargano, 1993; Macleod, 1989), and F-cued information can either consist of individual items or entire lists of words (MacLeod, 1999). Successful forgetting is exhibited when participants produce more R than F-cued words (Johnson, 1994).

Findings from directed forgetting studies contrast with some inadmissible information and belief perseverance findings because they usually find that people are able to suppress the to-be-forgotten information at test (Bjork, Bjork, & Anderson, 1998; Johnson, 1994). The proposed mechanism by which people are able to 'forget' the information they have

been instructed to is retrieval inhibition. More precisely, a process which inhibits subsequent retrieval of the to-be-forgotten information is initiated when participants are instructed to forget prior information and given new information to learn. Although the to-be-forgotten information is not directly retrievable, it remains at strength in memory and can be accessed by other measures such as recognition or word-fragment completion (Bjork & Bjork, 1996).

One prominent difference between the research paradigms described earlier in this section and directed forgetting studies, is the use of forgetting cues (Bjork et al., 1998). Participants in directed forgetting studies are told from the outset that they may receive an instruction to forget some of the material presented during the study, or told that some information had been incorrect and that they will now see the 'correct' materials for later study. This difference in the formulation of forget instructions may play a role in why some research areas find successful 'forgetting' of previously learned information and others do not. Most importantly, an instruction to disregard prior information in the context of a judgement experiment is not equivalent to an instruction to forget in the context of a memory experiment.

1.2.4. Conclusions

Methodologically, there are similarities between belief perseverance, disregarding invalid testimony, directed forgetting, and CIE approaches. Understanding the differences between these approaches and the variation in findings among these different approaches may help shed light on the conditions under which successful updating following a correction occurs. In belief perseverance and disregarding invalid testimony studies, the cue to forget, although clear, is usually implicit. Directed forgetting studies, by contrast, typically involve explicit cues to forget information. Such cues to forget usually do not imply 'forgetting' per se, but are instructions not to report the information at recall. Another important difference is the type of stimuli used. For example, belief perseverance and studies of disregarding invalid testimony use complex stimuli about social situations and events which necessarily invoke prior knowledge of causal relations. The word lists used in

directed forgetting studies are potentially less cognitively demanding, and easier to hold in working memory, than the complex social and event stimuli used in belief perseverance, invalid testimony, and CIE studies.

There are three main advantages that the CIE paradigm has over the above approaches that make it ideal for exploring novel research questions on the cognitive processing of corrections to misinformation. First, unlike the directed forgetting approach, the CIE task uses rich and complex descriptions of unfolding events that are more representative of the types of situations in which people might be asked to disregard information in the real-world. Second, rich and complex stimuli are formulated in a way that makes experimental manipulations more precise than in the courtroom approach. Third, unlike the belief perseverance approach, the CIE paradigm does not rely on the experimenter to discredit initially constructed beliefs. Instead, the correction is presented within the context of the scenario as it would be in a breaking news story. These factors make the CIE paradigm a promising method for studying how we reason about corrected information.

1.3. The Continued Influence Effect

As noted in the introduction to this chapter, the continued influence effect refers to the finding that causal misinformation is often influential beyond a clear and credible correction (see Ecker, Swire, & Lewandowksy, 2014; Lewandowsky et al., 2012; Seifert, 2002, 2014, for reviews of CIE research).

1.3.1. The continued influence paradigm

Continued influence studies examine corrections to misinformation using variants of a laboratory paradigm first developed by Wilkes and Leatherbarrow (1988; but also see Johnson & Seifert, 1994). A typical CIE task involves reading between 10 and 15 sequentially presented statements describing an unfolding event (i.e. an event extending in time with a sequence of different elements). The way in which event information is presented

resembles that of breaking news coverage. Target (mis)information[2] that allows inferences to be drawn about the outcome of the event is presented early in the sequence but then corrected later. Participants' inferential reasoning and factual memory for the story are then assessed through a series of open-ended questions.

The classic example of the CIE task is in Johnson and Seifert (1994), wherein participants were given an unfolding story about a warehouse fire. Target (mis)information which implies that carelessly stored flammable materials (oil paint and gas cylinders) are a likely cause of the fire is presented, and later corrected for some participants, but not others. Participants who received a correction learnt that no oil paint and gas cylinders had actually been stored in the warehouse, and therefore could not have caused the fire. After reading the report, participants answered causal inference questions (e.g. "what could have caused the explosions?"), and questions probing recall of the literal content of the story (e.g. "what was the cost of the damage done?"). Their responses were then categorized according to whether they were consistent with the explanation implied by the target (mis)information (e.g. "exploding gas cylinders"), or not (e.g. "electrical short circuit"). Participants were also asked to recount the correction (e.g. "what was the point of the second message from Police Investigator Lucas?"). The number of references to the corrected misinformation was then calculated in order to measure how much misinformation continued to influence participants' inferential reasoning about the story.

Continued influence experiments usually involve evaluating performance (i.e. the number of references to target (mis)information), on a misinformation-followed-by-correction condition to an upper or lower bound of comparison: either a condition in which the misinformation is presented but is not then retracted[3] (no correction condition), or a condition in which the misinformation is never presented (no misinformation condition). The no correction condition allows for measurement of a correction's effectiveness.

[2] (Mis) is parenthesized because in some control conditions the information is not corrected, meaning it cannot be considered misinformation from the participants' perspective.

[3] The terms 'retraction' and 'correction' are used interchangeably throughout this thesis and in the continued influence effect literature.

By contrast, the no misinformation condition arguably shows whether post-correction references to misinformation are reduced to a level comparable to never having seen the misinformation in the first place.

1.3.2. Main findings from CIE studies

The key finding from CIE studies is that corrections to misinformation are only partially effective. Some studies find no difference in the aggregate number of misinformation references between a condition featuring a correction and one in which a correction is presented (Johnson & Seifert, 1994). Usually a correction has some impact but fails to fully eliminate the misinformation's influence on subsequent causal inferences (Ecker et al., 2010; Ecker et al., 2011a; Ecker et al., 2011b; Guillory & Geraci, 2013; Rich & Zaragoza, 2016). Misinformation can have a lasting impact on people's reasoning even when people demonstrably remember that the information was corrected (Johnson & Seifert, 1994; Marsh, Meade, & Roediger, 2003), receive prior warnings about the persistence of misinformation (Ecker et al., 2010), and when misinformation is corrected immediately after it is presented (Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988).

1.3.3. Benefits and limitations of the CIE approach

The CIE is a robust finding that has been demonstrated using different scenarios (e.g. warehouse fire, plane crash, armed robbery), and types of causal misinformation; for example, neutral and emotionally laden (plane crash caused by bad weather or terrorist attack; Ecker et al., 2011), negatively - but not positively - valenced information (a politician was caught giving a bribe versus a donation; Guillory & Geraci, 2016), attitude-congruent racial information (Ecker, Lewandowsky, Fenton, & Martin, 2014), and information that implies rather than explicitly states the cause of an adverse outcome (Rich & Zaragoza, 2016).

As mentioned in the introduction to this Chapter, CIE studies involve examining how people process corrections to constructed misinformation. Using constructed misinformation circumvents some of the problems faced by

studies that examine processing of corrections to real-world misinformation, such as controlling for prior exposure, defensive processing, and guarding against potential floor/ceiling effects (Thorson, 2016). The CIE approach enables a better understanding of the cognitive mechanisms involved in processing corrections to misinformation that allows causal inferences to be drawn about an event or outcome. Presenting statements sequentially also makes it easier to separate out the manipulated pieces of information from the other information included in the story.

Despite the advantages of the CIE approach, inconsistencies in the approach and materials used could lead to variability in the effectiveness of a correction independently of the manipulated variables. For instance, the warehouse fire story often includes additional information that is congruent with the explanation that oil paint and gas cylinders caused the fire, such as the presence of ‘toxic fumes’, ‘explosions’, ‘oily smoke and sheets of flames’ and ‘an intense heat’. This may bias participants’ interpretation of the situation by making it appear that there is more evidence in favour of the explanation offered by the misinformation, than there is for the correction. Or in other words, if one starts with the hypothesis that carelessly stored oil paint and gas cylinders caused the fire, and learns about features of the incident supporting this hypothesis (e.g. there were toxic fumes), it might be reasonable to assume that the information provided makes the misinformation more likely than the correction.

Other features of the typical experimental task such as the specific questions used to elicit inferences or the strictness of the coding criteria could also moderate the strength and presence of the CIE. For example, people may be more likely to refer to misinformation if asked very specific (e.g. what was the possible cause of the fumes?) than more general causal inference questions (e.g. is there evidence of careless management in relation to this fire?) about the scenario. A related issue is that researchers who adopt the CIE methodology provide no threshold for observing the CIE. There is currently no specification of how many post-correction references to misinformation are considered necessary for the CIE to be observed, or for that matter, what the different levels of continued reliance on misinformation

mean. If the average number of post-correction references to misinformation are zero, this indicates that the correction has effectively eliminated reliance on misinformation. However, the situation is more ambiguous when post-correction references are low but still not zero (see, for example, Experiment 2 in Ecker et al., 2011a).

This is important because participants may refer to misinformation because of the pragmatic demand produced by certain types of questions, when they have no other information to rely on (Schwarz, 1996). A participant may ask themselves why the experimenter would ask about the corrected misinformation if it were not in some way relevant to the question. Such an interpretation could result in references to misinformation that are not strictly a consequence of continued reliance on outdated information, but rather, a willingness to provide the experimenter with the information they appear to be asking for. Such a demand effect may not arise simply from a desire to please the experimenter, but instead, because of the expectation that questions would only refer to relevant information presented in the scenario (cf. Grice, 1975). If the only potential causal information in the scenario is the misinformation, then the pragmatic interpretation would suggest that it must be the answer to the question by the experimenter, even if they know it is erroneous (see Bless, Strack, & Schwarz, 1993; Schwarz, 2014 for discussion of pragmatic demand effects in cognitive psychology research).

Studies adopting the CIE methodology have also tended to be restricted in terms of the experimental stimuli used and samples recruited. The limited number of scenarios used in CIE studies may not be generalisable to real-world situations in which misinformation is encountered. For example, apart from specific situations, it is rare that you would be told a piece of information is incorrect without any further explanation as to why it is incorrect, or any evidence to back up the claim that it is wrong. Furthermore, studies on the CIE have tended to examine different conditions of misinformation and its correction, in a single scenario (although see Johnson & Seifert; 1994; Wilkes & Leatherbarrow, 1988; Wilkes & Reynolds, 1999), which may only be representative of a small number of situations in which misinformation can be encountered naturally. The limited number of scenarios

used, and focus on examining the CIE in a single scenario, has clear implications for the ecological validity of CIE findings; particularly, if scenario interacts with the specific manipulations to the presentation of misinformation or a correction.

Another issue is that most CIE studies have been conducted in the lab with university students (although more recently researchers have moved to testing the CIE online: e.g. Guillory & Geraci, 2013; Rich & Zaragoza, 2016). University student sample demographics are not representative of the normal adult population because they are inherently biased in terms of age, experience, political beliefs, intellectual ability, ethnicity, and social class. Recruiting participants with biased demographics could result in misestimating the prevalence of the CIE and overlooking cognitive (or other) factors that might worsen the effect. For instance, age-related differences in memory for prose can be explained by differences in working memory capacity (Light & Anderson, 1985), and people with relatively low intelligence and poor perceptual abilities are more susceptible to the post-event misinformation effect (Zhu et al., 2010).

Despite these issues, CIE research has the potential to inform the types of cognitive mechanisms involved in both successful and unsuccessful correction processing. Knowledge of these mechanisms can, in turn, inform strategies for successful correction of misinformation regarding a range of societal issues (e.g. media, law, healthcare). In the next section I describe the two main theoretical positions and cognitive mechanisms that have been discussed in the literature thus far, and consider the relative evidence for each of these accounts.

1.4. Cognitive Mechanisms and Theoretical Accounts

Two cognitive mechanisms and theories have been proposed to account for the continued influence effect to date: The mental-model updating (Johnson & Seifert, 1994, 1999; Wilkes & Leatherbarrow, 1988; Wilkes & Reynolds, 1999), and retrieval failure accounts (Ecker et al., 2010; Jacoby, 1991). While the mental-model account postulates a failure to appropriately

update a causal mental model, the retrieval-failure account assumes a failure of strategic memory processes when retrieving misinformation.

1.4.1. Mental-model updating

As noted above, one account of the CIE assumes that people construct mental models of unfolding events. Mental models can be defined as representations of the world, or descriptions of the world, based on available semantic information and an individuals' knowledge. They 'represent distinct possibilities, that unfold in time in a kinematic sequence' (Johnson-Laird, 2010, p. 7), and are used to draw conclusions from available information.

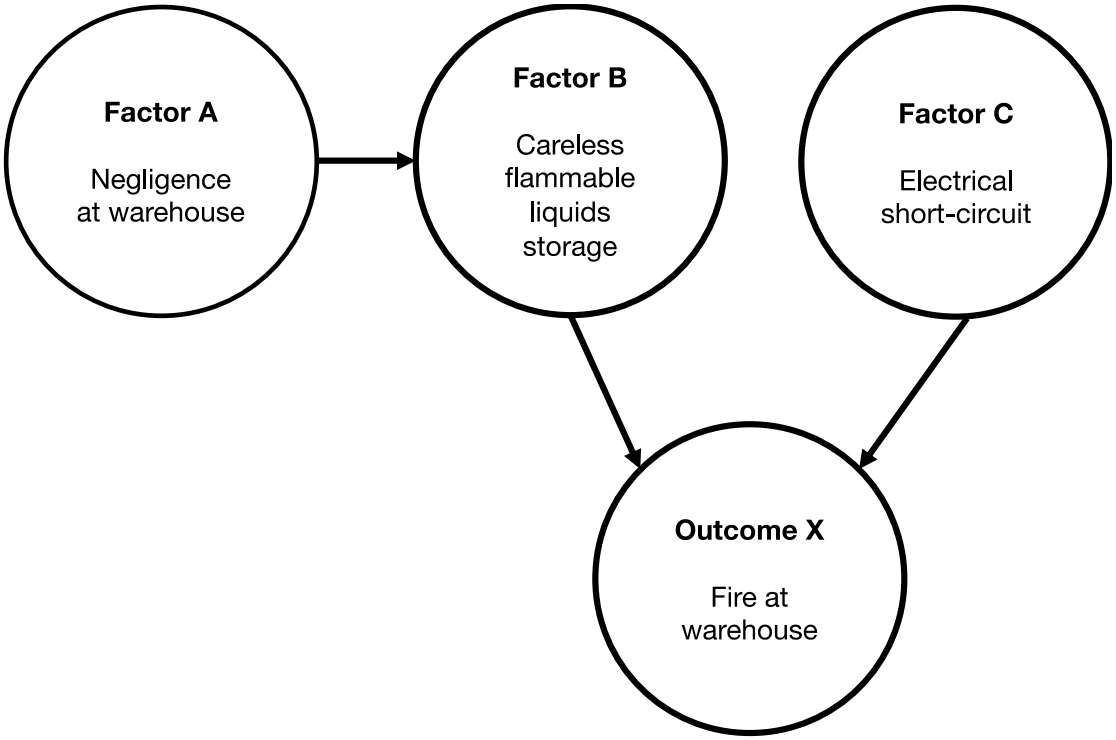

Figure 1 Schematic diagram of causal mental model for warehouse fire scenario

According to the mental-model view, people construct meaning from descriptions of events by organizing incoming information into a coherent mental representation in which causal connections between events are necessary and sufficient (Trabasso & van den Broek, 1985), and temporal, spatial, and causal aspects of the described situation are continually tracked (Zwaan, Magliano, & Graesser, 1995). Prior causal knowledge relating to physical (e.g. space and time) and intentional (e.g. beliefs, goals, desires and motivations) causality is used to draw inferences where information is missing (Pennington & Hastie, 1986, 1988, 1992), facilitating development of a model which is comprehensible and does not contain inconsistencies.

Figure 1 depicts a possible mental-model of the warehouse fire scenario in which corrected misinformation provides a causal explanation for the described outcome. The warehouse fire scenario could be represented such that factor A (e.g. negligence) resulted in factor B (e.g. inappropriate storage of flammable liquids), and the combination of factor B and factor C (e.g. an electrical short-circuit) caused outcome X (e.g. fire at the warehouse; example taken from Lewandowsky et al., 2012).

One reason that corrections to misinformation may be poorly encoded - or retrieved - is because they threaten the mental model's internal coherence (Johnson-Laird, 1980; Johnson & Seifert, 1994). For example, when reading the warehouse fire scenario described earlier, an individual might infer that a fire started by an electrical short circuit was a result of improper storage of flammable liquids. Correcting a key piece of causal information (i.e. stating that there were no oil paints or compressed gas present) that explains the outcome, results in an incoherent mental model of the described event. People therefore continue to rely on misinformation after a correction because they prefer to maintain a coherent but inaccurate model to an

incoherent, incomplete, but more accurate one. When participants are explicitly probed about the correction, they are aware of the update to the specifics (i.e. admitting that information about gas cylinders and paint cans was corrected), but continue to rely on the misinformation more broadly to answer inference questions.

1.4.1.1. Empirical evidence for the mental-model updating

Perhaps the most widely cited argument in favour of the model updating account is the finding that combining a correction with an alternative causal explanation[4] facilitates updating when correcting misinformation (Ecker, Hogan, & Lewandowsky, 2017; Ecker et al., 2011a; Ecker et al., 2010; Johnson & Seifert, 1994; Nyhan & Reifler, 2015; Rich & Zaragoza, 2016). Similar findings have also been obtained from courtroom analogue studies that examine guilt judgments after introduction of an alternative suspect (Tenney, Cleary, & Spellman, 2009). Providing an alternative explanation purportedly helps people 'fill the gap' in their model, facilitating or motivating global updating of the mental model of the event. This in turn reduces reliance on the original explanation offered by misinformation allowing people to disregard their initial model in favour of a new one.

Recent neuroimaging evidence also provides tentative support for the idea mental-model updating failures underlie the CIE. Gordon et al. (2017) examined the underlying neural substrates of the CIE by comparing

[4] Throughout the remainder of this thesis 'alternative explanation' will be used to specifically refer to 'alternative causal explanations'.

differences in neural processing during encoding and retrieval of retraction and non-retraction information. This was examined by having participants listen to brief fictional news reports that either involved a retraction of initial information or not, whilst undergoing functional magnetic resonance imaging (fMRI). The authors argued that the mental-model account predicts difficulties with encoding retracting information (either because it violates assumptions or disrupts model coherence), whereas the retrieval failure account predicts difficulties substituting the correct (retraction) for the incorrect (contradicted information), and should produce problems at retrieval.

These predictions were tested by examining brain activity in areas associated with violation of assumptions, shallower encoding of information, and memory suppression and/or substitution. Consistent with the mental-model predictions, encoding identical pieces of information elicited different brain activity depending on whether the information was processed as a retraction or not. Retractions elicited less activity in brain regions that have been previously associated with integration of continuous pieces of verbal information into a mental-model. These findings suggest that the CIE may be driven by a breakdown of narrative-level integration and coherence-building mechanisms and that information is encoded differently depending on whether it is a correction or not. There was little evidence that both types of information engage different neural mechanisms at retrieval.

Additional evidence for the mental-model accounts comes from the finding that working memory capacity – which plays a role in information integration and updating – predicts occurrence of the CIE (Brydges, Gignac, & Ecker, 2018). Furthermore, explicitly repeating misinformation during a correction can reduce reliance on misinformation by increasing its salience and highlighting the discrepancy between outdated and updated event interpretations (Ecker et al., 2017).

1.4.2. Retrieval failure

An alternative theory to explain the CIE is that it occurs because of a failure in controlled memory processes during retrieval. According to this view, both erroneous (i.e. misinformation) and correct (i.e. the correction)

information are stored in memory concurrently and the CIE arises when the erroneous information is activated but insufficiently suppressed (Ecker et al., 2011b; Swire, Ecker, & Lewandowsky, 2017)[5]. This argument takes root in dual-process accounts which differentiate between retrieval based on automatic (familiarity) and (recollection) strategic memory processes (Hasher & Zacks, 1979, 1984; Jacoby, 1996; Johnson, Hashtroudi, & Lindsay, 1993; Smith & DeCoster, 2000). From a dual-process perspective, the type of memory process that is activated upon retrieval of information could impact a correction's effectiveness. For instance, Ayers and Reder's (1998) activation-based framework proposes that valid and invalid information compete for automatic activation in memory but that strategic processes are required in order to retrieve contextual details of the information.

Automatic memory processes are thought to be driven by familiarity and afford rapid and context-free recognition of previously encountered information, reflecting a more global measure of stimulus recency and memory strength. By contrast, strategic memory processes rely on the slower process of recollection which allows for retrieval of qualitative or contextual details, such as the information's source, veracity, and spatiotemporal encoding context (see Yonelinas, 2002 for review). Strategic memory processes are easily compromised as they require more executive control and mental effort than automatic memory processes (Herron & Rugg, 2003; Jacoby, 1991). They are depleted with age (Craik & McDowd, 1987; McDermott & Chan, 2006; Swire et al., 2017), when attention is divided (Craik et al., 1996), and when there is a longer retention interval between study and test (Knowlton & Squire, 1995).

Ecker et al. (2010; 2011b; see also Lewandowsky et al., 2012) have argued that misinformation could occur because of failures in strategic monitoring at retrieval. This view is based on the idea that misinformation needs a "negation tag" to be linked with the original statement when it is corrected. An example would be the statement "playing Mozart to your child will boost its IQ – NOT TRUE" (Gilbert, Krull, & Malone, 1990). Corrected

[5] Note that both mental-model updating and retrieval failure mechanisms have been argued for by Ecker and colleagues.

statements may require strategic memory processes in order to successfully retrieve the veracity of the statement. The negation tag may be compromised if only automatic memory processes are employed at retrieval which results in misidentification of the misinformation as familiar. Another possibility is that the CIE arises from source confusion or misattribution (Johnson et al., 1993). For instance, participants given the warehouse fire scenario may remember the fire was thought to be caused by oil paint and gas cylinders, but incorrectly attribute this information to the second report.

1.4.2.1. Empirical evidence for retrieval failure

Assuming the dual-process perspective, activation of automatic retrieval processes could result in reliance on erroneous but familiar information. Evidence that the CIE is driven by a familiarity-based mechanism comes from Swire, Ecker, and Lewandowsky (2017). Swire et al. (2017) examined how factors known to affect strategic memory processes influence continued reliance on inaccurate information; namely age, detail in the correction, and time. They obtained pre-manipulation ratings of belief in statements of unclear veracity. Some statements were myths which were later corrected whilst others were facts that were affirmed. Swire et al. predicted that increasing the level of explanatory detail provided in the correction would support recollection of the correction and encourage detection of inconsistencies between inaccurate beliefs and the correction. They also predicted false acceptance of myths based on their familiarity would be more likely at longer retention intervals between encoding and retrieval and would be more apparent for older than for younger participants. Findings showed that providing a greater level of explanatory detail in the correction promoted more sustained belief change suggesting a boost in strategic memory processes at retrieval. Longer retention intervals between encoding and retrieval also resulted in more belief in corrected myths than shorter retention intervals. Acceptance of corrected myths was also higher for the older than the younger group. These findings suggest that the CIE may – in some cases - be familiarity-driven.

Further evidence that the CIE may be familiarity-driven comes from the finding that multiple repetitions of misinformation during encoding results in more of a CIE than a single repetition (Ecker et al., 2011b). Repeated statements are considered easier to process, and are also perceived as more truthful, than novel statements (Fazio, Brashier, Payne, & Marsh, 2015; Moons, Mackie, & Garcia-Marques, 2009), but only if statements are plausible (Pennycook, Cannon, & Rand, 2018). However, Ecker et al., (2011b) also found that multiple repetitions of a correction did not reduce the CIE below the level of a single presentation of misinformation and a correction. This is perhaps because corrections require strategic monitoring in order for their context to be recalled accurately and are therefore not enhanced by repetition. Consistent with this idea, Ecker et al. (2010) found that providing explicit warnings about the persistence of misinformation before exposure to misinformation reduced the CIE relative to providing general warnings, or no warnings. Ecker et al. argued that providing explicit warnings about the potential biasing effects of misinformation boosts strategic monitoring processes during retrieval of misinformation and its correction, therefore reducing the CIE.

There is also counter-evidence to the claim that repeating misinformation increases its familiarity resulting in more of a CIE (cf. Ecker et al., 2011b). Ecker et al. (2017) compared several correction conditions varying the extent to which the correction served as a reminder of initial misinformation. Corrections that served as explicit reminders of initial misinformation were the most successful at reducing reliance on misinformation. The authors interpreted this finding as consistent with the mental-model updating account in that an explicit reminder made both the falsity of misinformation and conflict between outdated and updated event representations more salient. However, these findings could equally be interpreted as supporting a retrieval-based account of the CIE in that an explicit reminder made the correction more memorable and therefore more effective at the time of retrieval.

1.4.3. Conclusions on cognitive mechanisms

To conclude, there is a lack of resolution regarding whether the CIE occurs because of a failure to update one's mental-model, or because of strategic memory processes when retrieving misinformation. Mental-model updating and selective retrieval accounts are often postulated as competing mechanisms (e.g. Gordon et al., 2017), and findings interpreted as supporting one or the other mechanism. However, it is often difficult to establish the relative influence of each of these mechanisms to the CIE. It could be the case that the particular mechanism that elicits the CIE varies between, or within, individuals at different time points, or depending on the specifics of the experimental situation.

1.5. Factors that Moderate the Continued Influence Effect

A number of factors have been identified that moderate the CIE (see Lewandowsky et al., 2012 for discussion). The main factors that have been identified in the literature will be considered here and discussed in terms of the potential constraints they place on when the CIE is likely to occur. Practical implications are also discussed.

1.5.1. Providing a causal alternative

As noted above, Johnson and Seifert (1994; Exp 3A) first established that pairing a correction with a causal-alternative which fills the explanatory gap left by invalidating the correction mitigates – but does not eliminate - the CIE. Participants read the warehouse fire story, described earlier, and initially learned that oil paint and gas cylinders were a likely cause of the fire. After initial misinformation was corrected, some participants then learned that gasoline-soaked rags and empty steel drums were found on the premises (i.e. implying that the fire was started intentionally rather by careless storage of flammable liquids as was originally inferred).

Studies including an alternative explanation condition have typically found that this halves the number of references to misinformation relative to a misinformation-correction condition (Ecker et al., 2011b; Ecker et al., 2010;

Johnson & Seifert, 1994; Nyhan & Reifler, 2015; Rapp & Kendeou, 2007), whilst others have found that alternative explanations almost eliminate the CIE completely (Ecker et al., 2011a). Differences in the capacity of an alternative explanation to reduce reliance on misinformation could suggest that alternative explanations are more effective in some scenarios than others.

There are some boundary conditions that apply to the causal-alternative strategy. For instance, the alternative explanation must also be instantiated within the context of the story rather than generated by participants themselves (Johnson & Seifert, 1994; Exp 2). It is also assumed that the alternative explanation must be plausible and account for the event features initially accounted for by the misinformation (Johnson & Seifert, 1994; Seifert, 2002). For example, the warehouse fire story included information about 'toxic fumes' and 'oily smoke' which could also be explained by someone intentionally dousing the warehouse in gasoline (i.e. the alternative explanation). Explaining that the fire was actually caused by a terrorist attack might be less plausible, given the other non-corrected details presented in the scenario.

Similar findings have been obtained from studies examining how introduction of an alternative suspect (and story) influences judgments of a defendant's guilt (e.g. Tenney, Cleary, & Spellman, 2009), and explanation-based refutations of earlier story information (Rapp & Kendeou, 2007). Providing a causal alternative has also been shown to improve the effectiveness of corrections in the political domain. For instance, Nyhan and Reifler (2015) gave participants a story which initially insinuated that a senator had resigned from office because he had embezzled money and committed tax fraud. This information was either denied or replaced with the 'causal alternative' that the senator resigned for personal reasons and had a prosecutor's letter confirming he had not been charged with any crimes. The innuendo was perceived as less likely to be true when a causal alternative was presented than when the innuendo was simply denied.

There are other factors that might moderate whether an alternative explanation can reduce the CIE, however. People have strong preferences

about what constitutes a ‘good’ explanation (see Lombrozo, 2016, for review). These preferences may influence the type of alternative explanations that are acceptable means of filling the gap left by invalidating the misinformation. For instance, simpler explanations (i.e. explanations that involve a single rather multiple causes to explain an outcome) are widely regarded as better than more complex ones (Lombrozo, 2007). This effect is mitigated or eliminated when the simple explanation is less probable than the more complex one (Lagnado, 1994; Lombrozo, 2007). Furthermore, explanations that involve fewer assumptions and have greater explanatory ‘scope’ are also favoured over explanations that have less scope (Khemlani, Sussman, & Oppenheimer, 2010; Read & Marcus-Newhall, 1993).

These factors impose some possible constraints on the types of explanations that are likely to improve the effectiveness of corrections. Practically speaking, it is often rare to find a single, coherent, and plausible alternative explanation to fill the gap left by invalidating a piece of causal information. There are often several competing explanations that may vary in terms of complexity, coherence, and plausibility.

1.5.2. Pre-exposure warnings

As noted in the previous section, forewarning people about the persistence of misinformation after a correction has also been shown to moderate the CIE. Ecker et al., (2010) demonstrated this in a study in which they gave some participants a specific warning explaining that people often continue to rely on outdated information - jurors do not ignore inadmissible evidence - before reading a scenario containing misinformation. Another group was given a more general warning that the media often report inaccurate information. Findings showed that providing a specific warning halved the number of references to misinformation produced, to provision of an alternative explanation. Although fewer references to misinformation were produced when participants received a general warning than just a correction, the general warning was not as effective as a specific warning. Specific warnings were even more effective when paired with an alternative

explanation, although this combination of mitigating factors still did not fully eliminate the CIE.

The benefits of warnings have also been observed for other judgment and memory biases. For instance, the imagination inflation effect – which occurs when people increase their confidence that an event occurred after imagining it – is attenuated when people are warned about the deleterious effects of imagining distant events (Landau & von Glahn, 2004); forewarnings also reduce false recognition of non-studied words (in the Dermot-Roediger-McDermott paradigm; Gallo, Roediger, & McDermott, 2001). Furthermore, both pre- and post-experiment warnings reduce the post-event misinformation effect and the benefits continued a week after testing (Chambers & Zaragoza, 2001).

One explanation for why pre-exposure warnings mitigate the CIE is that warnings enhance strategic retrieval of correction information. This allows the misinformation to be 'tagged' during rather than after encoding (Ecker et al., 2010). Tagging misinformation during encoding may make it possible focus on source monitoring, reducing the chance that the source of the memory is incorrectly attributed to another recollected experience (Johnson et al., 1993).

In terms of the practical application of such a counter-misinformation strategy, the timing of the presentation of the pre-exposure warning is likely important for enhancing strategic memory processes during retrieval of misinformation. Warnings that are presented well in advance of misinformation exposure may not enhance strategic retrieval of correction information or suppression of misinformation; either because the warnings are not recalled or cannot be effectively paired with misinformation and its correction.

1.5.3. Encoding strength

Outside of the laboratory, people may be exposed to misinformation on a disproportionate number of occasions to the correction. For instance, during the 'Vote Leave' campaign in the lead up the European Union (EU) referendum, people may have been exposed to the claim that leaving the EU

would save the National Health Service (NHS) £350 million pounds, on more than a handful of occasions. This information was criticised as being misleading throughout the campaign and it was then promptly discredited when the outcome of the referendum was announced when several prominent Leave campaign officials reneged on their claim. Most people who encountered this claim would have been exposed to it on numerous occasions - meaning that this erroneous information had more scope to be well encoded into memory. However, refutations of this information may not have been as widely reported and therefore not as well encoded into memory. Such an imbalance in the repetition of misinformation, and its correction or refutation will necessarily affect how well these pieces of information are encoded in memory, and therefore, how easily they are later retrieved from memory.

In line with this, empirical evidence indicates that encoding strength of misinformation and its correction impacts the CIE. The amount of effort employed by working memory – or cognitive load (Sweller, 1988) – during encoding of misinformation and its correction, can affect encoding strength. As noted previously, Ecker et al., (2011b) investigated how strength of encoding of misinformation and its correction affect the CIE by manipulating repetition of the misinformation and correction, and whether high (read aloud and memorise digits) or low (read aloud digits) load was imposed during encoding of misinformation and the correction. Stronger encoding (i.e. multiple repetitions or low cognitive load during encoding) of misinformation led to stronger continued influence of misinformation. Greater misinformation effects also required stronger retractions to be negated. Counter to expectations, however, the strength of the correction was inconsequential when misinformation was weakly encoded. That is if misinformation was only presented once then multiple corrections were no more effective than a single correction at reducing its influence. This finding has lead to recommendation that corrections should avoid repetition of misinformation in order to avoid increasing its impact (Lewandowsky et al., 2012). However, as noted previously, more recent evidence suggests that corrections which explicitly

repeat misinformation are more effective than those that avoid the misinformation altogether (Ecker et al., 2017).

1.5.4. Source credibility

Establishing a source's credibility is critical when deciding whether to believe the information conveyed to us by other people and may moderate the CIE. For example, jurors must establish a degree of belief in a witness' testimony in order to reach a verdict, and voters must place their confidence in the statements of politicians when deciding who to vote for. Source credibility typically refers to how believable a source of information is perceived to be (Birnbaum & Stegner, 1979), and is often orthogonally evaluated in terms of their trustworthiness and expertise (Birnbaum & Stegner, 1979; Birnbaum, Wong, & Wong, 1976; Harris, Hahn, Madsen, & Hsu, 2016; Hovland & Weiss, 1951; McGinnies & Ward, 1980; see Pornpitakpan, 2004 for review)[6]. Expertise refers to the source's capacity to convey accurate information whereas trustworthiness reflects their willingness to provide accurate information.

People may use a range of cues to evaluate the trustworthiness and expertise of a source. For example, studies in the legal domain have shown that witness calibration – the relationship between a witness' confidence and accuracy – influences judgements of credibility. That is, the credibility of highly confident witnesses who are shown to be wrong is penalised more than low confidence witnesses (Tenney, MacCoun, Spellman, & Hastie, 2007; Tenney, Spellman, & MacCoun, 2008). People may also use cues such as whether a witness contradicts themselves or is contradicted by another witness in order to assess the credibility. For instance, Connor Desai, Reimers, and Lagnado (2016) found that a prosecution witness' credibility was penalised when they claimed they had only drunk 4 pints but were contradicted by another prosecution witness who said they drank 8 pints. This penalisation of the key prosecution witness' credibility undermined their testimony and resulted in

[6] People also base credibility judgments on characteristics that seemingly have nothing to do with trustworthiness, such as attractiveness, listener, and situational characteristics (see Spellman & Tenney, 2010 for review)

lower ratings of the defendant's guilt compared to the non-contradictory testimony case.

The credibility of the misinformation's source and its correction may play a pivotal role in drawing inferences about the reliability of these pieces of information[7]. Courtroom studies have shown that witnesses who are shown to be wrong about something – whether trivial or not – lose credibility as perceived by the jury (Borckardt, Sprohge, & Nash, 2003; Tenney et al., 2007, 2008). Moreover, witnesses who contradict themselves are perceived as less credible than witnesses who do not contradict themselves (Berman & Cutler, 1996; Berman, Narby, & Cutler, 1995).

Contradiction is particularly relevant to CIE studies as misinformation and its correction are issued by the same source. A source who announces that they previously gave incorrect information may appear less credible than one who does not. Consistent with this, one CIE study found that distrust in the source of the correction was cited as a primary reason for disbelieving the correction (Guillory & Geraci, 2010). There is, however, a potential fallacy here in that someone who contradicts themselves should have *both* statements disbelieved, rather than just the latter statement. An initial source of information who later corrects themselves could result in that source being perceived as either unable or unwilling to provide accurate information. Alternatively, the later statement (correction) could increase credibility relative to the first statement (misinformation), given that the source now has access to new information (expertise) and has shown a willingness to contradict themselves (trustworthiness).

Either way, source credibility likely plays a role in how people process corrections to misinformation. Confirming this, Guillory and Geraci (2013) showed that corrections issued by a source who is considered to be highly trustworthy (i.e. willing to provide accurate information) more effectively reduces the CIE than a correction issued by a source high in expertise (i.e. access to accurate information). Participants were given a fictional news story about a politician running for re-election, which included target

[7] Reliability is used here to refer to the accuracy of a piece of information (Lagnado, Fenton, & Neil, 2013)

(mis)information about a politician receiving bribe money. The trustworthiness and expertise of the correction's source were independently manipulated. Findings showed that source expertise alone (e.g. the politician's campaign manager) was not sufficient to mitigate the CIE. However, source trustworthiness alone (e.g. politician's political opponent) decreased reliance on initial misinformation.

Taken together, these findings suggest that inferences about source credibility moderate how corrections to misinformation are processed. Source credibility may be used as a heuristic guide to establishing whether to accept or reject a given piece of information. Using source cues as a heuristic guide fits with Bayesian source credibility predictions (e.g. Hahn, Harris, & Corner, 2009; Harris et al., 2016). Lewandowsky et al., (2012) have also argued that source credibility (both high and low) could facilitate 'tagging' of correct and incorrect information and facilitating strategic retrieval of information from memory when this information is made salient.

1.5.5. Motivated reasoning

Another factor that may affect the magnitude of the CIE is motivated reasoning. How we process incoming information arguably depends on our motivations (via our goals, beliefs, and desires). One view is that motivational factors affect reasoning via strategies for accessing, constructing, and evaluating beliefs (Kunda, 1990). The motivation to arrive at a conclusion that fits our pre-existing attitudes and beliefs can, therefore, shape reasoning processes.

Ecker et al. (2014) provided evidence that motivated reasoning can influence how misinformation and its correction are initially encoded and later retrieved from memory. Participants in their study read a story about a robbery at a liquor store in which the suspects were described as Australian Aboriginal. Participants were divided into two groups: those who scored (relatively) high on a measure of racial prejudice towards Australian Aboriginals and those who scored low on the prejudice measure. When the story described the suspects as Australian Aboriginal, a correction failed to eliminate reliance on misinformation for both the high and low-prejudice

groups. However, when the misinformation described a citizen involved in preventing the robbery as Australian Aboriginal (i.e. stereotype-incongruent) a retraction eliminated reliance on misinformation for the high-prejudice but not the low prejudice group.

The fact that a retraction completely eliminated reliance on misinformation for the high-prejudice group when they received attitude-incongruent information (i.e. that the citizen preventing the robbery was Australian Aboriginal) could suggest that high-prejudice participants engaged in motivated reasoning. Supporting this, Ecker et al. reported that some participants rationalised that the Aboriginal man might have been an accomplice of the robber. The authors argued that, if the CIE can be assumed to arise from strategic memory failure, strategic monitoring could be improved when there is an attitude-based motivation to believe one version of events over another. More specifically, the high-prejudice group made certain to correct their initial attitude-incongruent interpretation of the event (i.e. that the citizen did not prevent the robbery).

These findings indicate the biasing effects that pre-existing attitudes, beliefs, and motivations have on how we initially encode and later retrieve information. However, critics of the motivated reasoning approach have disputed the notion that human beings are poor judgment and decision makers who are prone to motivational distortions and inherently irrational. Hahn and Harris (2014) have argued that in order for 'bias' to exist there must be systematic deviations from accuracy assessed against the appropriate normative standard rather than from the experimenter's intuition. Their argument is relevant for CIE which is often depicted as bias in that it reflects behaviour that systematically deviates from objective standard or norm (e.g. Lewandowsky et al., 2012).

Whether the CIE demonstrates a systematic deviation from a normative standard is debatable because it presupposes the normatively appropriate thing to do is favour the correction over the misinformation. There may be situations in which a legitimate strategy might be to place more weight on the misinformation than the correction. For example, if you distrust the source of the correction (as noted above) or the correction is inconsistent with

the current evidence then you may place more weight on the misinformation than the correction. Furthermore, the scenarios used in CIE studies do not provide any objectively 'true' information because they are fictional. This underscores the need to compare CIE behaviour against an appropriate normative standard rather than the experimenter's intuition (Harris & Hahn, 2014). Modelling CIE situations formally would help establish the types of inferences that might be legitimately permitted in different circumstances.

1.5.6. Communicative intentions

An additional moderating factor in the CIE could be the inferences people draw about the communicative intentions of the misinformation and correction. The Gricean (Grice, 1975) perspective on communication holds that people not only assess the literal meaning of information that occurs within the communicative context but also assess its communicative intention. According to the Gricean cooperative principle of communication people expect speakers to provide information that is relevant (*maxim of relation*), true (*maxim of quality*), and unambiguous (*maxim of manner*).

Conversational implications could affect which information appears relevant in the context of an experimental task. This has been shown in studies of judgmental biases (Krosnick, Li, & Lehman, 1990; Schwarz, Strack, Hilton, & Nadere, 1991). For example, Schwarz et al. (1991) found that people rely more on irrelevant (non-diagnostic) personality information when it is delivered by a human communicator than when it is presented as a random sample drawn from a computer database. This arguably occurs because people expect information from a human communicator to be truthful, relevant, and informative but do not hold the same assumptions for a computer. These findings suggest that participants might try to infer the experimenter's communicative intention from the information provided in the study and that this could render irrelevant information relevant in the eyes of the participant. This could result in judgmental errors relative to normative models that consider only the literal meaning of statements but not the implications for the communicative context.

Conversational implications could also pose problems for how people comprehend contradictory information. Seifert (2002; see also Johnson & Seifert, 1994) has argued that corrections are problematic for comprehension because people expect information relayed in a communicative context to be relevant and truthful. This view holds that the process of correction involves more than simply identifying previous information in memory and then negating it, but rather, the correction must address conversational implications of the contradiction in order to be fully understood.

Bush, Johnson, and Seifert (1994) tested this empirically by examining whether corrections that address the conversational implications of a contradiction are more effective than those that address only the literal implications of the contradiction. Participants read the warehouse fire scenario, described earlier in this chapter. One group of participants received an explanation that rendered the misinformation irrelevant (oil paints and gas cylinders were supposed to be delivered but were not), and a second group learned that misinformation was of poor informational quality (the closet actually contained non-flammable items). A third condition involved an 'enhanced negation' which addressed the literal implications of the contradiction by stating that oil paint and gas cylinders had never been present anywhere on the premises. The goal here was to address a literal interpretation of the contradiction that paint and compressed gasses were on the premises, but that they were stored elsewhere, and could thus still have caused the fire. The 'enhanced negation' condition tried to prevent participants from making the 'flammable liquids were elsewhere on the premises' inference.

Findings showed that explaining that misinformation was irrelevant or of poor informational quality attenuated the CIE relative a condition in which only a correction was presented but was not eliminated completely. Interestingly, the condition featuring enhanced negation (stating that oil paint and gas cylinders had never been present in the warehouse) actually increased references to misinformation! These findings suggest that correcting the literal content of misinformation is insufficient to produce an accurate understanding of the event. Addressing the conversational

implications of misinformation raised by a correction does mitigate – but does not eliminate - the continued influence and use of misinformation.

Overall, these findings suggest that assumptions about communicative intentions could moderate the effectiveness of a correction either because they make irrelevant information appear relevant in the experimental context or because they pose problems for understanding contradictory information provided in the context of communication.

1.5.7. Conclusions

In this section, I have discussed the main moderating factors of the CIE that have been discussed in the literature. The story that has emerged from this review is that the precise circumstances that bring about the CIE are poorly understood. It is unclear whether the CIE is a necessary consequence of correcting causal misinformation or whether it only occurs under certain conditions. This review has also drawn attention to the variability in the effectiveness of corrections - and other moderating factors such as alternative explanation – in reducing reliance on misinformation across studies that use similar manipulations. This suggests that other methodological characteristics of the study, such as the scenario used, could be an additional moderator of the effectiveness of corrections to misinformation in CIE studies.

1.6. Methodological Considerations

As mentioned earlier in this chapter, studies that have adopted the CIE experimental paradigm have found variance in the efficacy of corrections. Careful consideration of the experimental paradigm and methodological approach used in CIE studies is necessary to better understand how and why the CIE occurs. Factors such as the experimental stimuli (or scenario) used, sample size, and restricted demographics of the sample could moderate the presence and magnitude of the CIE, in addition to the variables that were manipulated in the study (e.g. pre-exposure warnings, valence of misinformation).

In order to examine whether the effectiveness of correction information varies across studies, I compiled a review of CIE studies that compared a

condition in which initial misinformation is corrected to a control condition in which it remained uncorrected (see Table 1 below). Table 1 includes information about the following study characteristics: 1) details of the authors of the paper, 2) which experiment the results refer to, 3) the correction and/or misinformation variables that were manipulated, 4) the scenario in which misinformation and the correction appear, 5) and whether the experiment was conducted in the lab or online. Table 1 also includes information about the study results, specifically: 1) the percentage reduction from uncorrected to corrected conditions in the average number of references to misinformation, 2) the effect size for the difference between corrected and uncorrected conditions, and 3) the sample size used in the experiment. It was not possible to include effect sizes if the appropriate descriptive information was not reported in the original paper.

The comparison between no-correction and misinformation-correction conditions was made - rather than the comparison between no-misinformation and misinformation-correction conditions- because it assesses the effectiveness of a correction. This is also the reason that the no-correction condition is used as the upper-bound of comparison throughout this thesis, rather than the no- misinformation condition, which, arguably, shows whether a correction reduces reliance on misinformation to a level comparable to having never been exposed to misinformation. The table also only includes studies which have used responses to open-ended inference questions in order to compute a measure of the extent to which a correction reduces the number of references to misinformation. There were some studies that did not fit this criterion and were therefore not included in the review. For instance, Gordon et al. (2017) tested the CIE with a series of 'comprehension probes' which required participants to indicate their level of agreement with a misinformation consistent statement. (This approach was used because participants completed the task while undergoing an fMRI scan). Similarly, Nyhan and Reifler (2015) examined how providing an alternative causal explanation reduces the impact of corrected misinformation but measured reliance via Likert scale responses.

One function of this table is to show that reductions in the CIE vary widely between 4-90%. This suggests that there is a lot of scope for affecting the CIE over and above the specific manipulations employed in a given study. The table also highlights the limited number of scenarios that have been tested and that manipulations are rarely compared across scenarios. Furthermore, the table shows that the warehouse fire scenario (discussed throughout this chapter) has been used repeatedly. As noted previously, one factor that might moderate the CIE is the scenario used in the study. There were actually more studies that have used variants of the warehouse fire scenario that are not included because they did not involve a 'no-correction' control condition (e.g. Guillory & Geraci, 2010).

Some experiments used different scenarios but could not be included because they did not include a no correction condition. For instance, Johnson and Seifert (1994; Exp 3B) examined reliance on misinformation for a scenario involving jewellery theft, and compared alternative explanation, negation (correction), and no misinformation conditions, but used the no-correction control in the warehouse fire story. Similarly, Wilkes and Leatherbarrow (1988) examined corrections to misinformation for two scenarios (warehouse fire, accident), but included a no-misinformation control in both their experiments. Wilkes and Leatherbarrow's results suggest some differences between scenarios. The proportion of spontaneous references to the 'targeted' information produced was higher for the accident report than the warehouse fire report. This suggests two things: First, that the explanations most easily brought to mind, without having been suggested by misinformation, differ depending on the specifics of the scenario. Second, that misinformation that is consistent with explanations that are more easily brought to mind, or more plausible, may be more difficult to correct than those that are not.

These findings, and the variability of the effectiveness of a correction across studies suggest that the scenario used may interact with the misinformation and/or correction. There may be a number of reasons for examining the CIE for a single scenario rather than comparing across scenarios. For instance, it may be time-consuming and difficult to construct

multiple scenarios that are comparable experimentally. There may also be a ‘file drawer’ issue in that some scenarios do not elicit the CIE and are therefore not used. This can be problematic for understanding the CIE as a phenomenon because specifics of the scenario may moderate whether the CIE is observed and to what degree. Comparing manipulations across different scenarios is a worthy endeavour which would help develop a better understanding of the circumstances that give rise to the CIE. Considering the methodological characteristics of the experimental procedure and whether this can artificially inflate or minimize the effect of interest, is important because CIE research may be used for policy recommendations (e.g. Ecker et al., 2014; Lewandowsky et al., 2012), but also because it hampers attempts to make scientific progress in the field.

Table 1 Characteristics, proportion reduction, and effect sizes of continued influence effect studies

Authors	Experiment #	Correction Variable	Misinformation Variable	Scenario	Online/Lab	Reduction (%)	Cohen's *d*	N	N p/cell
Bush, Johnson & Seifert 1994)	1	Just Correction		Warehouse Fire	Lab	46.67		116	Approx. 29
Bush, Johnson & Seifert 1994)	1	Enhanced Negation		Warehouse Fire	Lab	21.98		116	Approx. 29
Bush, Johnson & Seifert 1994)	1	Explanation (Quality)		Warehouse Fire	Lab	54.32		116	Approx. 29
Bush, Johnson & Seifert 1994)	1	Explanation (Relevance)		Warehouse Fire	Lab	57.04		116	Approx. 29
Johnson & Seifert (1994)	3A	Alternative Explanation		Warehouse Fire	Lab	39.81	0.88	81	Approx. 20
Johnson & Seifert (1994)	3A	Just Correction		Warehouse Fire	Lab	4.17	0.08	81	Approx. 20
Wilkes & Reynolds (1999)	1	Just Correction		Accident	Lab	21.96		36	Approx. 6
Wilkes & Reynolds (1999)	1	Just Correction		Warehouse Fire	Lab	27.34		36	Approx. 6
Ecker, Lewandowsky, & Tang (2010)	1	Alternative Explanation		Minibus Accident	Lab	56.12		125	25
Ecker, Lewandowsky, & Tang (2010)	1	Just Correction		Minibus Accident	Lab	20.17		125	25

Ecker, Lewandowsky, & Tang (2010)	1	General Pre-Exposure Warning		Minibus Accident	Lab	33.6		125	25
Ecker, Lewandowsky, & Tang (2010)	1	Specific Pre-Exposure Warning		Minibus Accident	Lab	58.1		125	25
Ecker, Lewandowsky, & Tang (2010)	2	Specific Pre-Exposure Warning + Alternative		Minibus Accident	Lab	77.59		67	No Correction = 25 Warning + Alternative = 42
Ecker, Lewandowsky & Apai (2011)	1	Alternative Explanation	Misinformation Emotionality (collapsed across)	Plane Crash	Lab	83.93	6.76	70	10
Ecker, Lewandowsky & Apai (2011)	1	Just Correction	Misinformation Emotionality (collapsed across)	Plane Crash	Lab	15.15	1.35	70	10
Ecker, Lewandowsky & Apai (2011)	2	Alternative Explanation	Misinformation Emotionality (collapsed across)	Plane Crash	Lab	71.79	1.66	112	16
Ecker, Lewandowsky & Apai (2011)	2	Just Correction	Misinformation Emotionality (collapsed across)	Plane Crash	Lab	68.5	1.38	112	16

Ecker, Lewandowsky & Apai (2011)	3	Alternative Explanation	Misinformation Emotionality (collapsed across)	Plane Crash	Lab	88.85	3.29	200	20
Ecker, Lewandowsky & Apai (2011)	3	Just Correction	Misinformation Emotionality (collapsed across)	Plane Crash	Lab	71.52	2.25	200	20
Ecker, Lewandowsky, Swire, & Chang (2011)	1	Repetition (1)	Repetition (1)	Warehouse Fire	Lab		0.89	161	23
Ecker, Lewandowsky, Swire, & Chang (2011)	1	Cognitive Load	Cognitive Load	Warehouse Fire	Lab		0.34	138	23
Ecker, Lewandowsky, Swire, & Chang (2011)	1	No Cognitive Load	No Cognitive Load	Warehouse Fire	Lab		0.47	138	23
Ecker, Lewandowsky, Swire, & Chang (2011)	1	Repetition (3)	Repetition (3)	Warehouse Fire	Lab		1.33	161	23
Guillory & Geraci (2013)	1	Source Credibility: High Expertise +High Trustworthiness		Re-election	Online	65.63	1.35	90	30

Guillory & Geraci (2013)	1	Source Credibility: Low Expertise & Low Trustworthiness	Re-election	Online	31.25	0.71	90	30
Guillory & Geraci (2013)	2	Source Credibility: High Expertise	Re-election	Online	31.25	0.74	90	30
Guillory & Geraci (2013)	2	Source Credibility: Low Expertise	Re-election	Online	43.75	0.97	90	30
Guillory & Geraci (2013)	3	Source Credibility: High Trustworthiness	Re-election	Online	46.88	1.11	90	30
Guillory & Geraci (2013)	3	Source Credibility: Low Trustworthiness	Re-election	Online	6.25	0.21	90	30
Ecker, Lewandowsky, Fenton, & Martin (2014)	1	Low Racial Prejudice - Attitude Congruent Correction	Liquor-store robbery	Lab			144	24
Ecker, Lewandowsky, Fenton, & Martin (2014)	1	High Racial Prejudice - Attitude Incongruent Correction	Liquor-store robbery	Lab			144	24

Ecker, Lewandowsky, Fenton, & Martin (2014)	2	Low Racial Prejudice - Attitude Incongruent Correction		Liquor-store robbery	Lab	100	25
Ecker, Lewandowsky, Fenton, & Martin (2014)	2	High Racial Prejudice - Attitude Congruent Correction		Liquor-store robbery	Lab	100	25
Guillory & Geraci (2016)	1	Just Correction	Valence: Neutral	Re-election	Lab	58	Within-Subjects 58
Guillory & Geraci (2016)	1	Just Correction	Valence: Positive	Re-election	Lab	58	Within-Subjects 58
Guillory & Geraci (2016)	1	Just Correction	Valence: Negative	Re-election	Lab	58	Within-Subjects 58
Rich & Zaragoza (2016)	1	Just Correction	Implied Cause	Jewellery theft	Lab	357	Corrected Implied = 42 Uncorrected Implied = 55
Rich & Zaragoza (2016)	1	Just Correction	Explicitly Stated Cause	Jewellery theft	Lab	357	Corrected Explicit = 49 Uncorrected Explicit = 59

Rich & Zaragoza (2016)	2	Alternative Explanation	Implied Cause	Jewellery theft	Lab				357	Uncorrected Implied = 60 Corrected Implied + Alternative = 56
Rich & Zaragoza (2016)	2	Alternative Explanation	Explicitly Stated Cause	Jewellery theft	Lab				357	Uncorrected Explicit = 61 Corrected Implied + Alternative = 51
Ecker, Hogan, & Lewandowsky (2017)	1	Reminder through Correction		Multiple (aggregated across scenarios)	Lab	32.76			60	Within-Subjects 60
Ecker, Hogan, & Lewandowsky (2017)	1	Reminder through Correction		Multiple (aggregated across scenarios	Lab	41.38			60	Within-Subjects 60
Ecker, Hogan, & Lewandowsky (2017)	1	Reminder through Correction		Multiple (aggregated across scenarios	Lab	53.45			60	Within-Subjects 60
					Mean	45.40	1.41			

Note: Only studies comparing 'No Correction' baseline to a form of correction are included here. Studies that do not measure

reliance on misinformation using the average number of references to misinformation on open-ended inference questions are excluded as well. Studies that did not include information about the condition means were also excluded because it was not possible to calculate the percentage reduction in average number of references to misinformation. The N p/cell was approximated for studies that did not specify exactly how many participants were allocated to each experimental condition.

1.7. Summary and Thesis Outline

Chapter 1 began by discussing the issue of misinformation in society and then discussed cognitive and memory approaches to studying how people disregard prior information (belief perseverance, disregarding invalid testimony, and directed forgetting). The CIE experimental paradigm was then proposed as the appropriate experimental methodology for studying how people process corrections to misinformation. This is because the CIE approach simultaneously uses rich scenarios whilst also carefully manipulating variables. Thus, the CIE approach makes it possible to study the cognitive processes involved in successful and unsuccessful corrections to misinformation for complex scenarios. After highlighting the benefits of this approach, I then discussed the limitations of this methodology. Studies on the CIE have also tended to use a limited number of scenarios and often do not compare effects across different scenarios. Furthermore, most CIE studies have mainly been conducted in the lab recruiting relatively small numbers of university students. It is important to address these limitations in order to be able to establish the precise set of circumstances that give rise to this phenomenon.

After initial discussion of the CIE paradigm and its limitations, I described the two mechanisms by which CIE is thought to occur, and discussed the empirical evidence for these two positions. Next, I described the main factors that are thought to moderate the CIE in order to establish what the CIE is and when it occurs. Through this review of the literature, I identified that the CIE may be partially explained by a distrust of the correction (either via source credibility heuristics or pragmatic inferences about the relevance and truthfulness of conveyed information). This analysis also suggested that factors which highlight the discrepancy between initial misinformation and the correction might be useful for reducing the CIE, either because they make people more aware of the inconsistency between initial and updated models or because they enhance strategic memory processes during retrieval (cf. Ecker et al. 2017).

The literature review has established that the CIE can be observed under some circumstances but not whether it is always guaranteed to occur. Furthermore, examination of the CIE paradigm and methodology typically used in these studies suggests that the limited number of scenarios used, small sample sizes in some studies, and focus on student populations, could pose problems for the validity and reliability of CIE findings. The aim of this thesis was to gain a better understanding of the circumstances that give rise to the CIE by establishing whether some methodological factors moderate the effect. Establishing this will advance understanding of the underlying cognitive processes involved in the CIE.

This thesis contains three empirical chapters that advance the CIE method through several steps. The overarching aim of Chapter 2 is to develop and validate a methodology for web-based testing of the CIE in order to collect data from larger samples who are more representative of the normal adult population. The specific goals are 1) to establish whether key CIE findings replicate, and 2) to explore the feasibility of converting open-ended questions to the type of closed-ended questions more typically seen online. The reason for converting open to closed-ended questions was to streamline the task for web-based testing because open-ended questions may be off-putting for participants completing experiments online. Chapter 2 also argues that web-based experiments provide a medium through which to target larger and more diverse samples. Research comparing responses to open- and closed-ended questions is reviewed in order to examine the feasibility of converting the open-ended questions to the closed format typically seen online, and whether different cognitive mechanisms are involved in responses to open- and closed questions. Chapter 2 establishes whether the web-based method of data collection is a viable means to test the CIE as this is the primary method of data collection in the remainder of the experiments reported in this thesis[8]. Experiments 2A and 2B reported in Chapter 2 also include a novel baseline condition in which

[8] The only exception to this is Experiment 4. The experimental design used in Experiment 4 meant that it was not feasible to conduct this study online.

misinformation is presented for the first time during the correction. This control condition was included in order establish whether a CIE still occurs when the misinformation is only encoded for the first time as it is being corrected. The aim of including this condition was to establish whether people still refer to misinformation even it is does not form the basis of an initial causal interpretation of an event. This inclusion of this condition also helps to determine whether the CIE can be explained in terms of the availability of the causal explanation offered by misinformation when answering inference questions.

Chapters 3 and 4 examine the effects of two potential moderators of the CIE across several scenarios that were designed specifically for this programme of research. Chapter 3 consists of three experiments: two of which are web-based and one that is laboratory-based. It investigates the claim that corrections that explain how or why misinformation occurred (e.g. intentional deception or unintentional error) are more effective than corrections that negate misinformation (as suggested by Bush, Johnson & Seifert, 1994; Johnson & Seifert, 1994). Chapter 4 consists of two web-based experiments that was designed to re-examine the finding that misinformation, which implies a likely cause of an adverse outcome, is more resistant to correction than misinformation which explicitly states a cause (as found by Rich & Zaragoza, 2016). In Chapter 5, the findings of this thesis are discussed in light of past research and theoretical implications for possible CIE mechanisms. Finally, practical implications are discussed.

2 Comparing the use of open and closed questions for web-based measures of the continued influence effect

2.1. Chapter Overview

Continued influence effect experiments have mainly been conducted in the laboratory (but see Guillory & Geraci, 2013; Hardwicke, 2016; Rich & Zaragoza, 2016, who have recently run CIE studies online). The CIE is also usually measured via responses to open-ended inference questions. Web-based data collection is the preferred method of data collection for the experiments reported in this thesis because it allows for testing of larger and more representative samples, and streamlines the research process. Chapter 2's aims were, therefore, threefold: First, to establish the feasibility of running CIE experiments online. Second, to compare open- and closed-ended inference and factual memory measures. Third, to examine the CIE in a novel control condition in which misinformation is only mentioned in the correction.

Chapter 2 consists of a paper that has been published in *Behavior Research Methods.* It is unchanged from the published version, except that the variable names have been changed to be more consistent with the remainder of this thesis. Some of the information included in the introductory section of this Chapter briefly repeats information included in Chapter 1 of this thesis but provides the immediate context for the work that follows. Chapter 2 reports the results of four experiments (Experiments 1A and 1B, 2A and 2B), that compare traditional open-ended responses to a closed-ended equivalent questionnaire. Experiments 1A and 1B aimed to replicate two key CIE findings: namely, that 1) a correction reduces but does not eliminate reliance on misinformation, and 2)

that an alternative explanation reduces reliance on misinformation beyond a simple correction.

Experiments 1A and 1B's results showed that a correction significantly reduced reliance on misinformation when open-ended measures were used (Experiment 1A), yet the difference was not significant with closed-ended measures (Experiment 1B). The second set of experiments were preregistered (https://osf.io/s39yr/) and examined whether differences between response measures were systematic or due to the small sample used in the first set of experiments. These experiments also included a novel baseline condition in which misinformation was presented for the first time during the correction. This control condition was included in order establish whether a CIE still occurs when the misinformation is encoded for the first time when its correction is presented.

Experiments 2A and 2B's results showed that misinformation which was only presented as part of a correction had as much of a CIE as misinformation presented early in a series of statements and only later corrected, for both open-ended (Experiment 2A), and closed-ended measures (Experiment 2B). The results of these experiments also showed that a correction did not significantly reduce reliance on misinformation when open-ended measures were used (Experiment 2A) but did so for closed-ended measures (Experiment 2B). The experimental stimuli, questionnaires, and full response coding criteria can be found in Appendices A, B, and C.

The results of the experiments reported in this chapter fit within the broader aims of the thesis for two main reasons: First, the results confirm that it is possible to perform complex memory-based experiments in which participants provide qualitative responses to questions online. Second, these results establish that the CIE can be partially explained by the availability of causal explanations when answering inference questions. The full implications of these results are discussed in Chapter 5 of this thesis.

2.2. Abstract

Open-ended questions, in which participants write or type their responses, are used in many areas of the behavioural sciences. Although effective in the lab, they are relatively untested in online experiments, and the quality of responses is largely unexplored. Closed-ended questions are easier to use online because they generally require only single key- or mouse-press responses and are less cognitively demanding but can bias responses. We compared data quality obtained using open and closed response formats using the *continued influence effect*, in which participants read a series of statements about an unfolding event, one of which is unambiguously corrected later. Participants typically continue to refer to the corrected misinformation when making inferential statements about the event. We implemented this basic procedure online (Experiment 1A, n = 78), comparing standard open-ended responses to an alternative procedure using closed-ended responses (Experiment 1B, n = 75). Finally, we replicated these findings in a larger preregistered study (Experiments 2A and 2B, n = 323). We observed the CIE in all conditions: Participants continued to refer to the misinformation following a correction, and references to target misinformation were broadly similar in number across open- and closed-ended questions. We found that participants' open-ended responses were relatively detailed (writing an average of 75 characters for inference questions), and almost all responses attempted to address the question. Responses for closed-ended questions were, however, faster. Overall, we suggest that with caution it may be possible to use either method for gathering CIE data.

2.3. Introduction

Over the past decade, many areas of research which have traditionally been conducted in the lab have moved to using web-based data collection (e.g. Peer, Brandimarte, Samat, & Acquisti, 2017; Simcox & Fiez, 2014; Stewart, Chandler, & Paolacci, 2017; Wolfe, 2017). Collecting data online has many

advantages for researchers, including ease and speed of participant recruitment, and a broader demographic of participants relative to lab-based students.

Part of the justification for this shift has been the finding that data quality from web-based studies is comparable to that obtained in the lab: The vast majority of web-based studies replicate existing findings (e.g. Crump, McDonnell, & Gureckis, 2013; Germine et al., 2012; Zwaan et al., 2017). However, the majority of these studies have been in areas where participants make single key- or mouse-press responses to stimuli. Less well explored are studies using more open-ended responses where participants write their answers to questions. These types of question are useful for assessing recall rather than recognition, and for examining spontaneous responses that are unbiased by experimenter expectations, and as such may be unavoidable for certain types of research.

There are reasons to predict that typed open-ended responses might be of lower quality than closed-ended responses. Among the few studies that have failed to replicate online have been those that require high levels of attention and engagement (Crump et al., 2013), and typing is both time-consuming and more physically effortful than pointing and clicking. Relatedly, participants who respond on mobile devices might struggle to make meaningful typed responses without undue effort.

Thus, researchers who typically run their studies with open-ended questions in the lab, and wish to move to running them online, have two options. They can either retain the open-ended question format or hope that online participants are at least as diligent as those in the lab, or they can use closed-ended questions in place of open-ended questions, but with the risk that participants will respond differently or draw on different memory or reasoning processes to answer the questions. We examine the relative feasibility of these two options using the continued influence effect, a paradigm which (a) is a

relatively well-used memory task, (b) has traditionally always used open-ended questions, and (c) is one that we have experience of running in the lab.

2.3.1. The continued influence effect

The *continued influence effect* of misinformation refers to the consistent finding that misinformation continues to influence people's beliefs and reasoning after it has been corrected (Chan, Jones, Hall Jamieson, & Albarracín, 2017; Ecker, Lewandowsky, & Apai, 2011; Ecker, Lewandowsky, Swire, & Chang, 2011; Ecker, Lewandowsky, & Tang, 2010; Gordon, Brooks, Quadflieg, Ecker, & Lewandowsky, 2017; Guillory & Geraci, 2016; Johnson & Seifert, 1994; Rich & Zaragoza, 2016; Wilkes & Leatherbarrow, 1988; for a review see Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012). Misinformation can have a lasting effect on people's reasoning even when they demonstrably remember that the information was corrected (Johnson & Seifert, 1994), and are given prior warnings about the persistence of misinformation (Ecker et al., 2010).

In the experimental task used to study the *continued influence effect* participants are presented with a series of 10-15 sequentially-presented statements describing an unfolding event. Target misinformation that allows inferences to be drawn about the cause of the event is presented early in the sequence and later corrected. Participants' inferential reasoning and factual memory for the event report are then assessed through a series of open-ended questions.

For example, in Johnson and Seifert (1994), participants read a story about a warehouse fire in which target information implies that carelessly stored flammable materials (oil paint and gas cylinders), are a likely cause of the fire. Later in the story, some participants learnt that no such materials had actually been stored in the warehouse and therefore could not have caused the fire. The ensuing questionnaire included indirect inference questions (e.g. "what could have caused the explosions?"), and direct questions probing recall of the literal content of the story (e.g. "what was the cost of the damage done?"). Responses

to inference questions are coded in order to measure whether the misinformation has been appropriately updated (no oil paint and gas cylinders were present in warehouse). Responses are categorized according to whether they are consistent with the explanation implied by the target (mis)information[9] (e.g. “exploding gas cylinders”), or not (e.g. “electrical short circuit”).

In a typical CIE experiment, performance on a misinformation-followed-by-correction condition is usually compared to one or more baselines: A condition in which the misinformation is presented but is not then retracted (no-correction condition), or a condition in which the misinformation is never presented (no-misinformation condition). The former allows assessment of the retraction’s effectiveness; the latter arguably shows whether the correction reduces reference to misinformation to a level comparable to never having been exposed to the misinformation (but see below).

The key finding from continued influence studies is that people continue to use the misinformation to answer inference questions even though it has been corrected. The most consistent pattern of findings is that references to previously corrected misinformation are elevated relative to a no-misinformation condition, and are either below, or in some cases indistinguishable from, references in the no-correction condition.

2.3.2. Using open- and closed-ended questions online

With a few recent exceptions (Guillory & Geraci, 2013, 2016; Rich & Zaragoza, 2016), research around reliance on misinformation has used open-ended questions administered in the lab (see Capella, Ophir, & Sutton, 2018, for overview of approaches to measuring misinformation beliefs). There are several good reasons for using them, particularly on memory-based tasks that involve comprehension or recall of previously studied text. First, responses to open-

[9] We use the term (mis)information throughout to refer to the original statement presented early in a CIE study that is later corrected. We parenthesize the (mis) because in some control conditions the information is not corrected, meaning it cannot be considered misinformation from the participants’ perspective.

ended questions are constructed rather than suggested by response options, and so avoid bias introduced by suggesting responses to participants. Second, open-ended questions also allow participants to give detailed responses about complex stimuli and permit a wide range of possible responses. Open-ended questions also resemble cued-recall tasks which mostly depend on controlled retrieval processes (Jacoby, 1996), and provide limited retrieval cues (Graesser, Ozuru, & Sullins, 2010). These factors are particularly important for memory-based tasks wherein answering questions requires active generation of previously studied text (Ozuru, Briner, Kurby, & McNamara, 2013).

For web-based testing, these advantages are balanced against the potential reduction in data quality when participants have to type extensive responses. The evidence around written responses is mixed. Grysman (2015a) found that participants on Amazon Mechanical Turk wrote shorter self-report event narratives than college participants completing online surveys, typing in the presence of a researcher, or giving verbal reports. Conversely, Behrend, Sharek, Meade, and Wiebe (2011) found no difference in the amount written in free-text responses between university-based and Mechanical Turk respondents.

A second potential effect is in missing data: Participants have anecdotally reported to us not enjoying typing open-ended responses. Open-ended questions could particularly discourage participants with lower levels of literacy or certain disabilities from expressing themselves in the written form, which could, in turn, increase selective dropout from some demographic groups (Berinsky, Margolis, & Sances, 2014). As well as losing whole participant datasets, open-ended questions in web surveys could also result in more individual missing data points than closed-ended questions (Reja, Manfreda, Hlebec, & Vehovar, 2003).

The alternative to using open-ended questions online is using closed-ended questions. These have many advantages, particularly in a context where there is less social pressure to perform diligently. Response options can also

inform participants about the researcher's knowledge and expectations about the world and suggest a range of reasonable responses (Schwarz, Hippler, Deutsch, & Strack, 1985; Schwarz, Knauper, Hippler, Neumann, & Clark, 1991; Schwarz, Strack, Müller, & Chassein, 1988).

There is also empirical evidence to suggest that open and closed responses are supported by different cognitive (Frew, Whynes, & Wolstenholme, 2003; Frew, Wolstenholme, & Whynes, 2004) or memory (Khoe, Kroll, Yonelinas, Dobbins, & Knight, 2000, see Yonelinas, 2002 for a review) processes. A straightforward conversion of open to closed-ended questions might, therefore, be impractical for testing novel scientific questions in a given domain.

This may be particularly relevant for the CIE. Repeated statements are easier to process and are subsequently perceived as more truthful than new statements (U. Ecker, Lewandowsky, Swire, & Chang, 2011; Fazio, Brashier, Payne, & Marsh, 2015; Moons, Mackie, & Garcia-Marques, 2009). Therefore, repeating misinformation in the response options could activate automatic (familiarity-based) rather than strategic (recollection-based) retrieval of studied text, which may not reflect how people reason about misinformation in the real world. Conversely, presenting corrections that explicitly repeat misinformation is more effective at reducing misinformation effects than presenting corrections that avoid repetition (Ecker et al., 2017). As such, substituting closed-ended questions for open-ended questions may have unpredictable consequences.

2.4. Overview of Experiments

The overarching aim of the experiments reported here was to examine open and closed questions in web-based memory and inference research. The more specific goals were to 1) to establish whether a well-known experimental task that elicits responses with open-ended questions replicates online, and 2) to explore the feasibility of converting open-ended questions to the type of closed-ended questions more typically seen online. In order to achieve this, two

experiments were designed in order to replicate the *continued influence effect*. Experiments 1A and 1B used the same experimental stimuli and subset of questions used in Johnson and Seifert (1994; Exp 3A), wherein participants read a report about a warehouse fire and answered questions that assessed inferential reasoning about the story, factual accuracy, and the ability to recall the correction or control information (critical information). Experiments 1A and 2A employed standard open-ended measures whereas a closed-ended analogue was used in Experiments 1B and 2B. Although reported as separate experiments, both Experiments 1A and 1B were run concurrently as one study, as were Experiments 2A and 2B, with participants randomly allocated to each experiment, as well as to experimental conditions within each experiment.

2.5. Experiment 1A: Method

2.5.1. Participants

A power analysis using the effect size observed in previous research using the same stimuli and experimental design (Johnson & Seifert, 1994; effect size obtained from means in Experiment 3A) indicated that a minimum of 69 participants were required (f = 0.39, 1-β = 0.80, α = 0.05). In total 78 US-based participants (28 females, aged between 19 and 62, M = 31.78, SD = 10.10) were recruited from AMT. Only participants with a Human Intelligence Task (HIT) approval rating greater than, or equal to, 99% were recruited for the experiment to ensure high-quality data without having to include attentional check questions (Peer, Vosgerau, & Acquisti, 2013). Participants were paid $2 and median completion time was 11 minutes.

2.5.2. Stimuli & Design

The experiment was programmed in Adobe Flash (Reimers & Stewart, 2007, 2015). Participants read one of 3 versions of a fictional news report about a warehouse fire that consisted of 15 discrete messages. The stimuli were identical to those used in Johnson and Seifert (1994, Experiment 3A). Fig. 2

illustrates how message content was varied across experimental conditions, as well as the message presentation format. The effect of correction information on reference to target (mis)information was assessed between groups; participants were randomly assigned to one of the 3 experimental groups: no correction (*n* = 32), correction (*n* = 21), and alternative explanation (*n* = 25).

Target (mis)information implying that carelessly stored oil paint and gas cylinders played a role in the fire, was presented at Message 6. This information was then corrected at Message 13 for the two conditions featuring a correction. Information implying that the fire was actually the result of arson (alternative explanation) was presented at Message 14; here the other two experimental groups learned that the storage hall contained stationery materials. The other messages provided further details of the incident and were identical in all 3 experimental conditions.

The questionnaire following the statements consisted of three question blocks: inference, factual, and critical information recall. Question order was randomized within inference and factual blocks, but not in the critical information recall block where questions were presented in a predefined order. Inference questions (e.g. “What was a possible cause of the fumes”) were presented first, followed by factual questions (e.g. “What business was the firm in?”), and after this the critical information recall questions (e.g. “What was the point of the second message from Police Investigator Lucas?”) were presented.

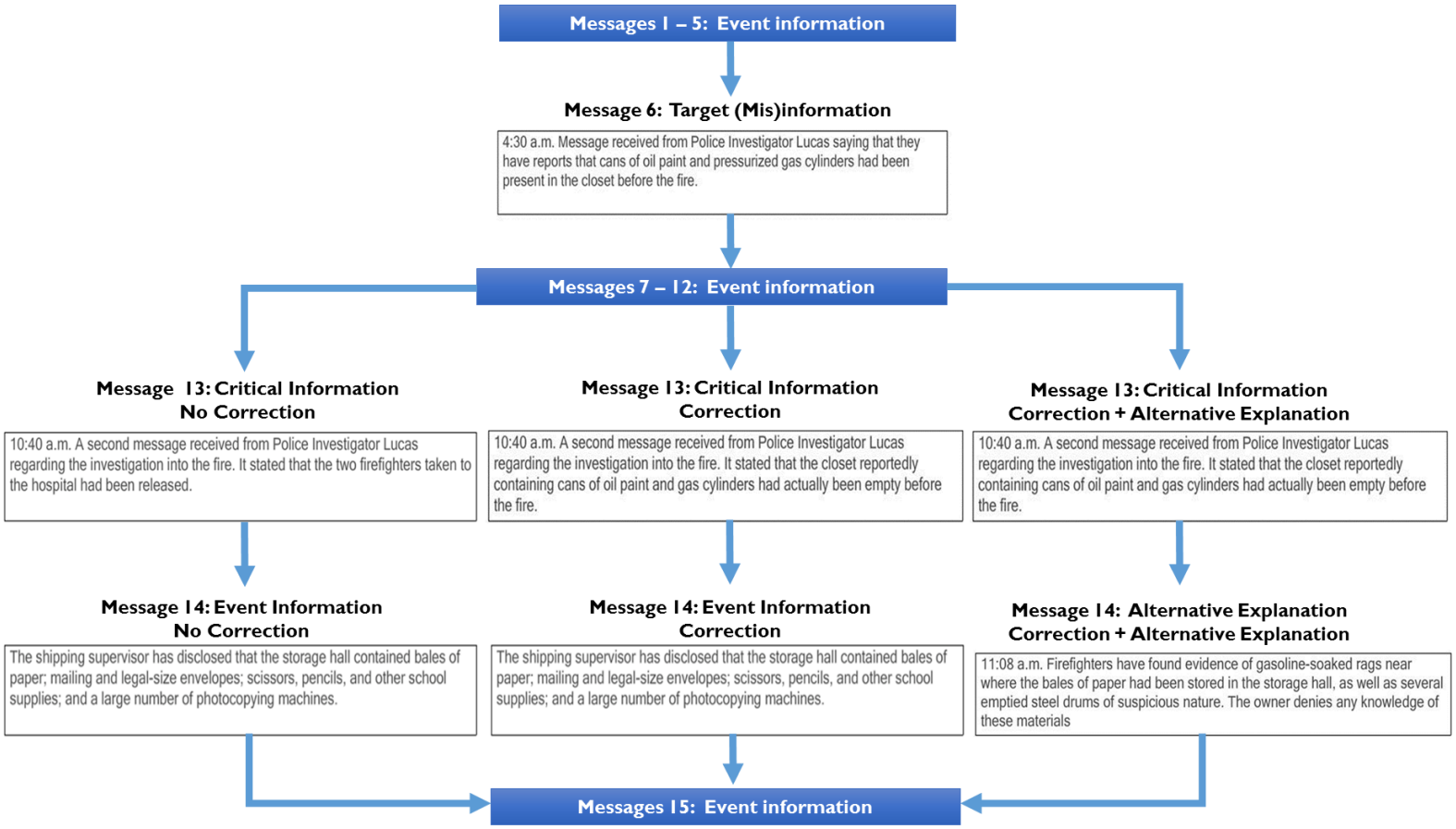

Figure 2 The continued influence effect task: Messages 1-5 provide general information about the event beginning with the fire being reported. Target (mis)information is presented at Message 6 and is then corrected for correction and correction + alternative explanation groups at Message 13. The correction + alternative explanation group then receive information providing a substitute account of the fire to 'fill the gap' left by invalidating the misinformation. This condition usually leads to a robust reduction in reference to misinformation.

There were three dependent measures: (1) reference to the target (mis)information in the inference questions, (2) factual recall, and (3) critical information recall. The first dependent measure assessed the extent to which the misinformation influenced interpretation of the news report, whereas the second assessed memory for the literal content of the report. The final measure specifically assessed understanding and accurate recall of the critical information that appeared at Message 13 (see Fig. 2). Although not all groups received a correction, the participants in all experimental groups were asked these questions so that the questions would not differ between the conditions. The stimuli were piloted on a small group of participants to check their average completion time and obtain feedback about the questionnaire. Following the pilot, the number of questions included in the inference and factual blocks was reduced from ten to six, because participants felt some questions were repetitive.

2.5.3. Procedure

Participants clicked on a link in AMT to enter the experimental site. After seeing details about the experiment, giving consent and receiving detailed instructions, they were told they would not be able to backtrack and that each message would appear for a minimum of 10 seconds before they could move on to the next message.

Immediately after reading the final statement participants were informed that they would see a series of inference-based questions. They were told to type their responses in the text box provided, giving as much detail as necessary writing in full sentences, writing at least 25 characters in order to be able to continue to the next question, and answering questions on the basis of their understanding about the report, and industrial fires in general. After this, they were informed that they would answer six factual question, which then followed. Next participants were instructed to answer the two critical information recall questions on the basis of what they remembered from the report. After completing the questionnaire participants were asked to provide their sex, age,

and highest level of education.

2.6. Results

2.6.1. Coding of responses

The main dependent variable extracted from responses to inference questions was ‘reference to target (mis)information’. References that explicitly stated, or strongly implied, that oil paint and gas cylinders caused, or contributed, to the fire were scored a 1 or were otherwise scored as 0. Table 2 shows an example of a response that was coded as a reference to target (mis)information and an example of a response that was not coded as such. There were several examples of references to flammable items but did not count as references to the corrected information. For example, stating that the fire spread quickly “Because there were a lot of flammable things in the shop”, would not be counted as a reference to the corrected information, as there is no specific reference to gas, paint, liquids, substances or the fact that they were (allegedly) in the closet. The maximum individual score across the inference questions was 6. Responses to factual questions were scored for accuracy; correct or partially correct responses were scored 1 and incorrect responses were scored 0. Again, the maximum factual score was 6. We also examined critical information recall, to check participant awareness of the correction to the misinformation or the control message, computed using two questions that assessed recall accuracy for critical information that appeared at Message 13. This meant that there were two correct responses depending on correction information condition. For participants in the no correction group the correct response was that the injured firefighters had been released from hospital and for the two conditions featuring a correction this was a correction of target (mis)information.

Table 2 Example of response coding in Experiment 1A

Question	Example of Response Scored 1	Example of Response Scored 0
Why did the fire spread so quickly?	Fire spread quickly due to gas cylinder explosion. Gas cylinders were stored inside the closet	The fire occurred in a stationery warehouse that housed envelopes and bales of paper that could easily ignite

Inter-coder reliability. All participants' responses to inference, factual, and critical information recall questions were independently coded by two trained coders. Inter-rater agreement was 0.88 and Cohen's K = 0.76±0.02, indicating a high level of agreement between coders, both of which are higher than the benchmark values of 0.7 and 0.6 (Krippendorff, 2012; Landis & Koch, 1977), and there was no systematic bias between raters, $\chi 2 = 0.29$, $p = 0.59$.

2.6.2. Inference scores

The overall effect of correction information on reference to target (mis)information was significant, F (2, 75) = 10.73, $p < .001$, η_p^2 = 0.22 [.07, .36]. Tukey corrected pairwise comparisons (shown in Panel A of Fig. 3) revealed that a correction, and a correction with an alternative explanation, significantly reduced reference to target (mis)information on inference questions.

A Bayesian analysis using BayesFactor package in R and default priors (Morey & Rouder, 2015) was performed to examine the relative predictive success of the comparisons between conditions. The BF_{10} for the first comparison was 28.93 – indicating strong evidence (Lee & Wagenmakers, 2014) in favour of the alternative that there is a difference between no correction and correction only groups. The BF_{10} for the comparison between no correction and alternative explanation groups was 209.03, again indicating very strong evidence in favour of the alternative. The $BF_{10\ was}$ 0.36 for the final comparison between correction only and alternative explanation groups indicating anecdotal evidence in favour of the null.

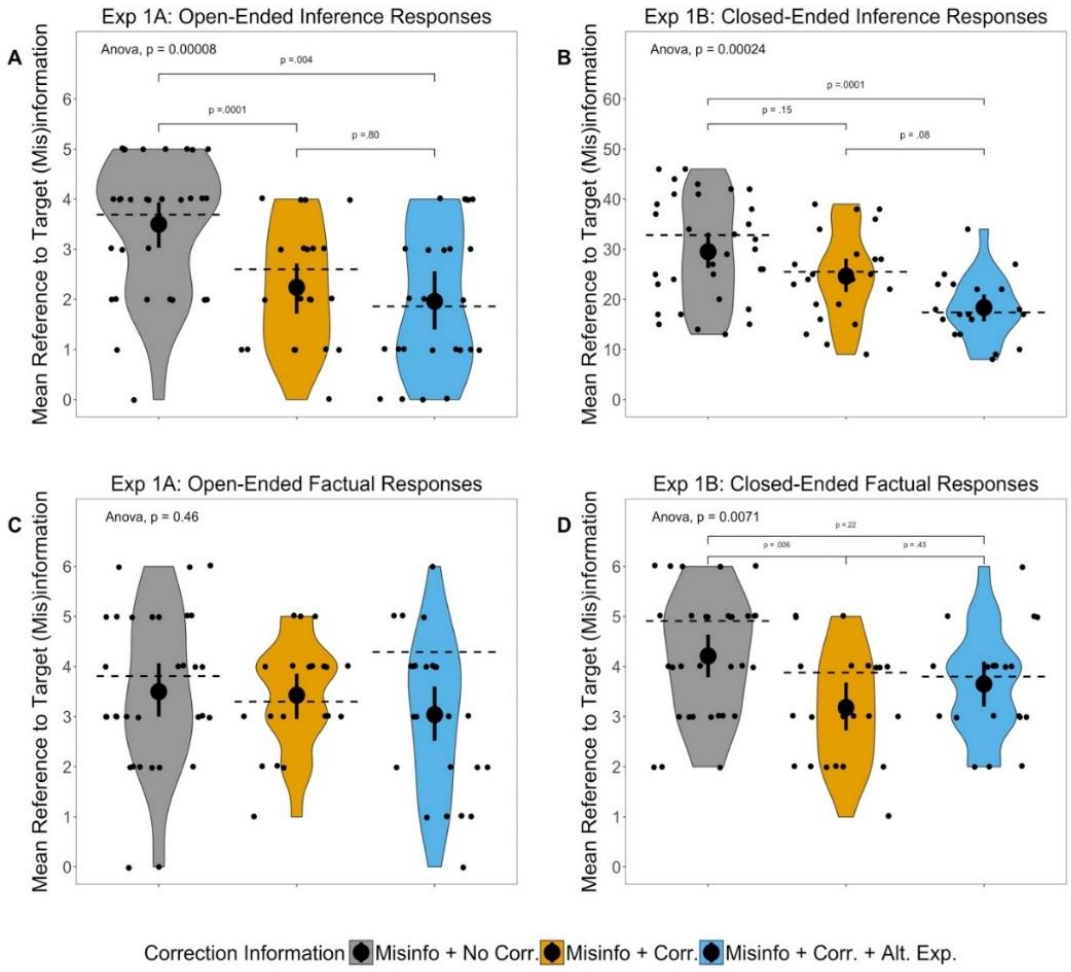

Figure 3 Effect of correction information on the number of (A) references to target (mis)information in Experiment 1A; (B) references to target misinformation in Experiment 1B; (C) accurately recalled facts in Experiment 1A; (D) accurately recalled facts in Experiment 1B. Error bars represent 95% confidence intervals of the mean. Brackets represent Dunnett's multiple comparison tests (which account for unequal group sizes) for significant omnibus tests. Dashed lines represent means after excluding participants who did not recall the critical information (i.e. scored 0 on the first correction recall question).

The Bayes factor analysis was mostly consistent with p-values and effect sizes. Both conditions featuring a correction led to a decrease in references to target (mis)information but the data for the two conditions featuring a correction

cannot distinguish between the null hypothesis and previous findings (i.e. that an alternative explanation substantially reduces reference to misinformation compared to a correction alone).

2.6.3. Recall Accuracy Scores

Factual responses were examined to establish whether differences in references to (mis)information could be explained by memory for the literal content of the report. Overall, participants accurately recalled a similar number of correct details across correction information conditions (Fig. 3, Panel C), and the omnibus test was not significant, F (2, 75) = 0.78, p =.46, η_p^2 = 0.02.

2.6.4. Response quality

Participants were required to write a minimum of 25 characters in response to questions. The number of characters written was examined as a measure of response quality. Participants wrote between 36-64% more on average than the minimum required 25 characters in response to inference (M = 69.45, SD = 40.49), factual (M = 39.09, SD = 15.85), and critical information recall questions (M = 66.72, SD = 42.76). There was - unsurprisingly - a positive correlation between time taken to complete the study and number of words characters written, r (76) = .31, p = .007.

2.7. Experiment 1B: Method

Experiment 1B examined the feasibility of converting open-ended questions to a comparable closed-ended form.

2.7.1. Participants

Seventy-five U.S. based (29 female, aged between 18 and 61, M = 34.31, SD = 10.54) participants were recruited from AMT. Participants were paid $2; the median completion time was 9 minutes.

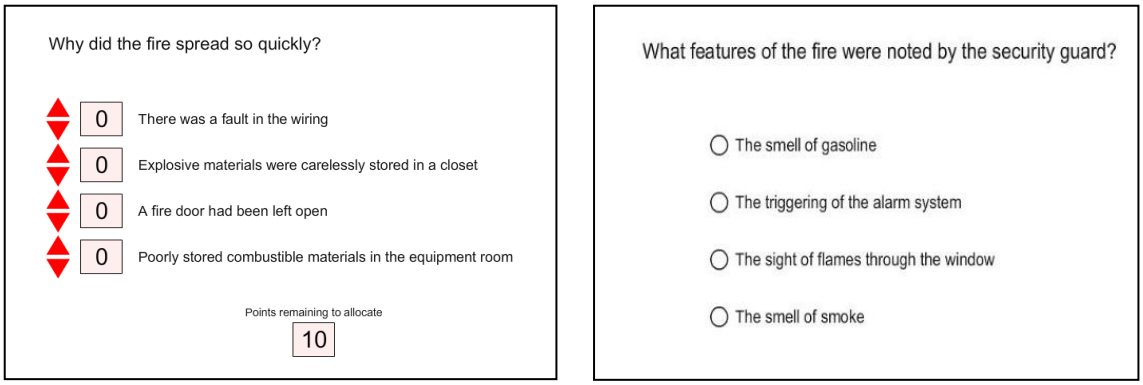

Figure 4 Screenshots of how inference (left) and factual (right) questions and response options were presented to participants in Experiment 1B. Participants used the red arrow features to allocate points to response alternatives to respond to inference questions. Factual questions were answered by selecting the 'correct' option based on the information in the report.

2.7.2. Design, Stimuli and Procedure

Experiment 1B used the same story/news feed stimuli and high-level design as Experiment 1; participants were randomly assigned to one of 3 experimental conditions: no correction (n = 33), correction only (n = 22), or alternative explanation (n = 20). The only difference between experiments was that closed-ended questions were used in the subsequent questionnaire. Fig. 4 shows how participants responded to inference and factual questions. For each question participants had 10 points which they could distribute across the 4 inference question response options to indicate which option/s best fit their understanding of the story. Response alternatives corresponded to 4 possible explanations for the fire. For example, when answering the question 'What could have caused the explosions?' participants could allocate points to a misinformation consistent option (e.g. 'Fire came in contact with compressed gas cylinders'), alternative explanation consistent option (e.g. 'Steel drums filled with liquid accelerants'), an option that was plausible given the story details but was not explicitly stated (e.g. 'Volatile compounds in photocopiers caught on fire'), or an option that was inconsistent with the story details (e.g. 'Cooking equipment

caught on fire'). The total number of points that could be allocated to a given explanatory theme was 60.

Response options were chosen in this way in order to give participants the opportunity to provide more nuanced responses than would be possible using multiple-choice or true/false alternatives. This approach allowed participants who were presented with misinformation and then a correction to choose an explanation which was consistent with the story but did not make use of the corrected information. If the *continued influence effect* is observed in response to closed-ended questions then the number of points allocated to misinformation consistent options in the conditions featuring a correction should be non-zero. Accuracy on factual questions was measured using 4AFC multiple-choice questions and participants responded by choosing the correct answer from a set of 4 possible options, which corresponded to the explanatory themes used for inference question response alternatives. Order of presentation of response alternatives for inference and factual questions was randomized across participants. Correction recall questions were open-ended and participants gave free-text responses in the same manner as in Experiment 1A.

2.8. Results

Individual inference, factual, and critical information recall scores (an analysis of the critical information recall responses is shown in the additional analyses in the Appendix D) were calculated for each participant. Since the maximum number of points that could be allocated to a given option explanation theme for each question was 10, the maximum inference score for an individual participant was 60. The maximum factual score was 6, and the maximum critical information recall score was 2. Critical information recall questions were open-ended, and responses were coded using the same criteria as in Experiment 1A.

2.8.1. Inference scores

A one-way ANOVA on reference to target (mis)information revealed a significant effect of correction information, F (2, 72) = 9.39, p <.001, η_p^2 = .21 [.05, .35]. Overall, the pattern of results for reference to target (mis)information in response to closed-ended questions was very similar to Experiment 1A (Fig. 3, Panel B). Although a correction with an alternative explanation significantly reduced reference to (mis)information, a correction on its own did not. The difference between the two conditions featuring a correction was also not significant.

The BF_{10} was 1.02 for the first comparison between the no correction and correction groups, indicating 'weak' or 'anecdotal' evidence in favour of the alternative, or arbitrary evidence for either hypothesis (Jarosz & Wiley, 2014). The BF_{10} was 250.81 for the second comparison between the no correction and alternative explanation groups indicating strong evidence for the alternative. The BF_{10} was 4.22 for the final comparison indicating substantial evidence in favour of the alternative.

The Bayes factor analysis was mostly consistent with p-values and effect sizes except that the Bayes factor for the comparison between correction and alternative explanation conditions suggested an effect whereas the p-value did not.

2.8.2. Recall accuracy scores

Analysis of factual scores indicated a significant difference between correction information groups, F (2, 72) = 5.30, p =.007, η_p^2 = .13 [.01, .26]. Fig. 3 (Panel D) shows that the allocation of points to the factually correct answer recalled from the report was significantly lower in the correction only condition than the no correction group but not the alternative explanation group. Poorer overall performance on factual questions for the correction only group was mainly attributable to incorrect responses to two questions. The first question asked

about the contents of the closet that had reportedly contained flammable materials, before the fire; the second asked about the time the fire was put out. Only a third (23% in the correction only and 25% in the alternative explanation group) answered this question correctly (i.e. that the storeroom was empty before the fire), whereas 86% of the no correction group correctly responded that oil paint and gas cylinders were in the storeroom before the fire. This is perhaps unsurprising: The correct answer for the no-correction condition ("paint and gas cylinders") was more salient and unambiguous than the correct answer for the other two conditions ("The storage closet was empty before the fire").

2.9. Discussion

The results for Experiments 1A and 1B suggest that both open- and closed-ended questions can successfully be used in online experiments with Amazon Mechanical Turk to measure differences in references to misinformation in a standard continued influence experiment. There was a clear *continued influence effect* of misinformation in all conditions of both experiments - a correction reduced but did not go anywhere near to eliminating, reference to misinformation on inference questions. In both studies references to (mis)information were significantly lower in the correction + alternative than in the no-correction condition, with correction condition between those two extremes (see Figure 3, Panels A and B). Although the pattern of significant results was slightly different (correction condition was significantly below no correction in Experiment 1A but not in Experiment 1B), this is consistent with the variability seen across experiments using the CIE, some of which have found a reduction in references to (mis)information following a correction (Ecker et al., 2010; Ecker et al., 2011a), and others of which have found no significant reduction (Johnson & Seifert, 1994).

With regard to motivation, we found that the vast majority of participants wrote reasonable responses to open-ended questions answers and were of a considerable length for the question, usually typing substantially more than the minimum number of characters required. We found that the absolute number of

references to the misinformation was comparable to that found in existing studies. That said, the open-ended questions had to be coded by hand, and for participants, the median completion time was 18% longer in Experiment 1A (11 minutes) than in Experiment 1B (9 minutes). This disparity in completion times only serves to emphasize that using closed-ended questions streamlines the data collection process compared to open-ended questions.

Taken as a whole, these findings show that reasonably complex experimental tasks that traditionally require participants to construct written responses can be implemented online either using the same type of open-ended questions or using comparable closed-ended questions.

2.10. Overview of Experiments 2A and 2B

The results of Experiments 1A and 1B are promising with regard to using open-ended questions in online research in general, and to examine phenomena such as the continued influence effect specifically. However, they had some limitations. The most salient limitation was in the sample size. Although the number of participants in each condition was comparable to those in many lab-based studies of the continued influence effect, the samples sizes were small. One of the advantages of using web-based procedures is that it is relatively straightforward to recruit large numbers of participants, so in Experiments 2A and 2B we replicated the key conditions of the previous studies with twice as many participants. We also pre-registered the method, directional hypotheses, and analysis plan (including planned analyses, data stopping rule, and exclusion criteria) prior to data collection; this information can be found at https://osf.io/cte3g/.

We also used this opportunity to include a second baseline condition. Several continued influence effect experiments have included control conditions of some form that make it possible to see whether references to the cause suggested by the misinformation following its correction are not only greater than zero but greater than the references to the same cause if the misinformation is

never presented. In this study, we did not believe that such a condition would be very informative because the strictness of the coding criteria meant that it would be unlikely that participants would spontaneously suggest paint or gas cylinders as contributing to the fire.[10]

Instead, Experiments 2A and 2B included a more directly comparable control group for whom a correction was presented without initial target misinformation. According to the mental-model-updating account of the continued influence effect, event information is integrated into a mental model that is updated when new information becomes available. Corrections may be poorly encoded or retrieved because they threaten the model's internal coherence (Ecker et al., 2010; Johnson & Seifert, 1994a; Johnson-Laird, 1980). If the continued influence effect arises because of a mental-model updating failure then presenting the misinformation only as part of a correction should not result in a continued influence effect because there will not be an opportunity to develop a mental-model involving misinformation. However, if participants continue to refer to misinformation for more superficial reasons (e.g. the cause presented in the misinformation is available in memory and is recalled without the context of it being corrected) then presenting the misinformation as part of the correction should lead to a comparable CIE to other conditions.

In these studies, we repeated the no-correction and correction conditions from Experiments 1A and 1B. In place of the correction + alternative explanation condition we had the no-mention condition, which was the same as the correction condition except we replaced the target (mis)information with a filler statement ("Message 6 - 4:30 a.m. Message received from Police Investigator Lucas saying that they have urged local residents to keep their windows and doors shut"). The wording of the correction message for this condition stated "a closet reportedly containing cans of oil paint and gas cylinders had actually been empty before the

[10] There is also a conceptual issue about whether references to the cause presented in the misinformation should be compared across correction and no-mention condition. In the former, the correction rules out the cause; in the latter, the cause is still possible.

fire" rather than referring to "the closet" so that the participants did not think they had missed some earlier information.

Beyond this, the general setup for Experiments 2A and 2B were the same as for Experiments 1A and 1B except for the following: We included an attention check (which appeared immediately after initial instructions and immediately before the warehouse fire report was presented) that tested participants' comprehension of the instructions via three multiple-choice questions. Participants were not excluded but were not allowed to proceed to the main experiment until they answered all three questions correctly, consistent with Crump et al.'s (2013) recommendations. As Adobe Flash, which we used for Experiment 1A and 1B, is being deprecated and is increasingly hard to use for web-based research, we implemented Experiments 2A and 2B using Qualtrics, which led to some superficial changes to the implementation. Most notable was that the point-allocation method for closed-ended inference questions required participants to type numbers of points to allocate, rather than adjusting the values using buttons. The sample size was also doubled in the second set of experiments[11].

2.11. Experiment 2A: Method

2.11.1. Participants

One-hundred and fifty-seven U.S. and U.K. based (66 female, aged between 18 and 64, M = 33.98, SD = 10.57) were recruited from AMT[12].

[11] Experiments 2A and 2B were actually conducted after the remaining experiments reported in this thesis. This is why the instructional attention checks used in these experiments differ to those used in experiments reported subsequently.

[12] Three of these participants were recruited from Prolific Academic. Data was collected from 159 participants but two participants were excluded because they gave nonsense answers to the questions (e.g. "because the wind is blow, love is fall, I think it is very interesting").

Participants took 16 minutes on average to complete the experiment and were paid $1.25[13].

2.11.2. Design and Procedure

Participants were randomly assigned to one of 3 experimental conditions: misinformation + no correction (*n* = 52), misinformation + correction (*n* = 52), or no misinformation + correction (*n* = 53).

[13] The modal completion time in Experiments 1 and 2 was below 10 minutes so the fee was reduced so that participants were paid the equivalent of the federal minimum wage in the U.S. ($7.25).

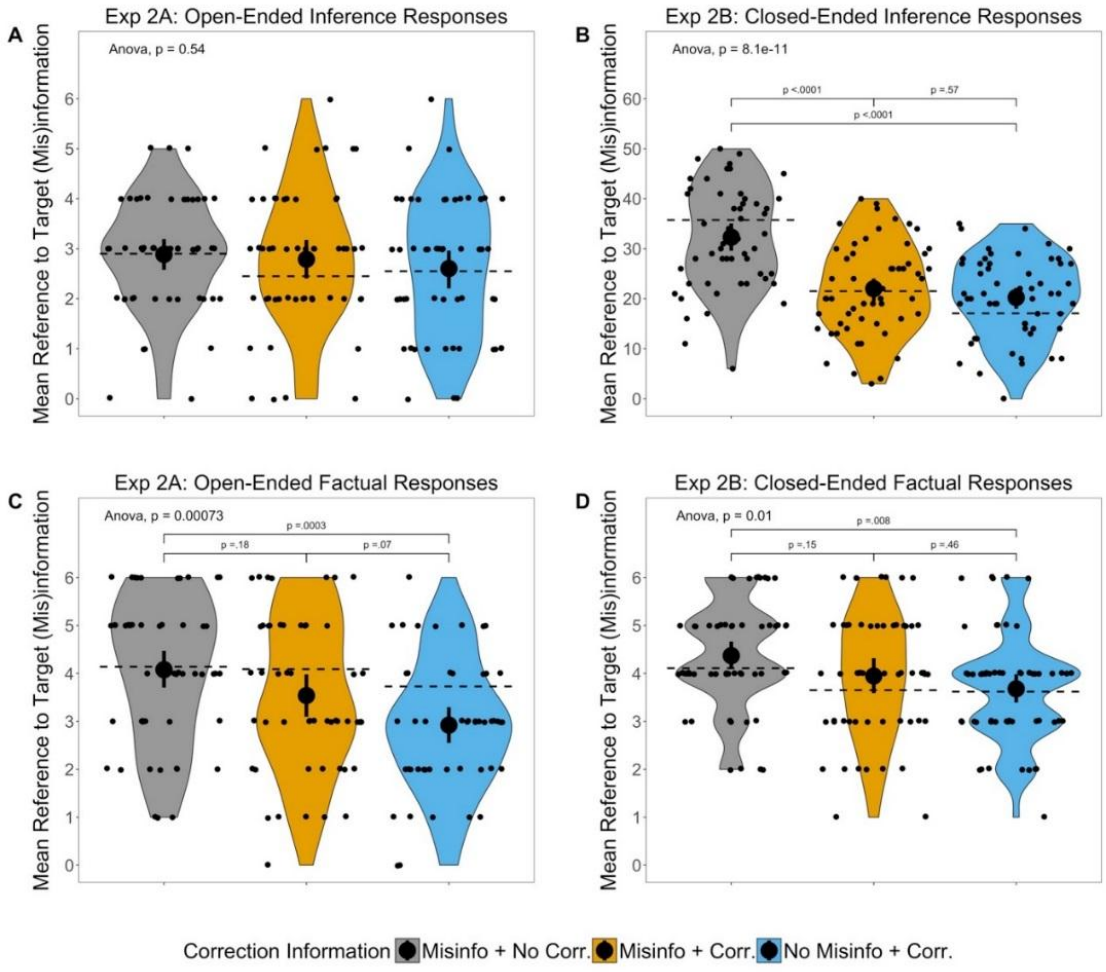

Figure 5 Effect of correction information on the number of (A) references to target (mis)information in Experiment 2A; (B) reference to target (mis)information in Experiment 2B; (C) accurately recalled factual details in Experiment 2A; (D) accurately recalled facts in Experiment 2B. Error bars represent 95% confidence interval of the mean. Brackets represent Tukey multiple comparison tests when the omnibus test was significant. Dashed lines represent means for restricted sample of participants who did not recall the critical information.

2.12. Results

Inter-coder reliability. Participants' responses to inference, factual, and critical information recall questions and were coded by one trained coder and 10% (N = 16) of responses were independently coded by a second trained coder. Inter-rater agreement was 1 and Cohen's K = 1±0, indicating, surprisingly, perfect agreement between coders.

2.12.1. Inference scores

Participants produced a similar number of references to target (mis)information across correction information conditions (Fig. 5, Panel A), and the omnibus test was not significant, F (2, 154) = 0.62, p = 0.54, η_p^2 = .01 [.00, .05]. Unlike Experiment 1A, a correction did not significantly reduce the number of references to target (mis)information relative to a control group who did not receive a correction. Moreover, participants who were not presented with initial misinformation but did receive a correction message made a similar number of misinformation references as participants who were first exposed to misinformation.

2.12.2. Recall accuracy scores

Participants' ability to accurately recall details from the report differed across correction information conditions (Fig. 5, Panel C), F (2, 154) = 8.12, p <.001, η_p^2 = .10 [.02, .18]. Tukey's test of multiple comparisons revealed that the group who received a correction without initial misinformation recalled significantly fewer details from the report than the group who saw uncorrected misinformation, while the other differences were non-significant, p's > .05.

2.12.3. Response quality

Participants wrote between 48-69% more on average than the minimum required 25 characters in response to inference (M = 80.76, SD = 56.38), factual (M = 48.15, SD = 24.86), and correction recall questions (M = 75.56, SD = 47.05). There was a positive correlation between time taken to complete

the study and number of characters written, r (155) = .34, p < .0001, showing that participants who took longer wrote more.

2.13. Experiment 2B: Method

2.13.1. Participants

One-hundred and sixty-six US and UK based (66 female, aged between 18 and 62, M = 35.04, SD = 10.36) participants were recruited from AMT[14]. Participants were paid $1.25; the average completion time was 13 minutes.

2.13.2. Design and Procedure

Experiment 2B used the same high-level design and procedure as Experiment 2A. Responses were closed-ended and responses were made in the same way as Experiment 1B. Participants were randomly assigned to one of 3 experimental conditions: misinformation + no correction (n = 54), misinformation + correction (n = 56), or no misinformation + correction (n = 56).

2.14. Results

2.14.1. Inference scores

There was a significant effect of correction information on references to target (mis)information for closed-ended measures (Fig. 5, Panel B), F (2, 163) = 26.90, p < .001, η_p^2 = .25 [.14, .35]. Tukey adjusted multiple comparisons further revealed that the group exposed to misinformation and its correction, and the group who saw only the correction without initial misinformation, resulted in significantly fewer references to target

[14] The recruited number of participants differed from the stopping rule specified in the pre-registration. In total 168 participants were recruited for the closed-ended condition due to an error. Ultimately, we decided to include the extra participants in the analysis rather exclude this data. However, responses from two participants were excluded; one because their participant took the HIT twice and another because they provided nonsense answers to the open-ended questions at the end of the study.

(mis)information than the uncorrected misinformation condition. The two groups who received correction information did not significantly differ.

2.14.2. Recall accuracy scores

Participants' responses to factual questions also showed a significant effect of condition (Fig. 5, Panel D), F (2, 163) = 4.70, p =.01, η_p^2 = .05 [.00, .13. Tukey's tests revealed that the factual responses from participants in the condition featuring a correction without initial misinformation were significantly lower than the group who saw uncorrected misinformation. The other differences were not significant (p's > .1). A closer inspection of the individual answers revealed that incorrect responses for the no misinformation + correction group were mainly attributable to the question asking about the contents of the closet before the fire.

2.14.3. Dropout analysis

Of the 375 people who started the study only 323 fully completed it (dropout rate 13%). Of those who completed the study 4 (1.23%) were excluded prior to analysis because they gave nonsense open-ended responses (e.g. " 21st-century fox, the biggest movie in theatre"). The majority of participants who dropped out did so immediately after entering their worker ID and before being assigned to a condition (41%). Of the remaining dropout participants who were assigned to a condition 27% were assigned to one of the open-ended conditions and dropped out during the first question block. A further 16% were assigned to one of the closed-ended conditions and dropped out when asked to answer the open-ended critical information recall questions. The remaining 14% were assigned to a closed-ended condition and dropped out as soon as they reached the first question block. The dropout breakdown suggests that many people dropped out because they were unhappy about having to give open-ended responses. Some participants who were assigned to closed-ended conditions dropped out when faced with open-ended questions despite the fact that the progress bar showed that they had almost completed the study.

2.15. Discussion

Experiments 2A and 2B again showed clear evidence of a continued influence effect. As in Experiments 1A and 1B, participants continued to refer to misinformation after it had been corrected. As with the previous two experiments, the effects of a correction differed slightly across conditions. This time the reduction in references to (mis)information was significant for the closed-ended questions, but not for the open-ended questions. As noted earlier, this is consistent with findings that a correction sometimes reduces references to misinformation relative to no correction, and sometimes does not (e.g. Ecker et al., 2010).

Experiments 2A and 2B also included a novel control condition in which participants were not exposed to initial misinformation but were exposed to its correction. Contrary to expectations, the new condition resulted in a statistically equivalent number of references to target (mis)information as the group who were exposed to both misinformation and its correction. This finding suggests that the continued influence effect might not reflect a model-updating failure, but rather, a decontextualized recall process.

2.16. General Discussion

Four experiments examined the feasibility of collecting data on the continued influence effect online, comparing the efficacy of using traditional open-ended questions versus adapting to use closed-ended questions. For both types of elicitation procedure, we observed clear continued influence effects: Following an unambiguous correction of earlier misinformation, participants continued to refer to the misinformation when answering inferential questions. As such, these studies provide clear evidence that both open-ended and closed-ended questions can be used in online experiments.

2.16.1. The continued influence effect

Across all four studies, we found that participants continued to use misinformation that had been subsequently corrected. This occurred even though a majority of participants recalled the correction. We found mixed

results when examining whether a correction had any effect at all in reducing references to misinformation. Experiments using similar designs have both found (Ecker, Lewandowsky, & Tang, 2010; Ecker, Lewandowsky, & Apai, 2011) and failed to find (Johnson & Seifert, 1994), an effect of a correction. Overall, we found limited evidence for an effect of a correction for the open-ended questions, but substantial evidence for an effect of a correction using closed-ended questions. For open-ended questions, it appears that any effect of a correction on reference to misinformation - at least using this scenario - is relatively small, and would be hard to detect consistently using the small sample sizes that have traditionally been used in this area. This may explain the variability in findings in the literature.

A correction with an alternative explanation appeared (at least numerically) to be more effective in reducing reliance on misinformation than a correction alone. Furthermore, given that Experiment 1B's results were actually more consistent with the original finding (Johnson & Seifert, 1994), the differences between past and present work are most likely unsystematic and therefore unrelated to the online testing environment or question type.

Finally, with regard to the main results, in Experiments 2A and 2B we found using a novel condition, that misinformation which was only presented as part of a correction had as much of a continuing influence effect as misinformation presented early in a series of statements and only later corrected. This has both theoretical and practical implications. Theoretically, it suggests that - under some circumstances - the CIE may not be the result of participants' unwillingness to give up an existing mental model without an alternative explanation (U. Ecker, Lewandowsky, & Apai, 2011; U. Ecker, Lewandowsky, Swire, et al., 2011; Johnson & Seifert, 1994). Instead, it might be that participants search their memory for possible causes when asked inferential questions, but fail to retrieve the information correcting the misinformation.

2.16.2. Open and closed questions and the CIE

The pattern of results in response to inference questions was qualitatively very similar across both open and closed-ended questions. This finding is particularly interesting in light of the fact that responses to open and closed questions might be supported by different underlying retrieval processes (Fisher, Brewer, & Mitchell, 2009; Ozuru et al., 2013; Shapiro, 2006). Crucially, the response options used in Experiments 1B and 2B required participants to make a more considered judgment than multiple-choice or yes/no questions, which may have encouraged recall rather than a familiarity-based heuristic. It is also interesting that participants still referred to the incorrect misinformation despite the fact there was another response option that was consistent with the report, although not explicitly stated.

Another important observation was that there was an effect of correction information on responses to closed factual questions but not open questions. The difference between conditions is significant because it was partly attributable to a question which probed participants' verbatim memory about the correction. Many participants in both conditions featuring a correction answered this question incorrectly despite the fact that options clearly distinguished between the correct and incorrect answers, given what participants had read. This question asked what the contents of the closet was before the fire so it not hard to see why participants who have continued to rely on corrected misinformation might answer this question incorrectly. The fact that there were differences between conditions highlights the importance of carefully wording questions and responses in order to avoid bias.

It is also worth noting that floor effects were not observed (i.e. misinformation was still influential for both groups that received a correction) despite the fact the current study did not include a distractor task and participants answered inference questions directly after reading the news report (so theoretically should have better memory for the report details).

A brief note on the use of closed-ended questions and response alternatives: there is the possibility that presenting a closed-list of options reminded participants of the arson materials explanation and inhibited

responses consistent with the oil paint and gas cylinders' explanation. The closed-list of options which repeated the misinformation could have increased its familiarity and made it more likely to be accepted as true (e.g. Ecker et al., 2011b). For the group that received a simple correction the other options had not been explicitly stated in the story. Participants may not have fully read or understood the question block instructions and therefore perceived the task as choosing the option that appeared in the story, irrespective of the correction. In contrast, participants in the alternative explanation group were able to better detect the discrepancy between the misinformation and its correction because of the option alluding to arson materials. Although the response alternatives provided a plausible response that was consistent with the details of the fire story, there were no options that made it possible to rule out that participants just do not consider the correction when responding. The response alternatives provided forced participants to choose from four explanations, which may have not reflected participants' understanding of the event, but nonetheless was the option that was most consistent with what participants had read. This explanation is also consistent with previous studies showing that the response options chosen by the researcher can be used by the participants to infer which information the participant considers relevant (Schwarz et al., 1985, Schwarz et al., 1991).

2.16.3. Open and closed-ended questions in web-based research

As well as looking directly at the continued influence effect, we also examined the extent to which participants recruited via Amazon Mechanical Turk could provide high-quality data from open-ended questions. We found high levels of diligence - participants typed much more than required, in order to give full answers to the questions, spent more time reading statements than required, and - with a small number of exceptions - engaged well with the task and attempted to answer the questions set.

We found that drop-out increased where participants had to give open-ended responses. This may suggest that some participants dislike typing open-ended responses, to the extent that they choose not to participate. (It could be that participants find it too much effort, or that they do not feel

confident giving written answers, or that it feels more personal having to type an answer oneself.) Alternatively, it may be that some participants because of the device they are using would struggle to provide open-ended responses, and so drop out when faced with open-ended questions. Either way, it is striking that we had over 4% of participants in Experiment 2B who read all the statements, and gave answers for all the closed-ended questions, but dropped out in the final few questions when asked to type their response for the final two critical information awareness questions. There are ethical implications of having participants spend ten minutes on a task before dropping out, so the requirement for typed answers should be presented prominently before participants begin the experiment.

We found that participants' recall of the correction to the misinformation was worse than in previous lab-based studies. We found that only a little over half the participants across conditions in our studies correctly reported the correction when prompted. This figure is poor when compared to 95% (correction only) and 75% (alternative explanation) found in Johnson and Seifert's (1994; Exp 3A) laboratory-based experiment. It is possible that this is the result of poor attention and recall of the correction, but we believe it is more likely that it is a response issue where participants had retained the information but did not realise that the questions was asking them to report it when asked about whether they were aware of any inconsistencies or corrections. (In other unpublished research, we have found that simply labelling the relevant statement "Correction:" greatly increased participants' reference to it when asked about any corrections.) Although this did not affect the continued influence effect, in future research we would recommend making instructions around the correction-awareness question particularly clear and explicit. This advice would, we imagine, generalise to any questions which may be ambiguous, and which require a precise answer.

In choosing whether to use open-ended questions or to adapt them to closed-ended questions for use online, there are several pros and cons to weigh up. Open-ended questions allow a consistency of methodology with traditional lab-based approaches - meaning there is no risk of participants switching to using different strategies or processes as they might with closed

questions. We have shown that participants generally engage well and give good responses to open-ended questions. It is also much easier to spot and exclude participants who respond with minimal effort, as their written answers tend to be nonsense or copied and pasted from elsewhere. For closed-ended responses, attention or consistency checks or other measures of participant engagement are more likely to be necessary. That said, closed-ended questions are, we have found, substantially faster to complete, meaning researchers on a budget could test more participants or ask more questions, they require no time to manually code, participants are less likely to drop out with them, and - at least in the area of research used here - they provide comparable results to open-ended questions.

2.17. Conclusion

In conclusion, the *continued influence effect* can be added to the existing list of psychological findings that have been successfully replicated online. Data obtained online are of sufficiently high quality to examine original research questions and are comparable to data collected in the laboratory. Furthermore, the influence of misinformation can be examined using closed-ended questions with direct choices between options. Nevertheless, as with any methodological tools researchers should proceed with caution and ensure that sufficient piloting is conducted prior to extensive testing. More generally, the research reported here suggests that open-ended written responses can be collected via the web and Amazon Mechanical Turk.

3 Explanatory corrections to misinformation across multiple scenarios

3.1. Chapter Overview

As noted in the introductory chapter to this thesis, previous CIE research has found that two of the most effective strategies for reducing reliance on misinformation are to provide an explicit pre-exposure warning about the possibility of being misled, or provide a plausible alternative explanation for the corrected information (e.g. Lewandowsky et al., 2012). The combination of pre-exposure warnings and provision of an alternative explanation can further reduce the CIE but still fails to eliminate it completely[15] (Ecker et al., 2010). This raises questions about the capacity for pre-exposure warnings and alternative explanations to be effective strategies for reducing the CIE in the real-world. An additional issue is that implementation of these strategies may not always be possible outside the lab. It is not often possible to provide a single, plausible alternative explanation to replace the misinformation. For instance, the claim that there is a causal link between autism and the MMR vaccination might have been more successfully corrected had the causes of autism been better understood and explained to the public (e.g. Ecker et al., 2014). Likewise, it may be difficult to provide timely pre-exposure warnings about the possibility of being misled.

Given that neither pre-exposure warnings nor alternative explanations fully eliminates the CIE, and that their implementation may only be possible on occasion, it is of practical importance to develop new strategies for

[1515] Lewandowsky et al., (2012) also discuss a third factor – repetition of the correction. There is mixed evidence regarding the effectiveness of repeated corrections for reducing the impact of misinformation (Ecker et al., 2017; Ecker et al., 2011), so it is not included as a factor here.

reducing continued reliance on misinformation. Establishing novel strategies for enhancing the effectiveness of corrections to misinformation not only has clear implications for counter-misinformation campaigns, but also may help to distinguish between mental-model updating and retrieval failure accounts of the CIE (discussed in detail in Chapter 1 of this thesis).

One potential strategy for reducing the CIE is to explain where or how misinformation initially occurred and is therefore no longer valid. There are three main reasons that this strategy may be effective at reducing or eliminating reliance on misinformation. First, experimental studies in the legal domain suggest that explaining why invalid information is unreliable can reduce its impact on later judgments (e.g. Kassin & Sommer, 1997). Second, there is evidence that corrections which address the conversational implications (i.e. that address assumptions about the intended meaning of the statement as opposed to the literal meaning) of contradictory statements are more effective than negations at reducing the CIE (Bush et al., 1994). Third, refutations that provide sufficient explanation to suggest updating is necessary increase the likelihood of text representation revisions (e.g. Rapp & Kendeou, 2007).

The three experiments reported in this chapter examine whether explaining how misinformation occurred (e.g. unintentional error or an intentional lie) can enhance a correction's effectiveness relative to a negation. The results of these experiments provide evidence that corrections which explain how misinformation occurred are no more effective at reducing the CIE than corrections which negate the misinformation. These results demonstrate that the CIE is not a necessary consequence of the correction of misinformation, but instead, may be constrained to ambiguous scenarios in which corrections leave open the possibility that the explanation offered by misinformation is still valid. This finding has both practical and theoretical implications. Practically, it suggests that misinformation is more likely to have an impact when circumstances are ambiguous and corrections do not sufficiently invalidate misinformation explanation. Theoretically, the findings suggest two things: First, that corrections do not necessarily have to fill the explanatory gap left by invalidating misinformation with an alternative

explanation (see Chapter 1 for further discussion of alternative explanations and Chapter 2 for experimental evidence that alternative explanations do not eliminate the CIE); Second, that corrections which address the conversational or literal implications of contradictory information are equally effective. These findings also suggest that corrections that rule out the causal explanation offered by misinformation can substantially attenuate the CIE.

3.2. Introduction

The CIE has proved difficult to eliminate (Ecker et al., 2014; Lewandowsky et al., 2012; Seifert, 2002, 2014). Continued influence studies have invariably found that corrections to misinformation reduce but rarely eliminate the influence of misinformation completely (see also the results reported in Chapter 2). In a typical CIE experiment, participants read a description of an unfolding event and then answer a series of causal inference questions about the scenario. A common scenario presented to participants involves a fire at a stationery warehouse in which initial misinformation implies that carelessly stored oil paint and gas cylinders are a likely cause of the fire (Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988). Participants for whom misinformation is later corrected (i.e. there were no oil paints and gas cylinders present) often continue to use the corrected information to answer subsequent causal inference questions (e.g. what could have caused the explosions?).

Previous CIE studies have identified that two of the most effective strategies for reducing the CIE are to provide pre-exposure warnings about the possibility of being misled or provide a plausible alternative explanation for the misinformation (e.g. Ecker et al., 2011a; Ecker et al., 2010; Johnson & Seifert, 1994), 2). The CIE can be further reduced – but not eliminated – by combining these strategies (Ecker et al., 2010). Given that it may not always be possible to provide a single plausible causal alternative explanation, and that it may be difficult to provide timely pre-exposure warnings, it is necessary to establish other means of reducing or eliminating the CIE.

One reason that corrections do not eliminate reliance on misinformation in CIE studies could be that laboratory implementations of

corrections do not correspond with the types of corrections that people encounter in everyday life. Continued influence experiments are usually necessarily sparse in the information they present and participants intrinsically have the opportunity to construct their own interpretations from the information presented. One potential account of the CIE is that it occurs in lab-based scenarios because participants have no reason to place more weight in the correction than on misinformation. This is because they do not know why erroneous information was initially presented nor what led to its correction. For instance, the correction used in the warehouse fire story asserts that the closet reportedly containing flammable liquids was actually empty before the fire, but does not explain why someone thought flammable liquids were there in the first place, and then changed their mind (e.g. Johnson & Seifert, 1994). When people encounter corrections in the real-world they are often presented with much richer information and usually receive an explanation for why misinformation is incorrect. Provided that the misinformation is not already congruous with a pre-existing world-view or attitude (e.g. belief in the causal link between the MMR vaccination and autism may be congruous with a distrust of big pharmaceutical companies), offering an explanation for why misinformation is wrong should allow them to place more weight on the correction than on misinformation. For instance, learning that a scientific article was retracted because the data were fabricated or that a news story is corrected because of a proof-reading error should help people to disregard prior incorrect information in favour of newer information.

Accordingly, one of the reasons that corrections may be ineffective in CIE studies could be that they do not provide sufficient grounds to disregard earlier information. The introduction to this chapter provides an overview of research on how attempting to remove the influence of previously presented information functions when participants are a) told to ignore previously presented information in courtroom simulation studies, b) shown further information that discredits earlier misinformation, and c) received an explicit correction to misinformation.

3.2.1. Instructions to disregard prior information

Applied research examining juror decision-making has yielded similar findings to CIE studies. These studies often show that inadmissible evidence (i.e. information that cannot be presented to the jury) has a reliable impact on judgments or guilt and verdicts despite corrective judicial admonition (Carretta & Moreland, 1983; Fein, McCloskey, & Tomlinson, 1997; Kassin & Sukel, 1997; Steblay, Hosch, Culhane, & McWethy, 2006; Thompson, Fong, & Rosenhan, 1981). The main factor that is thought to increase juror compliance with instructions to disregard inadmissible evidence is whether the judge provides a rationale for the inadmissibility ruling (see Steblay et al., 2006 for review of studies on instructions to ignore inadmissible information). For example, Kassin and Sommers (1997) found that mock-jurors who learned that a key piece of incriminating (wire-tap) evidence was inadmissible because it would harm the defendant's right to due process (e.g. a taped confession was secured without a warrant) were more likely to convict a defendant than mock-jurors who were told that the evidence was unreliable (e.g. the tape was inaudible). Later work by Sommers and Kassin (2001) also showed that participants who selectively complied with judicial instructions (i.e. disregarding inadmissible information when it was unreliable but not when it violated due process) also scored high on 'need for cognition' - a personality factor reflecting an inclination toward effortful cognitive activities (Cacioppo & Petty, 1982).

One explanation for these findings could be that there is a stronger motivation to disregard earlier information when the reasons given state why the information is unreliable than when they leave some room for the information to still be true (see Schul & Mayo, 2014, for a similar argument). However, it is worth noting that the instruction to disregard information because of due process situation differs from the CIE, because due process conditions do not speak to the truth or otherwise of information. If anything, these situations may suggest the to-be-disregarded information is still relevant to the judgment or decision at hand. Therefore, providing a rationale for a

correction that renders the misinformation irrelevant should reduce its influence.

3.2.2. Discrediting evidence

Experimental studies with a legal flavour also suggest that people more readily revise their beliefs about witness testimony when the witness' credibility is called into question (Hatvany & Strack, 1980; Lagnado & Harvey, 2008; Schul & Manzury, 1990; Weinberg & Baron, 1982; see Whitley, 1987, for meta-analysis of discredited eyewitness testimony studies). For example, Lagnado and Harvey (2008) found that people relied less on eyewitness identification testimony when they were told the witness had a 'longstanding grudge' against the suspect.

In a similar vein, Lagnado, Fenton, and Neil (2013) found that discrediting an 'intentionally deceptive' alibi had more of an impact on guilt ratings than a discredit which characterised the alibi as an 'honest mistake'. They compared guilt ratings for a suspect whose grandmother had provided alibi evidence which was subsequently discredited as an honest mistake (*the grandmother could not remember the night in question*) or as deception (*there was evidence that the grandmother was somewhere else on the night in question*). Guilt ratings were measured at baseline, after the alibi information, and again after the alibi was discredited. Findings showed that guilt ratings were significantly reduced immediately after the alibi information but then increased again after the alibi was discredited. Participants also rated the 'honest mistake' alibi as more believable than the 'deception' alibi.

These findings demonstrate that people are sensitive to the manner in which information is discredited and are more likely to revise their initial belief when given reason to question the initial misinformation. The findings also suggest that people are more likely to discount earlier information if they discover that deception was involved. Compared to legal reasoning studies, Lagnado and Harvey (2008) and Lagnado et al.'s (2013) studies were sparse in information, which further indicates that people make intuitive judgments

about the reliability of information based on how credible they perceive the source of that information to be.

3.2.3. Distrust

One way that explanations can serve to enhance the effectiveness of a correction is that they elicit distrust in the misinformation. One CIE study found that more than a third of participants in their study thought the correction was a cover-up for the truth when asked to describe why they believed there had been a correction to earlier information (Guillory & Geraci, 2010). This suggests that distrust in the correction to misinformation could contribute to the ineffectiveness of corrections. Distrust could either be generated endogenously within the scenario - causing the participant to question why the source of misinformation would contradict themselves – or exogenously, such that the participant questions why the experimenter provided information that they later said was irrelevant.

Distrust in the source of the correction could also be a reason that negation of misinformation is insufficient to eliminate its influence (e.g. Johnson & Seifert, 1994). Guillory and Geraci (2013) tested this idea by varying the credibility of the source of the correction. They gave participants a scenario in which the misinformation alleged that a politician running for re-election had accepted a bribe. This information was subsequently corrected by sources that varied in terms of their trustworthiness (i.e. willingness to convey accurate information) and expertise (i.e. capacity to convey accurate information). Unsurprisingly, findings showed that a correction was more effective if it was issued by a highly trustworthy source (e.g. a religious leader or the politician's opponent) than when it came from a source low in trustworthiness (e.g. the politician's wife). Source expertise was not sufficient to reduce reliance on initial misinformation.

Legal decision-making studies have also found that arousing suspicions about the reasons for introduction of misleading or inadmissible information can reduce its impact on later judgments. Fein, McCloskey, and Tomlinson (1997) asked participants to play the role of jurors in a murder trial.

Prior to reading the trial transcript participants read a (fictional) newspaper article about the murder that provided evidence against the suspect. One group also read a newspaper article in which the defendant's attorney called into question the press' motives for printing incriminating information. All of the groups then indicated whether or not they would convict the defendant. Findings showed that participants who were led to be suspicious[16] about the introduction of pre-trial publicity information were no more likely to convict the defendant, and substantially less confident in the defendant's guilt, than participants who did not receive pre-trial publicity. These results were also replicated for a case in which a key witness' testimony was ruled inadmissible. Together these findings suggest that the CIE could be reduced when the correction provides a reason to disbelieve the misinformation by calling into question its relevance and validity.

3.2.4. Detailed refutations

Research on detailed refutations in text comprehension is another strand of research that suggests explanations for why misinformation is wrong might reduce its continuing impact. In text comprehension research, refutation texts include statements that explicitly refute incorrect beliefs and explain correct principles. For instance, Rapp and Kendeou (2007) presented participants with short stories that included behavioural evidence for particular traits (e.g. that the protagonist was a messy person), which was either supported with an additional trait statement, refuted in a non-explanatory way as incorrect, or refuted with an explanation of why an incorrect interpretation of the behavioural evidence was possible. Findings showed that the explanation-based refutations resulted in more successful revision than the non-explanatory counterpart. In a similar vein, Swire, Ecker, and Lewandowsky (2017) found that level of explanatory detail facilitates belief change following exposure to myths. Participants read myths – such as "Liars sometimes give themselves away by physical 'tells' such as looking to the

[16] The authors define suspicion as 'actively entertaining multiple, plausible rival, hypotheses about the motives underlying behaviour and considering the notion that the person is trying to hide something that has the potential to discredit the apparent meaning of the behaviour' (p. 1217).

right or not looking you in the eye" – as well as facts of unclear veracity. Myths were subsequently corrected whilst the facts were affirmed with varying degrees of explanatory detail provided in corrections. Findings showed that providing a greater amount of explanatory detail promoted more sustained belief change over a three-week period than non-explanatory corrective information. These findings collectively suggest that explanations for why misinformation is incorrect may reduce or eliminate the CIE by encouraging deeper and more elaborate processing of the correction which later enhances strategic retrieval of that information.

3.2.5. Conversational implications

Another way that corrections to misinformation can be improved is if they address the conversational implications of the contradiction between misinformation and a correction. The CIE may be understood by the pragmatic inferences people draw about the conversational implications of initial misinformation (Grice, 1975; Seifert, 2002, 2014). Grice's (1975) account of conversational logic suggests that corrections *ought* to be challenging for our interpretation of human generated information. On this view, the contradiction may be poorly understood when the correction to misinformation only addresses the literal content of misinformation (e.g. there were no flammable liquids on the premises) and not the conversational implications of misinformation (i.e. why was the information conveyed in the first place). More specifically, the Gricean perspective asserts that conversational conventions are important for assessing the truth (*maxim of quality*) and relevance (*maxim of relation*) of statements. Therefore, corrections that do not address the maxims of relation and quality may be particularly difficult for people to understand. This is because people make inferences that consider pragmatic information as well as logic. Logically, when a statement is later corrected, the original statement is expected to be disregarded and replaced with an updated version of events. Pragmatically, the original statement and its correction usually have the same status – they are the reported beliefs of an individual.

Without a compelling justification for the original error, participants can come up with plausible explanations that give the original misinformation more weight. For instance, people might entertain the possibility that the person issuing the correction was paid off or that a superior told them to lie in issuing the correction because the misinformation made someone look bad. There may also be a meta-narrative issue, in which the participants question why the experimenter presented incorrect information given they know it is wrong which renders the information relevant in the eyes of the participant (Bless, Strack, & Schwarz, 1993; Schwarz, 2014). Corrections may therefore be problematic because they imply that the speaker believed that both the misinformation and its correction are true and accurate. Corrections which violate conversational principles and cause problems for interpretation thereby reinforce the validity of the misinformation. Several experimental studies have demonstrated the impact of conversational conventions on reasoning (Igou & Bless, 2003, 2005; Krosnick, Li, & Lehman, 1990; Schwarz, Strack, Hilton, & Naderer, 1991; Tetlock, Lerner, & Boettger, 1996).

There is also evidence from studies adopting the CIE approach that addressing the conversational implications of a correction can reduce reliance on misinformation. Bush, Johnson, and Seifert (1994) examined whether corrections that addressed the conversational implications of the contradiction reduced reliance on misinformation more than a correction that addressed only the literal implications. Participants read the warehouse fire scenario (described earlier) and either received a correction that explained why misinformation was uninformative ("the storeroom had actually contained cans of coffee and soda canisters and not flammable liquids") or no longer relevant ("a delivery of paint and gas cylinders was expected but never arrived)"[17]. Findings showed that both types of 'explanatory corrections' were more effective at reducing the impact of misinformation when compared to a simple negation. Interestingly, however, Bush et al., also found that ruling out the involvement of the misinformation ("there was clear evidence that no paint or gas were ever on the premises"), without providing an explanation, fared

[17] It may be worth noting that both explanations could be classed as 'poor quality information'. These explanations instead appear to contrast misinterpretation of the situation and a communicative error.

worse than a correction on its own (but see our results below). These findings further indicate that the person issuing the correction must explain why the original information should no longer be believed in order to increase the chances the correction is understood.

3.3. Overview of Experiments

The findings discussed in the introduction to this chapter suggest that providing an explanation for how the misinformation initially occurred could reduce its post-correction influence. The three experiments reported in this chapter were designed to examine whether corrections that explain why misinformation is incorrect are more effective at reducing the CIE than a correction which negates misinformation. Two types of explanatory corrections were used in the experiments reported here[18]. The first explained misinformation as an (unintentional) error and the second explained misinformation as an (intentional) lie. These explanations should address the conversational implications of the contradiction between misinformation and the correction and help participants to place more weight in the correction. The explanatory corrections used in the present set of experiments are modelled on situations in which testimony is discredited by showing that the witness is a liar or had made a misidentification. All three experiments used a variant of the CIE task in which information about a description of an unfolding event is presented as a series of discrete messages. The experiments were performed both online (see Chapter 2 introduction for discussion of conducting web-based CIE experiments) and in the laboratory.

3.4. Experiment 3

In light of the literature discussed in the introduction to this chapter, and the results of the experiments reported in Chapter 2 of this thesis, there were two main predictions. First, that a correction would reduce, but not eliminate, reference to target (mis)information compared to no correction. That is, the number of references to target (mis)information would be significantly lower in

[18] The two types of explanations used in these experiments represent an infinitesimally small subset of the possible explanations one could reasonably provide.

corrected misinformation groups than the uncorrected group, but would still be greater than zero. Second, a correction would more effectively reduce reliance on misinformation when the correction compellingly explains where the target (mis)information originated. This means that the number of references to target (mis)information will be significantly lower in the explanatory correction groups than the non-explanatory correction group.

There was also a tentative prediction that corrections in the two explanatory conditions would differ in terms of their scope for reducing the CIE. It was expected that people would be more sensitive to a correction explained as a lie than a correction explained as an error. This tentative prediction is made on the basis of work showing that human beings are social animals who are well prepared to detect deception (Schul, Mayo, & Burnstein, 2004). Furthermore, as noted earlier, work on discrediting alibi testimony has shown that people find an intentionally deceptive alibi less believable than one that was an honest mistake (Lagnado et al., 2011). Moreover, some evolutionary psychologists have argued that the ability to detect dishonesty facilitates reasoning in social contexts (e.g. Cosmides & Tooby, 1992; Hartwig & Bond, 2011). Following this line of reasoning, correcting a proven lie might be more effective than correcting an inadvertent error.

3.5. Method

3.5.1. Participants

A power analysis indicated a minimum of 280 participants would be required to detect a medium-sized effect (f = 0.25, 1- β = 0.95, α = 0.05). In total 365 U.S. based (169 female, 196 male, M_{age} = 39.38, aged between 21 and 72) participants were recruited via Amazon Mechanical Turk. Only participants with a human intelligence task (HIT) approval rating great than, or equal to 99%, were recruited for the experiment in order to further safeguard against poor quality data. More participants than required were initially recruited in anticipation that it would be necessary to exclude a substantial number of participants. Of this number, 126 (35%) participants failed an

attention check question included in the inference and recall question block[19]. The participants who failed the attention check question were ultimately not excluded because their exclusion did not change the results. Participants were paid $1.50 (approx. £1.07) and took an average of 18 minutes to complete the study.

3.5.2. Design

The effect of correction information on reference to target (mis)information was assessed between-groups; participants were randomly assigned to either the no correction (*n* = 95), correction (*n* = 95), correction + error explanation (*n* = 79), or correction + lie explanation (*n* = 96) groups.[20]

3.5.3. Stimuli

The stimuli were generated using Qualtrics (Qualtrics, Provo, UT). Figure 6 illustrates the content of the messages and demonstrates how they were presented to participants in Experiment 3. Participants read one of four versions of a fictional news report about a fire at a stationery warehouse consisting of 12 individually presented statements. The stimuli were modified from those used in previous research (Guillory & Geraci, 2010; Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988), and from the experiments reported in Chapter 2 (Experiments 1A and 1B, 2A and 2B), in the following ways. First, the number of messages was reduced from 15 to 12 to streamline the task for web-based testing. Reducing the number of messages may also increase the probability that participants would recall the correction and decrease the possibility that people use misinformation to answer the inference questions because they do not remember the correction but do remember the misinformation. Second, the report was presented in the style of a series 'Tweets' from the social media platform Twitter (as in Hardwicke, 2016). The 'Tweets' originated from the same fictional news outlet, called

[19] This question was randomly interspersed in the inference and factual recall question block. The question asked participants to indicate approximately how many Twitter messages they had just read and in parentheses prompted participants to choose the option 'more than 40'. Participants would then be excluded on the basis that they did not follow the instructions in parentheses.

[20] Group sizes were unequal to a programming error in Qualtrics.

“news now” and each message was no longer than 140 characters (see Figure 7). The ‘Twitter’ presentation format was used in order to make the news report appear more authentic and resemble a breaking news report. The third modification was the additional event information presented either side of misinformation and the correction was re-written so that it was not congruent (although not incongruent) with causal explanation offered by misinformation. Statements such as “Three warehouse workers working overtime, have been taken to St. Columbus Hospital, due to smoke inhalation” were used in place of statements like: “Two firefighters are reported to have been taken to the hospital as a result of breathing toxic fumes that built up in the area in which they were working”. The reason for this change to the stimuli was to avoid participants reasoning that ‘toxic fumes’ would not have been present if oil paint and gas cylinders were not involved. Using misinformation congruent statements to make up the rest of the scenario could be one reason that some studies have found a strong continued reliance on misinformation (e.g. Johnson & Seifert, 1994). There was one statement that was congruent with the explanation that oil paint and gas cylinders were causally involved in the fire (“Thick, oily smoke + sheets of flames hinder firefighters’ efforts, intense heat has made the fire difficult to bring under control”). This statement was not changed so that there was at least some opportunity for participants to develop a model in which the misinformation provided a causal explanation.

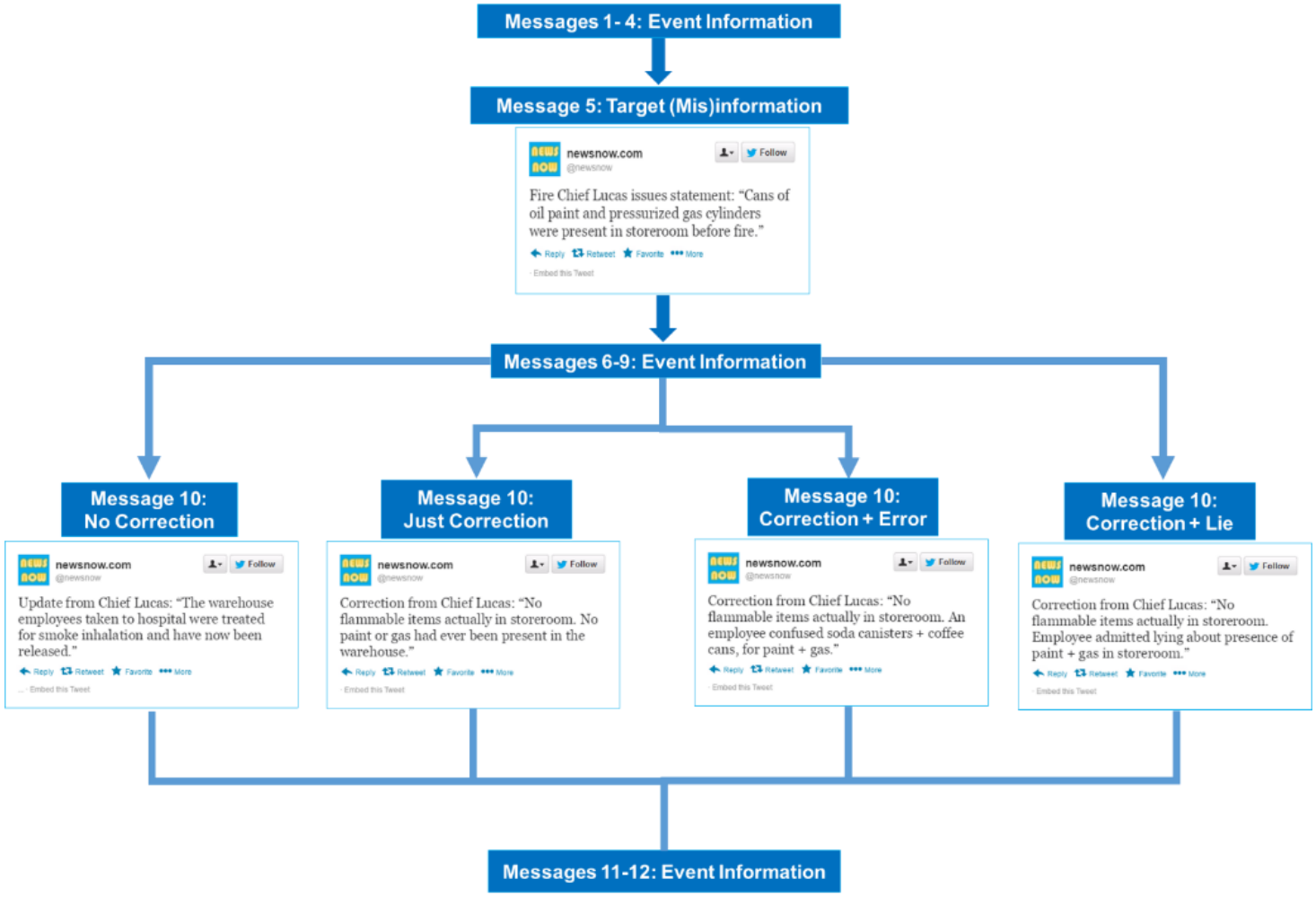

Figure 6 Schematic diagram depicting experimental design in Experiment 3

Figure 7 Example Misinformation 'Tweet' used in Experiment 3

In terms of the experimental manipulations, target (mis)information (Message 5) stated that carelessly stored oil paint and pressurized gas cylinders were present in a storeroom before the fire. This information was later corrected (Message 10), for the three conditions featuring a correction but remained uncorrected for the no correction group, who provided a baseline for the inference test. The no correction group instead saw a control statement indicating that warehouse workers taken to hospital had been released. The correction + error explanation group learned that the target (mis)information had been corrected because an employee had made an error and confused soda canisters and coffee cans in the storeroom for paint and gas (similar to Bush et al. 1994's 'explain quality' condition). The correction + lie explanatory correction group learned that misinformation was incorrect because an employee had lied about the presence oil paint and gas cylinders in the storeroom (later studies provided a motivation for the lie that the employee was unhappy). The remaining (10) messages provided further details of the incident and were identical in all four experimental conditions.

3.5.4. Procedure

Participants clicked on a link in Amazon Mechanical Turk to enter the experimental site. They subsequently: read details about the experiment and gave consent to take part, they then received instructions that the study

explored the factors that affect people's judgments about news reports and that their task was to read a brief report about an investigation into a fire and complete a short questionnaire about the report, and then provide demographic information. Participants were told they would not be able to backtrack and that each message would appear for a minimum of 5 seconds before they could move on to the next message. They then completed the first instructional attentional check (e.g. Oppenheimer, Meyvis, & Davidenko, 2009), before starting the experiment. Participants (N = 4) who did not respond appropriately, indicating that they had not read the instructions properly, were not able to complete the study and received a message thanking them for their time.

After completing the first instructional attention check, participants read one of four versions of the warehouse fire report (depending on the condition they had been allocated to). The 12 messages making up the report were presented individually and appeared on the screen for a minimum of 5 seconds before participants could move on to the next message. Immediately after reading the report, participants were taken to the questionnaire instruction page that informed participants that they would now see 15 questions about the report. Questionnaire instructions made it clear to participants that they did not have to rely on the material presented in the messages to answer inference questions (see Appendix G). The questionnaire consisted of 16 questions: 7 inference questions, 7 factual recall questions, and 2 questions probing recall of critical information presented at Message 10 (see Appendix F). Inference and factual recall question blocks were intermixed and presented in a random order except the question probing the most likely cause of the fire, which always came last.

Inference questions probed participants' understanding of the news report (e.g. "Is there any evidence of careless management in relation to this fire?"), and included a question querying the most likely cause of the fire. Factual recall questions enquired about the literal details of the report (e.g. "Which hospital were the workers taken to?"). Two further questions assessed recall ("What was the point of the second message from Fire Chief Lucas") and ("Were you aware of any corrections or contradictions in the messages

you read'), of the critical information (i.e. Message 10). Participants in all four conditions were able to answer the first question from the report that they read. The latter question, however, was only relevant to participants in conditions featuring a correction. Participants typed a response to each of 16 questions in a text box, were required to use a minimum of 25 characters, and encouraged to answer using full sentences.

One of the questions included in this block was the second instructional attention check which asked participants to indicate approximately how many 'Twitter' messages they thought they had read. In parentheses the question then asked participants to respond with 'more than 40'. If participants chose anything other than the response indicated in parentheses this was taken as evidence of inattention to instructions. Unlike the first instructional attention check participants were not immediately excluded from the study. This question was included as a means of excluding participants before analysis[21]. After answering this block of questions, participants were informed they would answer 2 more questions on the basis of what they remembered from the report. After completing the questionnaire participants were asked to provide their sex, age, and highest level of education.

3.6. Results

3.6.1. Coding of Responses

Responses to three types of questions were used in the analysis. Participants answered the 7 inference questions on the basis of their understanding of the report. Responses to inference questions were coded as reference to target (mis)information (i.e. references to information that in most conditions was corrected were given a score of 1) if they explicitly stated, or strongly implied, that oil paint and gas cylinders caused or contributed to the fire and were scored 0 otherwise. This gave a minimum inference score of 0 and maximum of 7. The factual recall questions could be answered by recalling the literal details of the report. Each response was coded as 1 if the

[21] Although ultimately it was decided not to exclude these participants since excluding them did not change the results.

detail was fully or partially recalled and scored 0 if it was not accurately recalled. This gave a minimum recall accuracy score of 0 and maximum score of 7. Critical information recall scores were computed using the same criteria; the maximum individual critical information score was 2[22]. An additional measure was computed from the factual recall which asked about the contents of the storeroom before the fire. Responses to this question were coded to examine the presence of misinformation in participants' literal recall of the event information. Responses indicating that oil paint and gas cylinders were in the storeroom before the fire were scored 1 and were otherwise scored 0.

Inter-coder reliability. Responses were coded by a trained coder. A second, independent judge then coded 10% of participants' responses (*n* = 36). Inter-rater agreement was 0.88 and Cohen's K = 0.76±0.03, indicating a high level of agreement between coders, both of which are higher than the benchmark values of 0.7 and 0.6 (Krippendorff, 2012; Landis, & Koch, 1977), and there was no systematic bias between raters, χ^2 = 2.45, *p* = .12.

Table 3 Example response coding for inference questions in Experiment 3

Inference Question	Response Scored 1	Response Scored 0
How could the fire at the warehouse have been avoided?	The fire at the warehouse could have been avoided by keeping accelerants and explosives such as pressurized gas and flammable paints in a designated contained area, per fire safety code	Whether or not the fire could have been avoided would depend on whether the facility was compliant with safety regulations during their inspection.

[22] Previous CIE studies refer to this measure as 'awareness of the correction/retraction'. The average of the two 'awareness' questions is typically compared for conditions featuring a correction to misinformation. This approach ignores that the 'no correction' condition. Ultimately, only the first question was analysed ("What was the point of the second message from Fire Chief Lucas") in order to compare recall of the message presented at the same serial position in each version of the report, irrespective of message content.

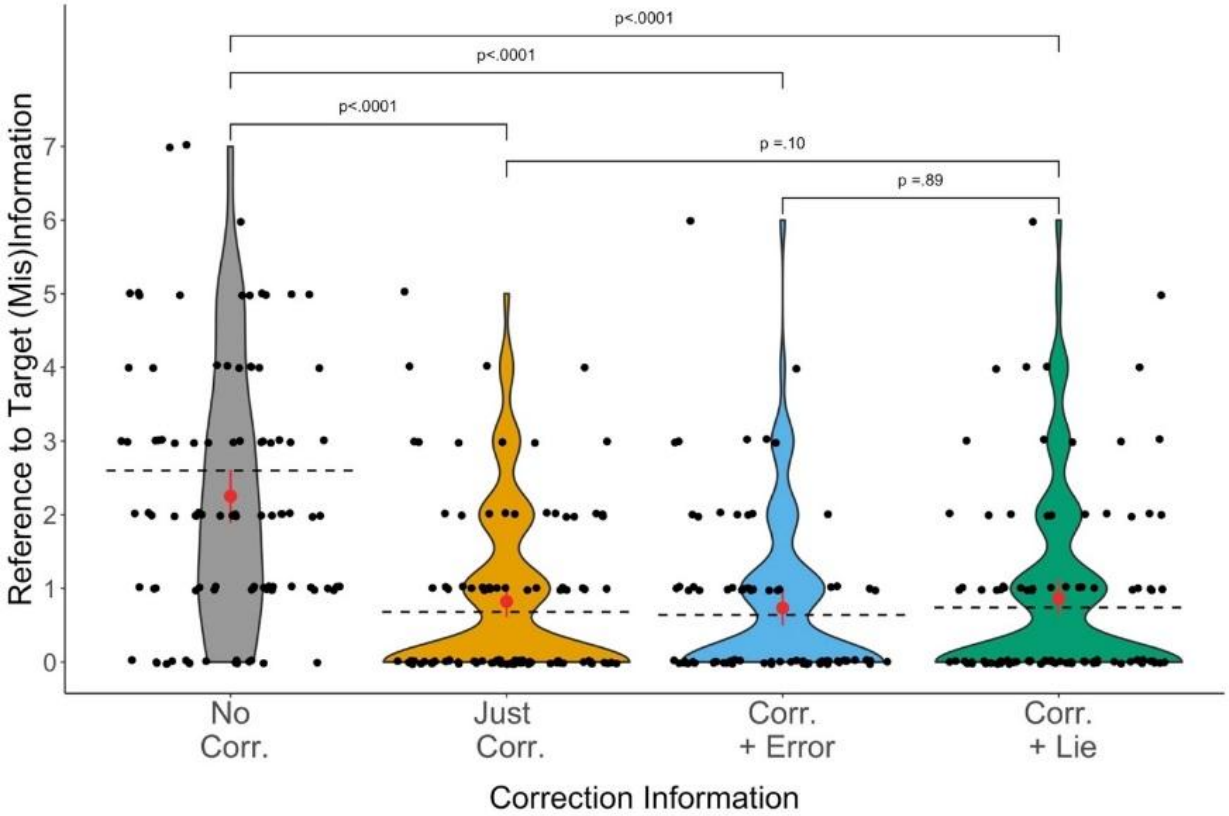

Figure 8 Distribution and probability density of references to target (mis)information by correction information condition in Experiment 3. Red points represent mean and 95% confidence interval of the mean. Dashed lines represent condition means after excluding participants who did not recall the critical information presented at Message 10. Lines are based on data from 25 (no correction), 88 (just correction), 71 (correction + error), and 83 (correction + lie) participants, respectively. Brackets show significance of Dunnett least-square mean comparisons which control for unequal group sizes.

3.6.2. Inference Scores

Figure 8 shows the distributional characteristics of inference scores across correction information conditions are shown. The no correction group served as an empirical baseline for interpretation of the three groups that received a correction to the target (mis)information. The majority of participants in correction groups did not refer to misinformation in response to any of the inference questions. All three types of correction substantially reduced reference to misinformation. All three types of corrections clearly reduced the number of references to target (mis)information relative to the condition where the target information was uncorrected.

Previous CIE studies have examined differences in the number of references to target (mis)information between conditions using ANOVA. I decided that a different analytical approach was more appropriate for the following reasons. Inference scores – or the number of references to misinformation – are non-negative integer values and therefore constitute count data. General linear models (such as OLS regression and factorial ANOVA) are often used to analyse count data but may produce biased estimates and inferences - particularly when the data are characterised by excessive zeroes as is the case with the present data (Atkins & Gallop, 2007; Baguley, 2012). Poisson regression models are more appropriate for count data because they rely on a Poisson distribution rather than a normal distribution as their probability model (Atkins & Gallop, 2007). Zero-inflated Poisson (ZIP) regression is an extension that can be used to account for excessive zeros (i.e. cases where the participant did not make any misinformation references) because it directly models the zeros in the structural part of the model. The ZIP model therefore has two parts, a Poisson count model, and a logit model for predicting extra zeros. It assumes that there are two processes, that a participant has referred to misinformation or they have not. If they have referred to misinformation then it is a count process. In this case the count process was modelled with a negative

binomial model[23] (Long, 1997).

A model comparison approach was used whereby a null model (including only the intercept) was compared to a model that included correction information as a predictor. The model including correction information fit the data significantly better than a null model (i.e. intercept-only model), $\chi^2(3) = 17.14$, $p = .0007$[24]. Dunnett's multiple comparisons tests of the estimated marginal means further confirmed that all three types of correction (explanatory and non-explanatory) significantly reduced the number of references to target (mis)information (see Table 4 below). The differences between the three correction information conditions were not significant. These results show that correction information approximately halved the number of references to target (mis)information. The majority of the participants in the conditions featuring a correction did not refer to misinformation in response to any inference questions. Participants who did refer to misinformation either generally referred to it once or twice suggesting that participants were more likely to refer to misinformation in response to some questions than others (see question analysis below).

Table 4 Estimated marginal means by correction information condition in Experiment 3

Correction Information	Estimated marginal mean	Std. error	Group
Corr. + Error	0.73	0.12	a
Just Corr.	0.82	0.12	a
Corr. + Lie	0.86	0.13	a
No Corr.	2.25	0.19	b

[23] Negative binomial models are appropriate when the data are overdispersed (i.e. the variance is greater than the mean).

[24] The chi-square statistic represents the deviance goodness of fit test for Poisson regression. Deviance is a measure of how well the model fits the data or how close the model predictions are to the observed values. To obtain the chi-square statistic a likelihood ratio test comparing a null model to a saturated model is performed.

Note: Confidence level used: 0.95. Confidence-level adjustment: Dunnett method for 4 estimates. P value adjustment: Dunnett method for 6 tests. Significance level used: alpha = 0.05. Group represents significance of comparisons. Groups that share the same letter are not significantly different from each other and groups with different letters have significantly different means.

Question analysis. The distribution (see Figure 8) of inference scores suggested that participants might be mostly referring to misinformation in response to a subset of inference questions. In order to further examine how participants were responding to inference questions, and whether some questions were more likely to elicit references to misinformation than others, the proportion of references to misinformation was computed as a function of condition and inference question. Figure 9 shows that participants mainly referred to misinformation in response to the question about the most likely cause of the fire, and to a lesser extent, the question asking how the fire could have been prevented. Proportionally, references to misinformation on other questions were relatively low for all three conditions featuring a correction. This suggests that some 'types' of questions are more likely to elicit references to misinformation than others which could in turn influence the strength and presence on the CIE. Furthermore, this analysis indicates that some questions are more diagnostic of a continued influence effect than others and the questions selected can modulate the strength of the observed effect.

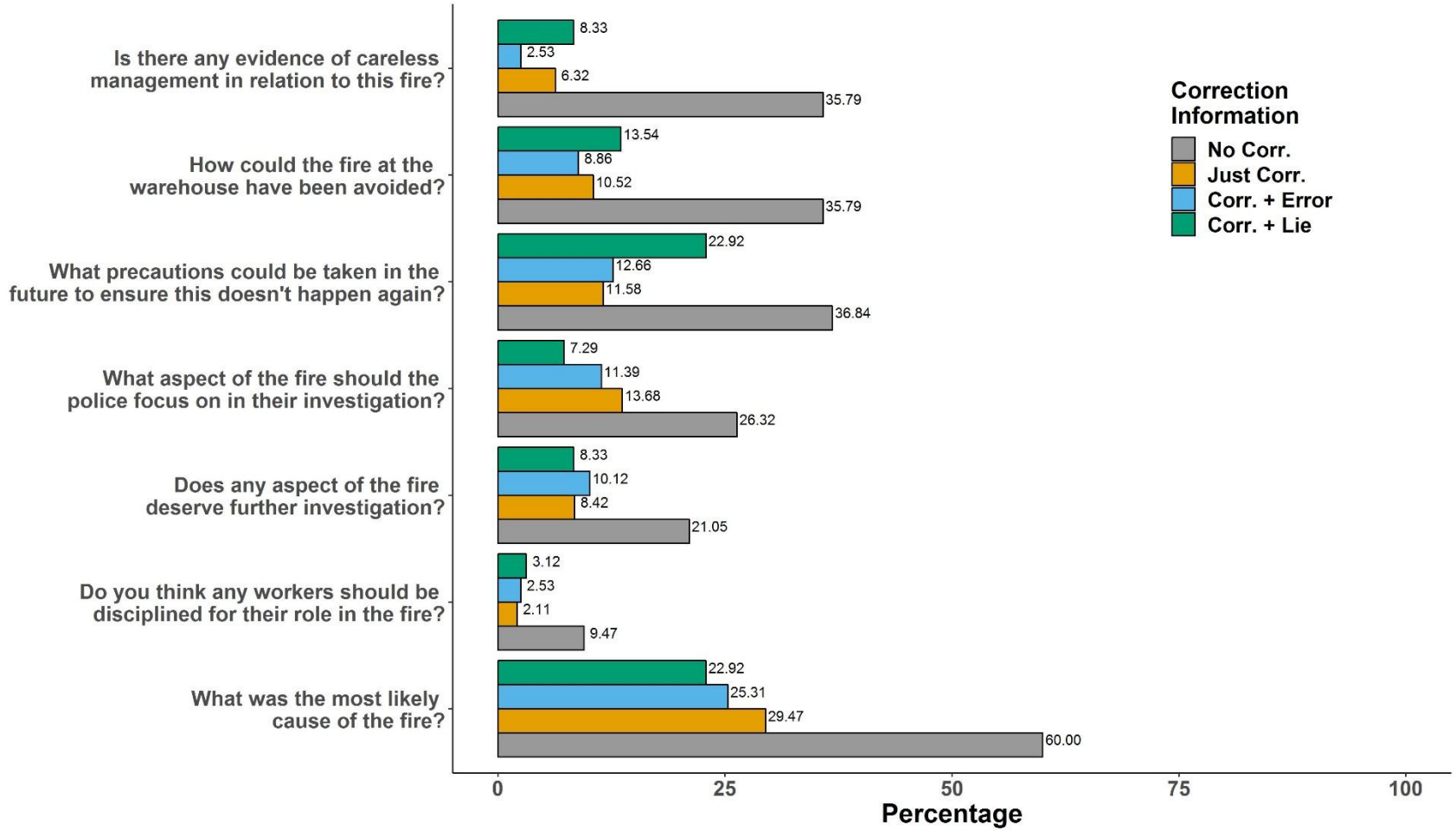

Figure 9 Proportion of references to target (mis)information by question and correction information condition in Experiment 3

Correction acknowledgment. One of key claims from the CIE literature is that people often continue to rely on misinformation despite clearly understanding and recalling that the misinformation was corrected (e.g. Johnson & Seifert, 1994). In order to examine to extent to which this claim was true of the current study, the proportion of participants who correctly recalled the correction yet still referred to misinformation in response to at least one inference question was computed (see Figure 10). Responses were categorized as referring to misinformation if there was at least one reference to misinformation in response to one of the seven inference questions. A small majority of participants (51-53%) accurately recalled the correction and did not make any uncontroverted references to misinformation in response to any of the inference questions. A substantial minority of participants across all three correction information conditions (34-40%) recalled the correction and made at least one reference to misinformation on inference questions, and therefore exhibited a continuing influence of misinformation. The remaining participants (7-14%) did not recall the correction. Thus overall, we observed that a substantial proportion (around a third) of participants who received a correction referred to misinformation despite acknowledging that the information was corrected.

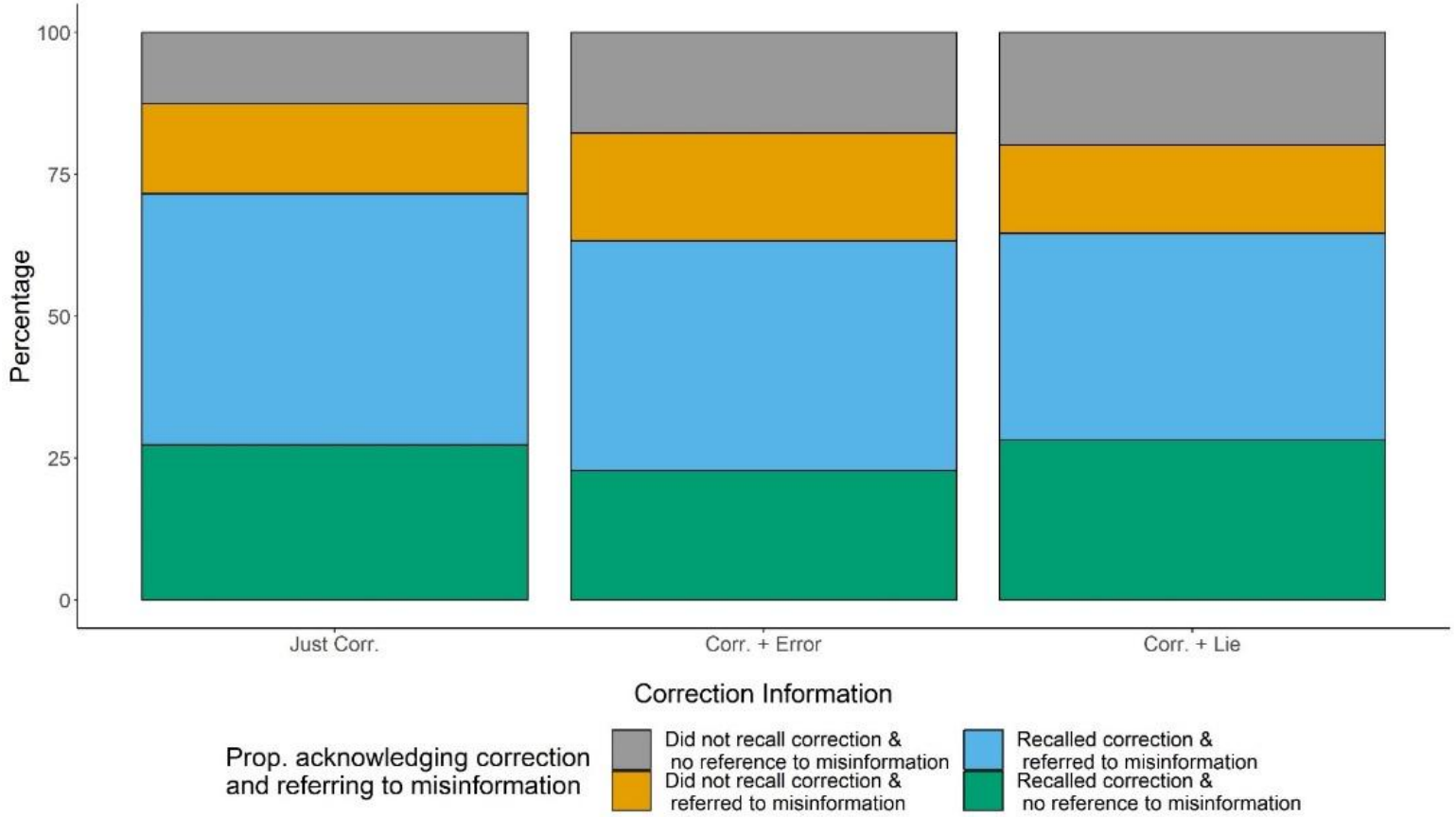

Figure 10 Proportion of participants from correction groups who recalled the correction and referred to target (mis)information as a function of correction information condition in Experiment 3

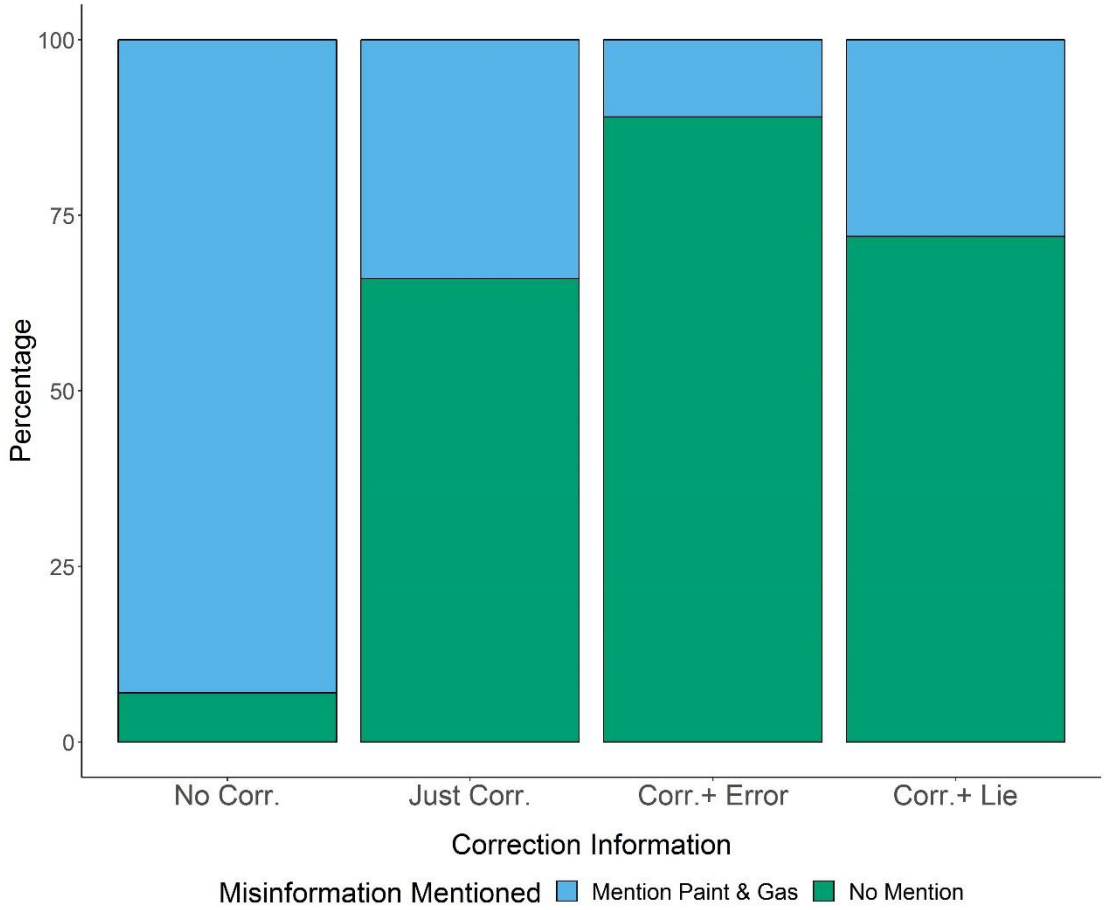

Figure 11 Proportion of references to presence of oil paint and gas cylinders in storeroom before the fire in response to the recall question probing this knowledge in Experiment 3

3.6.3. Recall Accuracy Scores

The manipulation of correction information was not expected to have any effect on participants' ability to accurately recall report details. Recall accuracy scores are also non-negative integers or count data but were not characterized by excessive zeros so a Poisson regression model was fit to the data. The model including correction information was not a significantly better fit for the data than the null model, χ^2 (3) = 2.36, p = .50. This means that the conditions' factual recall did not differ and that the differences between conditions observed for inference scores can therefore not be attributed to overall differences in factual recall of story information.

Misinformation mentions in recall. Responses to the recall question probing factual recall of the contents of the storeroom before the fire were compared across correction information conditions. The no correction

condition was not included in this analysis in order to just compare conditions featuring a correction. There was a significant association between correction information condition and mention of flammable substances, χ^2 (2) = 15.16, p < .001. Figure 11 shows the proportion of responses mentioning flammable substances in the storeroom by correction information condition. The correction + error group were less likely to incorrectly state that gas and oil paint had been in the storeroom before the fire than the two other corrected groups. This suggests that correction type can influence factual recall while having no differential effect on inferential use of misinformation. This may have implications for the mechanisms by which the CIE works.

3.6.4. Critical Information Recall

Two questions assessed participants' awareness of, and ability to recall, the correction information (presented at Message 10). This measure is usually assessed to examine whether there are differences in awareness and recall of the correction across conditions featuring a correction. However, this approach does not consider whether participants in the control (no correction) group could accurately recall the control information shown to them at Message 10. In order to make this comparison, only the first question ("What was the point of the second message from Fire Chief Lucas"?) was analysed. Instead of comparing whether conditions featuring a correction differ with respect to awareness and recall of the correction information, the test performed examined whether participants in all conditions recalled critical information from the same serial position in the story.

A chi-square test of independence on these data revealed a significant association between correction information condition and critical information recall, χ^2 (3) = 47.94, p < .001. This was primarily due to the low proportion of the no correction group (26.3%) who accurately recalled the critical information (i.e. that injured firefighters were released from hospital). In contrast over half of the participants in the groups featuring a correction recalled the critical correction information: correction + error (63.2%), correction + lie (64.5%) and correction only (71.6%). The difference in conditions may be explained by the difference in the salience of the critical

information (i.e. correction of misinformation vs. update about the firefighter's injuries). The test was not significant when looking just at the conditions featuring a correction, χ^2 (2) = 1.62, p = .44.

3.7. Summary

Experiment 3 set out to test whether corrections that explain why misinformation is incorrect and how it occurred are more effective at reducing reliance on misinformation than corrections which negate misinformation (non-explanatory correction). Participants read a fictional report about a warehouse fire in which target (mis)information implied that careless storage of oil paint and gas cylinders were a likely cause of the fire. Later in the report participants in conditions featuring a correction learned that no flammable items were present before the fire. The explanatory groups also learned that misinformation occurred because of an error (an employee mistook non-flammable liquids for flammable liquids), or because of a lie (an employee lied about the presence of flammable liquids in the storeroom).

It was predicted that corrections which explained why misinformation was incorrect would more effectively reduce reliance on misinformation than non-explanatory corrections. There was also a tentative prediction that a lie explanation would be more effective than the error explanation. Contrary to predictions, there was no evidence that explanatory corrections were more effective than a non-explanatory correction at reducing the CIE. Explanatory and non-explanatory corrections to misinformation resulted in a comparable number of references to misinformation. There was also no difference between error and lie explanations in the number of post-correction references to misinformation.

Explanatory and non-explanatory corrections reduced references to misinformation by 62% and 68% relative to the no correction condition. These results are consistent with previous CIE studies showing that, on average, a correction reduces reference to misinformation approximately by half (see Table 1 in Chapter 1). Findings suggest that a correction which explained that misinformation occurred because of an honest mistake or deception offered

no additional advantage in reducing references to misinformation compared to a non-explanatory correction.

Experiment 3's results are inconsistent with previous findings (Bush et al., 1994) showing that explanatory corrections are more effective than non-explanatory corrections. Prior work found that explanatory corrections reduced reliance on misinformation by 14-20% more than corrections which negated misinformation. Experiment 3's results demonstrate that a clear correction without an explanation can be just as effective in correcting misinformation as including a reason for the misinformation being presented initially. This occurred despite the fact that explanatory corrections should encourage and require further processing of the correction because there is more information to comprehend. It is not entirely clear why this was the case, but the fact that there was some information available in the correction condition ("no paint and gas had ever been present in the warehouse"), that was not available in the explanatory correction messages could have rendered all three corrections equally effective.

Experiment 3's results did show evidence of a continued influence effect of misinformation. The majority of participants in conditions featuring a correction made at least reference to misinformation despite an unambiguous correction. Furthermore, a sizeable proportion of the correction groups (34-40%) made at least one reference to misinformation whilst also acknowledging that the misinformation had been corrected. This is a novel finding and is usually not reported despite being a central claim from the CIE literature (e.g. Johnson & Seifert, 1994). Analysis of the references to misinformation across individual questions also revealed that participants were most likely to refer to the misinformation when asked what the most likely cause of the fire was. This strongly suggests that corrected misinformation plays a role in participants' causal understanding of the warehouse fire and is consistent with the mental-model updating account of the CIE (Johnson & Seifert, 1994; Seifert, 2002).

One limitation of the present work, and of CIE research in general, is that findings are demonstrated for a single story. Those that have compared

across scenarios do sometimes find that the strength of the correction differs somewhat between scenarios (discussed in more detail in Chapter 1). Experiment 3's results could therefore be scenario specific. The effect (or lack thereof) of explanatory and non-explanatory corrections could interact with the specific details of the scenario presented to participants. The particular scenario used could also moderate the strength of a correction which could in turn impact the validity and generalisability of the findings. Experiment 4's aim was therefore to examine whether the effectiveness of explanatory corrections extended to different scenarios with the same underlying structure.

3.8. Experiment 4

Experiment 4 further explored the effect of explanatory corrections in multiple scenarios. In order to address this, Experiment 4 tested the effect of explanatory corrections across four different scenarios (reports). If the effect is robust it should extend to scenarios that have the same underlying structure but different content. In Experiment 4 the explanatory and non-explanatory corrections were also more closely matched so that they both had the same base correction and that the only difference was whether an explanation was provided. The explanatory corrections used in Experiment 3 did not include the information that "no paint and gas had ever been present in the warehouse". This was changed in Experiment 4 to ensure that the lack of difference between conditions was not due to this discrepancy between the base correction information presented in each correction information condition. The scenarios used in Experiment 4 also included a statement providing information about potential causes of the outcome described in the report to allow participants to answer inferential questions, even though the misinformation explanation had been invalidated (see Figure 13). Unlike the other experiments reported in this thesis, Experiment 4 was conducted in the lab. The reason for this was that Experiment 4 employed a Latin-square design in which participants took part in all four correction information conditions and saw four different scenarios. This meant that the study took substantially longer to complete than the previous experiments reported in this

thesis. In order to minimise attrition and maximise attentiveness throughout, the study was completed in the lab.

3.9. Method

3.9.1. Participants

A power analysis using a medium effect size (f = .25, α = .05, 1-β = .95), for the effect of correction information (df = 3, k = 1, number of measurements = 4[25]), indicated it would be necessary to collect data from a minimum of 36 participants[26]. In total data were collected from 37 participants (21 female, aged 19 to 57, M = 26.92, SD = 9.83) recruited from the City, University of London, subject pool (https://city.sona-systems.com/). Twenty-one participants were paid £8 in return for participation; the remaining participants received course credits. Participants took 49 minutes on average to complete the study.

3.9.2. Design

Experiment 3 used a 4 x 4 Latin square experimental design (Bradley, 1958). A Latin square is a specific randomized block design which has a three-way layout (Kirk, 2013). This design included two blocking variables; one assigned to the rows of the square (group: group 1, group 2, group 3, group 4) and one to the columns (event narrative: fire, crash, injury, missing person), and is represented in Table 5. This meant that the 4 correction information conditions could be tested across 4 different scenarios, by randomly assigning participants to 1 of 4 possible groups or rows of the square. Although a Latin square is similar to a three-way ANOVA, it is more parsimonious because it allows the effects of two blocking variables (e.g. group and scenario) to be isolated from the effect of correction information

[25] Number of measures refers to the number of inference scores that were computed for each participant, based on the number of scenarios they read, and is required to compute power for a repeated measures ANOVA.

[26] This sample size was computed to look for the main effect of correction information with four measurements reflecting the four different scenarios participants read. The time of designing the study I was not aware that there was a method of calculating a sample size for Latin-square designs and GPower does not include an option to compute sample sizes based on Latin square designs. Therefore, the sample size estimate is possibly incorrect.

(i.e. the treatment variable), to achieve greater power to reject the null hypothesis (Kirk, 2012). It is recommended that only the main effects from the Latin square are estimated. This is due to the fact that the square is in an incomplete factorial design and not all of the cells are represented. There were 16 cells in total in Experiment 3 (4 x 4 x 4). This design is efficient when it is unfeasible to test all possible cells in a three-way factorial design. Requiring participants to respond to all 16 versions would be time-consuming and impractical, as well as encouraging immediate sequential carryover effects (Bradley, 1958). The second blocking variable group was added such that each participant only responded once to the 4 different scenarios.

Table 5 Scenario and Correction Information Implemented in Latin Square

Group	**Fire**	**Crash**	**Injury**	**Missing Person**
1	No Correction	Just Correction	Correction + Error	Correction + Lie
2	Just Correction	Correction + Lie	No Correction	Correction + Error
3	Correction + Error	No Correction	Correction + Lie	Just Correction
4	Correction + Lie	Correction + Error	Just Correction	No Correction

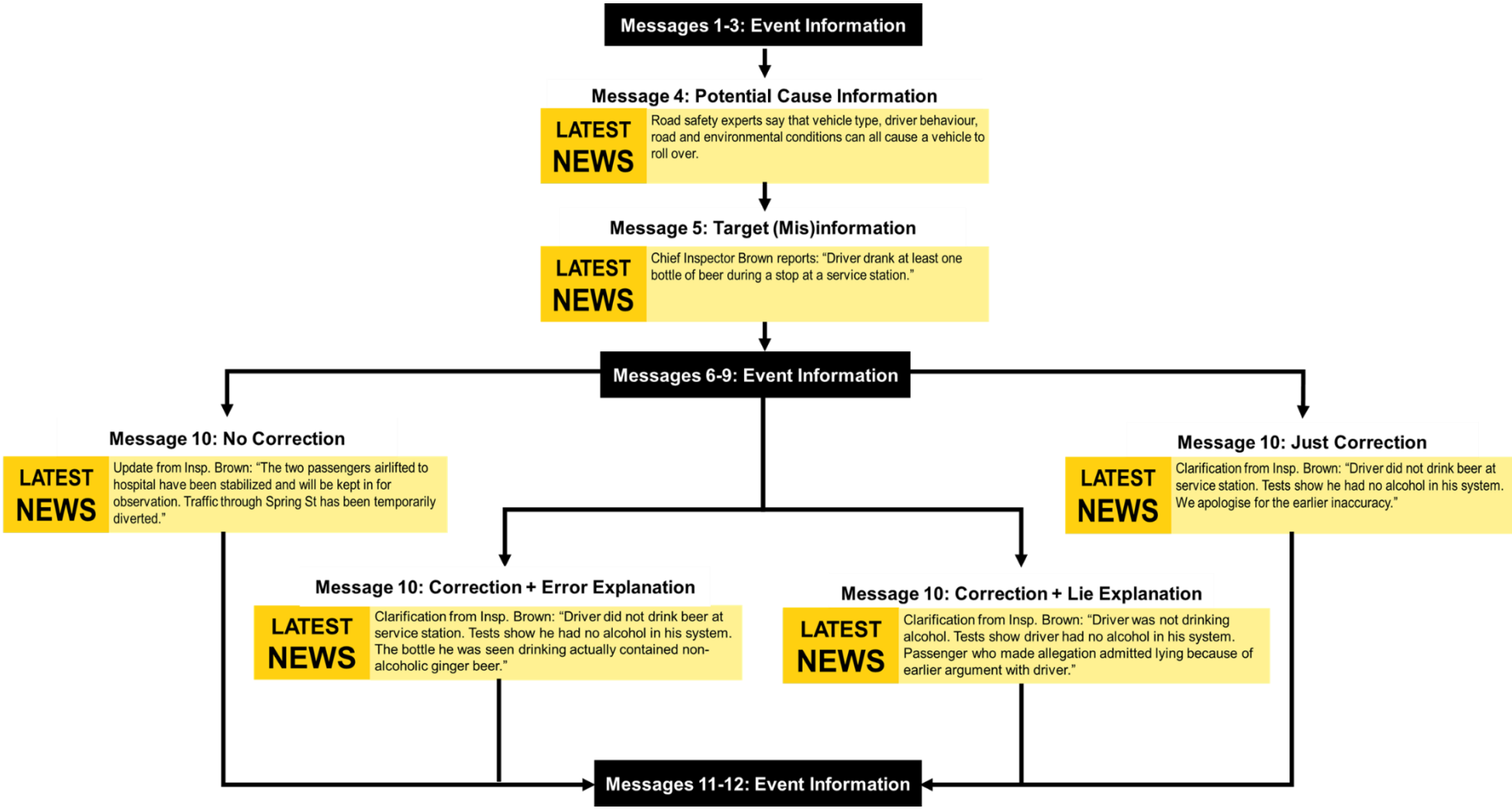

Figure 12 Schematic design showing the high-level design and content of the 'crash' scenario used in Experiment 4

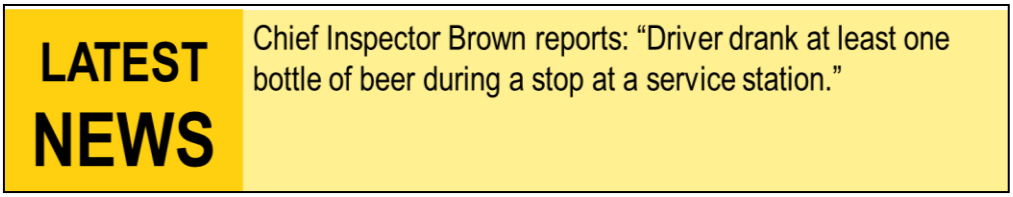

Figure 13 Example statement from 'Crash' report in Experiment 4

3.9.3. Stimuli

The experimental stimuli consisted of four different scenarios containing 12 individually presented messages (see Appendix H for full details of the scenarios). The messages were presented in the form of individual breaking news statements that appeared to originate from the same fictional news source (see Figure 13). The maximum character length per message was changed to 250 and the maximum number of words was 35. The messages were approximately matched for number of characters and words across experimental conditions. The change in presentation format from Experiment 3 for individual messages was made in order to allow more freedom to increase length of the messages – real 'Tweets' were limited to 140 characters at the time the experiment was run. All four scenarios were constructed so that they had the same underlying structure but appeared as superficially distinct stories. Three of the scenarios (i.e. fire, crash, and missing person) used in Experiment 4 were based on scenarios used in previous CIE studies but were distinct from the original stories (Ecker et al., 2011b; Ecker et al., 2010; Johnson & Seifert, 1994).

Several modifications were made to the fire report used in Experiment 3. First, Message 4 provided general information about potential causes for the outcome (e.g. industrial fires are often caused by electrical issues). This was new to Experiment 4 and was included in the reports so that participants always had some information to answer inferential questions even if the posited cause had been corrected (as it was for three of the conditions). This message was changed to allow participants the opportunity to generate alternative explanations for possible causes of the fire after the misinformation

was invalidated. Second, the correction message in Experiment 3 offered additional information that was not stated in the explanatory correction messages (i.e. that no paint or gas had ever been present in the warehouse) whereas the explanatory correction messages did not provide this information. The explanatory correction messages included the same information as the non-explanatory correction, as well as the respective explanation, in order to rule out the possibility that the lack of difference between correction information conditions in Experiment 3 was due to differences in the information included in the correction messages. Third, the 'Correction + Lie' message was also amended so that the correction message referred to an 'unhappy employee' in order to provide some context for lying about the presence of flammable substances in the storeroom. Finally, the additional information messages included in the report were amended to better match the length of the critical (correction) messages.

The fourth scenario was a novel report, concerning a person found with a head injury, which has not been used in previous studies on the continued influence effect. Figure 12 depicts represents the different correction information conditions for the 'crash' scenario used in Experiment 3 and shows how stimuli were presented to participants. Misinformation always appeared at Message 5 and provided information about a likely cause of the outcome (i.e. how the woman sustained head injuries). The correction (or control) information always appeared at Message 10.

3.9.4. Procedure

Participants were tested in individual cubicles and completed the experiment on a computer. The experiment was implemented via Qualtrics. They were informed by the experimenter that they were taking part in a study investigating how well people understand and remember information presented in news reports, that they would read four different reports and answer questions about the reports and were informed about the minimum requirements for responses (i.e. minimum of 25 characters). The written instructions appeared on the screen whilst participants were verbally instructed about the task. Once it was clear that participants understood the

task they were allowed to begin the experiment. They first had to click the >> button which took them to the first report, after which they responded to 14 intermixed, and randomly presented, inference and fact recall questions, and then 2 critical information recall questions which were always presented in the same order. This basic procedure was then repeated three further times until participants had read and answered questions about all four reports. Participants finally answered a series of debrief questions and were then debriefed about the purpose of the experiment, in person.

3.10. Results

3.10.1. Coding of Responses

The main dependent measure extracted from the responses was 'reference to target (mis)information'. Responses which explicitly stated, or strongly implied, that the target information was causally involved in the event were scored 1 and otherwise scored 0 (e.g. that the van crashed because the driver was drunk, that the woman was assaulted by the man seen running away). Consistent with Experiment 3, the maximum individual inference score for each scenario that each participant saw was 7. The same coding criteria used in Experiment 3 were applied to code fact recall and critical information recall responses.

Inter-coder reliability. A trained coder coded all inference, fact recall, and correction recall question responses; 10% (n = 4, items = 64) of the responses were then coded by a second trained coder. Cohen's K was run to determine the level of agreement between the two raters. There was a high level of agreement between the two raters, $K = 0.83$, $p < .001$, 91.7%.

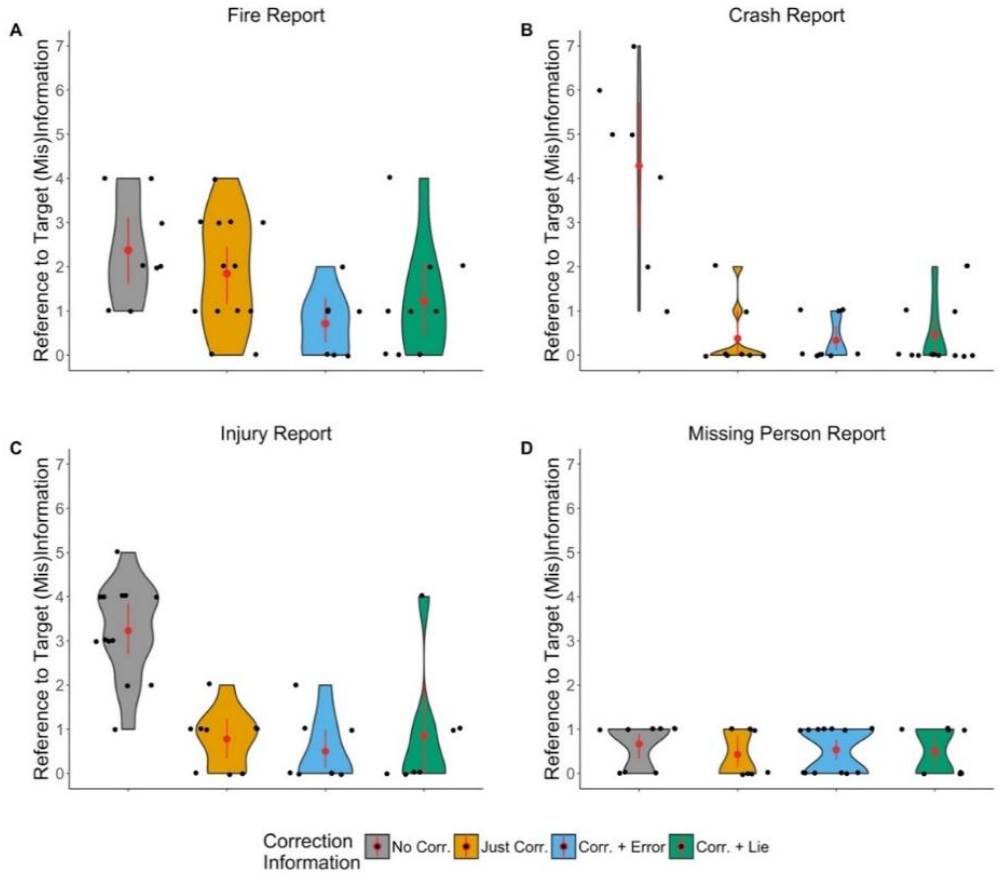

Figure 14 Distribution and probability density of references to target (mis)information by correction information and scenario in Experiment 4. Red points represent mean and error bars represent 95% confidence interval of the mean.

3.10.2. Inference Scores

Figure 14 shows the distributional characteristics for inference scores as a function of correction information and scenario. One assumption of Latin-square designs is that row, column, and treatment variables do not interact. Figure 9 clearly shows that the number of references to target (mis)information depended on whether or correction information was presented and which report or scenario the correction information appeared in. The figure also shows that the means for each scenario were based on a very low number of observations and that there were an unequal number of observations in each cell of the Latin-square. Correction information was also

more effective in some scenarios than in others. The magnitude of the difference between corrected and uncorrected groups appears to be substantially larger for the crash report than for the fire report, whereas the injury report to be somewhere in between the two.

A Poisson mixed effects model was fit to inference scores in order to account for the repeated measures nature of the count data. Participant was treated as a random variable allowing intercepts to vary between participants. A model which included row (group), column (scenario), and treatment effects (correction information), was significantly better than a model which included only the intercept, χ^2 (9) = 108.21, $p < .001$[27].

The fact that group was included in the best fitting model supports visual inspection of the data indicating that there was an interaction between correction information and scenario. The nature of the Latin-square design meant that any test of an interaction between correction information and scenario would be biased. Follow-up tests of correction information, scenario, and group were not performed because the overall differences between levels of these factors were not particularly informative. Breaking down inference scores by scenario would mean that inferences would be based on a very low number of observations and therefore unreliable. It was deemed best to rely on visual representation of the data for inference and the results are therefore described qualitatively.

The warehouse fire and injury reports present a classic continued influence effect of misinformation, in which participants in correction conditions continue to refer to misinformation but not as much as in the uncorrected condition. The missing person report seems to show that participants did not attend to the misinformation, so did not refer to it even when it was not corrected. As such, this does not tell us much about the continued influence effect. The crash report, on the other hand, gives a good example of a relatively effective correction. Under a third of participants in the

[27] Predictors were added incrementally. Correction information was added first, scenario second, and group last. Models were then sequentially compared to examine whether that predictor improved the model fit. The chi-square values reported here are for the comparison of the full model including all of the predictors against the null model where the intercept was fixed.

correction conditions referred to the misinformation, whereas all participants in the uncorrected condition referred to the misinformation.

3.10.3. Recall Accuracy Scores

The manipulations of scenario and correction information were not expected to influence recall accuracy. A mixed effects Poisson regression model was fit to recall accuracy scores. This confirmed that the number of accurately recalled details was not predicted by correction information, scenario, or group (*p*'s > .3).

3.11. Summary

Experiment 4 was designed to replicate and extend Experiment 3's finding that including an explanation for why misinformation is incorrect is no more effective than clearly negating misinformation. Findings were replicated for the warehouse fire and head injury reports. Specifically, corrections failed to eliminate the influence of misinformation and there was no difference between explanatory and non-explanatory corrections when a CIE was present. There was little to no evidence of a CIE in the crash report. That is, participants produced a substantial number of references to uncorrected target (mis)information in response to the crash report, but very few participants referred to misinformation after it had been corrected.

Despite the fact that the reports were intended to have the same underlying structure, the uniform pattern of misinformation references in the missing person scenario suggests there were unintended differences between the reports. The crash report resulted in a substantial number of misinformation references in uncorrected conditions but almost none in corrected conditions. This contrasts with the warehouse fire and head injury reports where the difference between corrected and uncorrected conditions was much smaller. This suggests that the scenarios differed in some fundamental respect despite being designed to be structurally very similar. The difference between scenarios seemed to be related to the 'base' correction information rather than between the misinformation used in the different reports. In other words, correction information had a larger impact in

the crash reports than the warehouse fire or head injury report. This suggests that the difference between the reports is not related to the causal relevance of the misinformation per se, but rather, the scope for the correction information to render the misinformation irrelevant. More specifically, the crash scenario corrected the inference that the crash was caused by drink-driving by providing evidence that this could not have been the case (i.e. the test confirmed the driver did not have alcohol in his system). In contrast, the correction in the warehouse fire report merely stated that the flammable substances were never in the warehouse but did not completely rule out flammable liquids as a cause of the fire. Similarly, the head injury report refutes the idea that the man seeing running away assaulted the woman but does not provide evidence that this did not occur (e.g. CCTV footage).

Experiment 4's results showed a classic CIE in the warehouse fire and head injury scenarios. There was no evidence of a CIE in the crash scenario suggesting the CIE is not inevitable. Experiment 4's results were very similar to results previously obtained online providing further proof that cognitively demanding memory-based tasks can be conducted online. Experiment 4's results suggested that the CIE occurs when reasoning about some descriptions of events but not for others. These results provide preliminary evidence that the magnitude of the difference between corrected and uncorrected conditions was much larger in the crash report than the fire or injury reports. Since it is not possible to test for interactions with Latin-square designs, and because there were a low number of observations in each cell of the square, it is difficult to draw firm conclusions from Experiment 4's results.

3.12. Experiment 5

Experiment 5 was designed to directly compare a scenario in which the CIE was eliminated (crash scenario) and a scenario in which the effect was present (warehouse fire scenario), in order to examine whether there was an interaction between scenario and correction information. There were two main reasons it was important to check whether the differences in the CIE across scenarios that was observed in Experiment 4 are robust. First, the sample sizes when breaking down the scenarios were very low (between 7 and 13

per cell). Second, the Latin-square design used in Experiment 4 meant that it was not possible to statistically test for an interaction between scenario and correction information condition. The upshot of this is that it was not possible to establish with any certainty whether the CIE is observed for some scenarios but not others. Experiment 5 therefore aimed to address Experiment 4's limitations by: 1) increasing the sample size and, 2) using a balanced and completely between-subjects design in order to specifically test for the interaction between scenario and correction information.

Experiment 5 was run online and so included several additional measures in order to further ensure that participants attended to the task and answered questions properly. An additional test examined whether participants had encoded the critical information presented at Message 10 (i.e. the correction or control message). The test was included in addition to open-ended critical information recall prompts as a further measure of whether participants noted the correction (see Figure 15). The recognition test comprised of a two-alternative-forced-choice (2AFC) recognition question asking participants to choose which of two statements had appeared in the report they had just read in addition to the open-ended critical information recall questions. The recognition test used was based on the 'modified recognition test' developed by McCloskey and Zaragoza (1985)[28] . For the recognition test used in Experiment 5, choosing the critical 'lure' message would indicate that a given participant had not properly encoded the information and was choosing randomly between the two options. If the participant had not encoded the misinformation then they would also be unlikely to give reasonable responses to inference questions and were therefore excluded. The actual correction and critical 'lure' messages were designed to be similar enough so as to not make the task easy but distinct

[28] The 'modified test' was developed in order to challenge the idea that misleading information presented at test in the post-event misinformation paradigm modifies memory for the original event (Loftus, 1975), overwrites the original 'correct' information resulting in the 'misinformation effect'. The recognition test typically used in the post-event misinformation paradigm includes the original information and the misleading information whereas the modified test includes the original information and a novel piece of information. The idea was that performance (i.e. correctly choosing the original information) should be equivalent to a control group who did not receive misinformation if their memory for the original information is not impaired. McCloskey and Zaragoza did not find a misleading information effect using this procedure.

enough so as the task was not too difficult.

3.13. Method

3.13.1. Participants

A power analysis indicated that a minimum of 110 participants (*f* = 0.40, 1- β = 0.95, α = 0.05) would be required in order to detect a main effect of correction information (*df* = 3, *k* = 8)[29]. In total of 163 participants completed the experiment via Amazon Mechanical Turk. One participant was excluded prior to analysis because they failed a recognition test examining memory for the correction message (see Figure 15). Following these exclusions, a total of 158 (69 females, aged 21 to 76, *M* = 39.62, *SD* = 11.21), participants were included in the final analysis. Participants were paid $1.50 and took 18 minutes on average to complete the experiment. Participants were also incentivised with the potential to receive an additional $10 if they achieved a high level of accuracy on the questionnaire (N = 2)[30].

3.13.2. Stimuli & Design

The experimental stimuli were generated via Qualtrics (Qualtrics, Provo, UT) and consisted of the warehouse fire and crash reports that were used in Experiment 4 (see Fig. 12 for example of experimental manipulation of crash report). A 2 (Scenario: Fire, Crash) x 4 (Correction Information: No Correction, Correction Only, Correction + Error, Correction + Lie) between-subjects factorial design was used such that there were 4 versions of the warehouse fire

[29]The effect size for the effect of correction information in Experiment 3 was used to estimate the sample size for Experiment 5. The effect size used (η_p^2 = 0.14) was incorrect and should have actually been η_p^2 = 0.18. The power calculation was therefore based on the wrong effect size and according to GPower the sample size should have actually been N = 83. However, this number seemed incredibly small and would have resulted in a very low number of participants in each condition after exclusions.

[30] The specifics of what constitutes a high level of accuracy were not described to participants. The incentive was intended to generally motivate participants to fully read the messages included in the report and respond accurately to all questions.

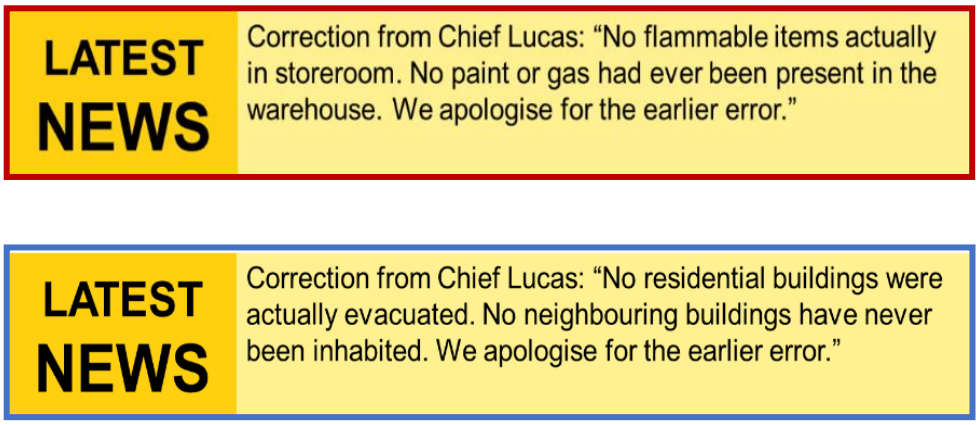

Figure 15 Example of the response options for the correction recognition test used for the fire scenario in Experiment 5. The response options represent those given to participants in the 'just correction' condition. The top panel (outlined in red) shows the actual correction message participants read in the report and the bottom panel is the 'lure' message.

report and 4 versions of the crash report. Each report consisted of 12 individually presented statements, in which target (mis)information was presented at Message 5 and critical (correction) information was presented at Message 10. Participants were randomly allocated to one of the 8 experimental conditions: no correction: warehouse fire (n = 14), crash (n = 20), correction: fire (n = 17), crash (n = 27), correction + error explanation: warehouse fire (n = 21), crash (n = 24), correction + lie explanation: fire (n = 18), crash (n = 17). The dependent variables of interest were the same as the other experiments reported in this chapter.

3.13.3. Procedure

Participants clicked on a link in Amazon Mechanical Turk to enter the experimental site. The Amazon Mechanical Turk advertisement also informed participants that there they had the opportunity to receive an additional bonus of $10 for providing accurate responses. The bonus was included in order to incentivize participants to carefully read the statements making up the report and to reduce the possibility that any continued influence effect observed was not due to misreading or misunderstanding of the messages included in the

report[31]. Experiment 5 included a set of instructional attention check questions which participants answered immediately after reading the instructions. These questions were ultimately not used to exclude participants but gave a sense of how well instructions are attended to. Only 98 (62%) of participants that were included in the final analysis answered all three questions correctly. Ultimately, it was not considered appropriate to exclude these participants because many researchers advise against excluding participants who fail instruction checks as this can introduce demographic bias in the data (Berinsky, Margolis, & Sances, 2014, 2016; Hauser & Schwarz, 2016).

Following this, the experimental procedure was much the same as Experiment 3: participants read the messages one at a time, answered intermixed inference and recall questions, and then answered the two critical information recall questions. Experiment 5 also included a test of whether participants had encoded the critical information presented at Message 10. Participants completed a modified recognition test (described in above). After completing the recognition test, participants answered debrief questions and were then debriefed about nature of the study.

3.14. Results

3.14.1. Coding of Responses

The coding criteria used for all measures (reference to target (mis)information, factual recall accuracy, critical information recall) were identical to those used to code the warehouse fire and crash reports in Experiment 4. Table 6 provides examples of coding for one of the questions following the crash report and one question following the fire report.

Inter-coder reliability. Responses were coded by a trained coder. A second, independent judge then coded approximately 10% of participants' responses (n = 17). Inter-rater agreement was 0.90 and Cohen's K = 0.81±0.05, indicating a very high level of agreement between coders, and

[31] After reading the instructions, and before starting the experiment, participants answered a three-part instructional attention check which can be seen in Fig. 10. A four-alternative forced choice (4AFC) method was used and response options included the correct answer as stated in the instructions and three incorrect answers.

there was also no systematic bias between raters, χ^2 = 1.33, *p* = .24.

Table 6 Example coded responses to inference questions in Experiment 5

Scenario	Question	Example of Response Coded as 1	Example of Response Coded as 0
Crash	How could this accident have been avoided?	Possibly by the driver not drinking and dulling his reflexes.	It is not known at this time, but road conditions might have played a part in the crash.
Warehouse Fire	What precautions could be taken in the future to ensure this doesn't happen again?	Flammable items should be placed in a safe storage place.	They need to have better safety system and a more effective plan of attack.

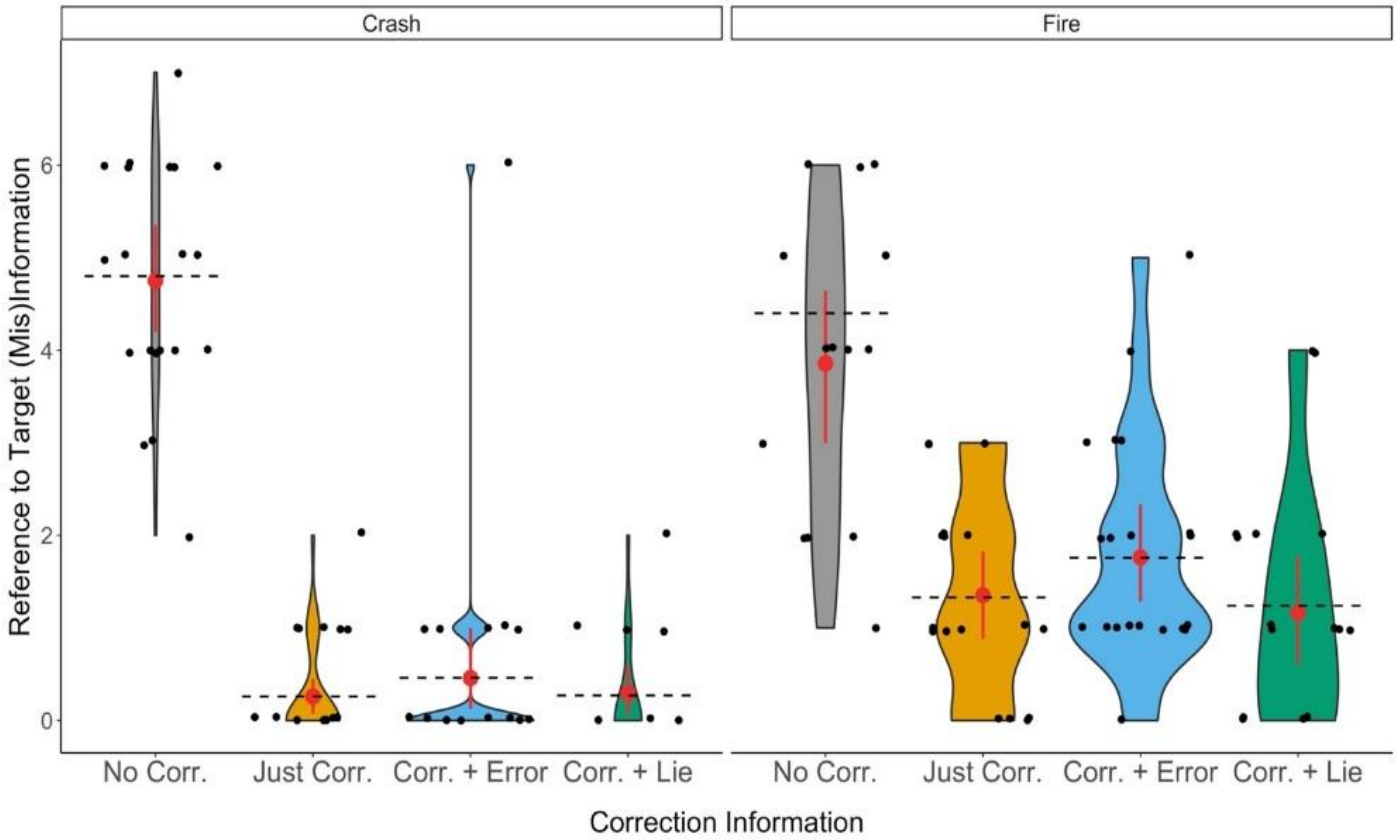

Figure 16 Distribution and probability density of references to target (mis)information by correction information condition in Experiment 4. Red points represent mean and error bars represent 95% confidence intervals of the mean. Dashed lines represent condition means after excluding participants who did not recall the critical information, based on data from 10 (crash: no corr.), 13 (crash: just corr.), 20 (crash: corr. + error), 10 (crash: corr. + lie), 5, (fire: no corr.), 12 (fire: just corr.), 15 (fire: corr. + error), and 12 (fire: corr. + lie) participants, respectively.

3.14.2. Inference Scores

Figure 16 shows the distributional characteristics of inference scores as a function of correction information, separately for the crash and fire reports. It is clear from Figure 16 that all three types of correction substantially reduced reference to misinformation (almost to zero) for the crash report but not for the fire report. Comparatively the proportion reduction in reference to misinformation was 90-95% for the crash report and 55-70% for the warehouse fire report.

Likelihood ratio tests[32] indicated that the best fitting model for the data was one that included an interaction between correction information and scenario, χ^2 (3) = 36.02, $p < .0001$. The letters in the final column of Table 5 represent the significance of Dunnett's multiple comparison tests. Groups that share the same letter are not significantly different from each other. The table shows that all three types of correction reduced the number of references to target (mis)information for both the warehouse fire and crash reports. The differences between the three types of corrections were not significantly different from each other within each scenario. When comparing across scenarios, participants produced significantly fewer references to misinformation following a correction that explained misinformation as an error in the crash than the fire report. The two conditions featuring just a correction also differed between the two different reports. The analysis therefore confirms the pattern of results that can be seen in the violin plots.

[32] Here, a negative binomial model was deemed a best fit for the data because the inference score variance was greater than the mean.

Table 7 Marginal inference score means and post-hoc comparisons Experiment 5

Correction Information	Scenario	Estimated marginal mean	Group
Just Corr.	Crash	0.26	a
Corr. + Lie	Crash	0.29	ab
Corr. + Error	Crash	0.46	abc
Corr. + Lie	Fire	1.17	bcd
Just Corr.	Fire	1.35	cd
Corr. + Error	Fire	1.76	d
No Corr.	Fire	3.86	e
No Corr.	Crash	4.75	e

Note: Confidence level used: 0.95. Confidence level adjustment: Dunnett method for 8 estimates. P value adjustment: Dunnett method for 28 tests. Significance level used: alpha = 0.05 are shown in the final column. The comparisons that share the same letter group are not significantly different from each other.

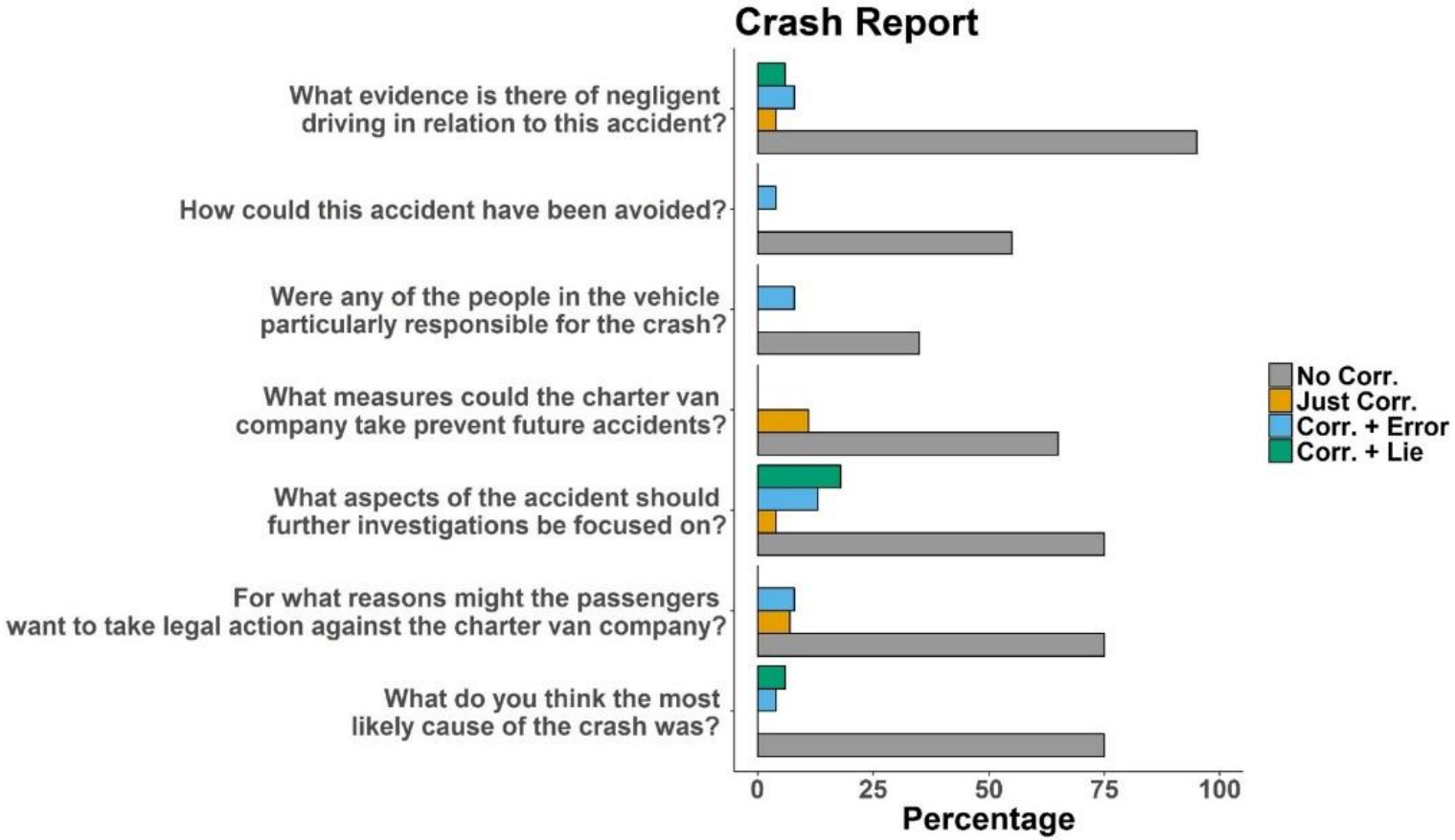

Figure 17 Proportion of references to target (mis)information by inference question and correction information for the crash scenario in Experiment 5.

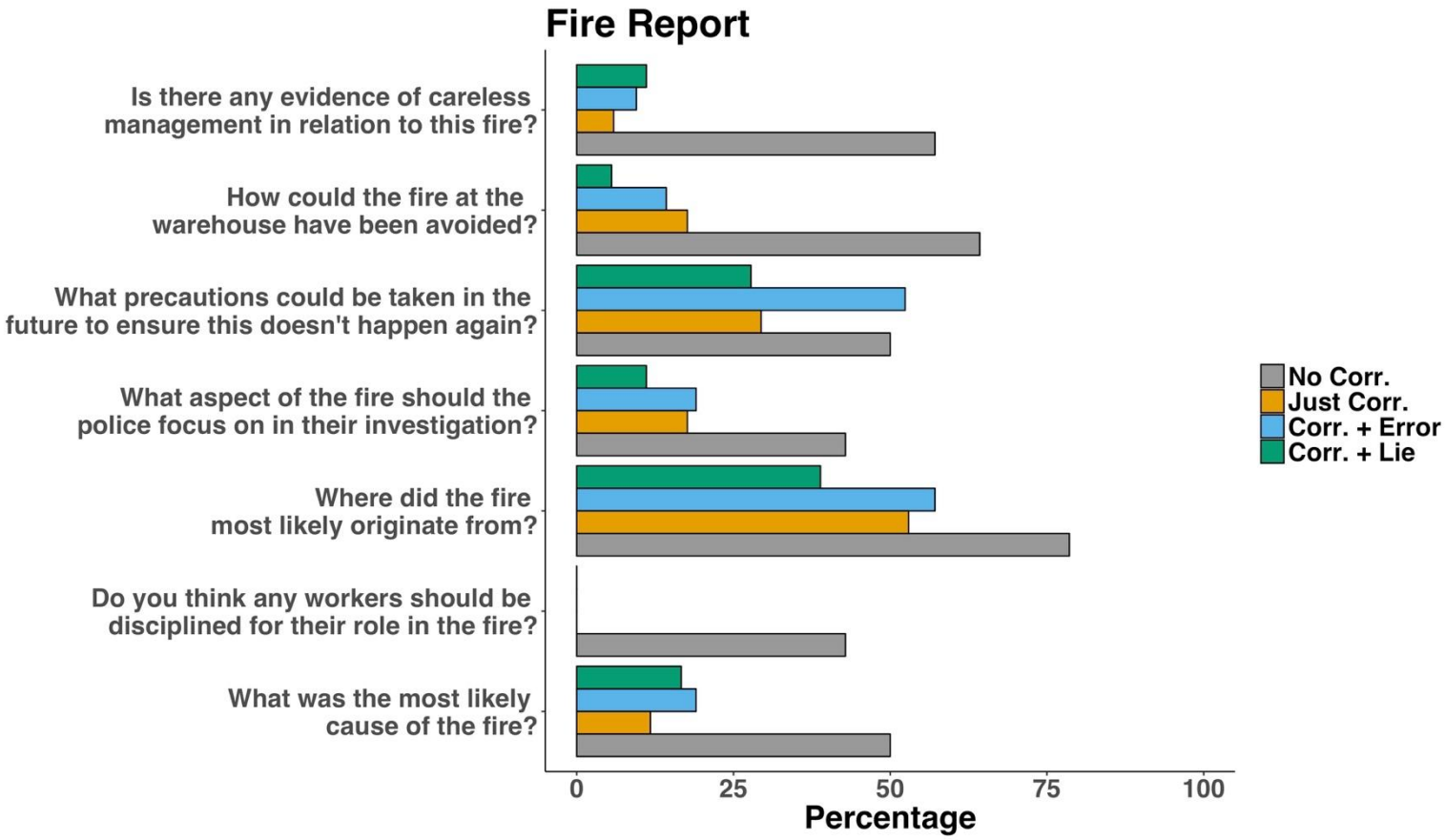

Figure 18 Proportion of references to target (mis)information by inference question and correction information for the fire scenario in Experiment 5.

Question analysis. The proportion of references to misinformation was examined separately for each question that followed the crash (Figure 17) and warehouse fire (Figure 18) reports. Figure 17 shows that most questions about the crash report did not elicit references to misinformation, which is unsurprising given that no CIE was observed for this scenario. Figure 18, on the other hand, shows that the questions most likely to elicit a reference to misinformation asked about the origins of the fire and what precautions could be taken to prevent future fires. There were no consistent patterns in regards to explanatory and non-explanatory conditions.

Correction acknowledgement. The proportion of participants who referred to misinformation even though they acknowledged that the information had been corrected for each is shown in Figure 19. The figure shows that the proportion of participants who made at least one uncontroverted reference to misinformation whilst also acknowledging that the information was corrected was substantially higher for the fire report (60-90%) than the crash report (18-25%). This is not surprising given that no continued influence effect was observed for the crash report. One participant who read the fire report responded that "… Gas and paint cans should be stored in a different place and labelled with big letters …" Then in response to the question asking about the point of the second message from Fire Chief Lucas they said: "The point of the second message from Fire Chief Lucas is to explain that the cause of the fire was not from gas cans and paint cans. Fire Chief Lucas wanted to clear that statement because any further news will continue to use that statement and believe this was the cause when it was a different cause to the fire".

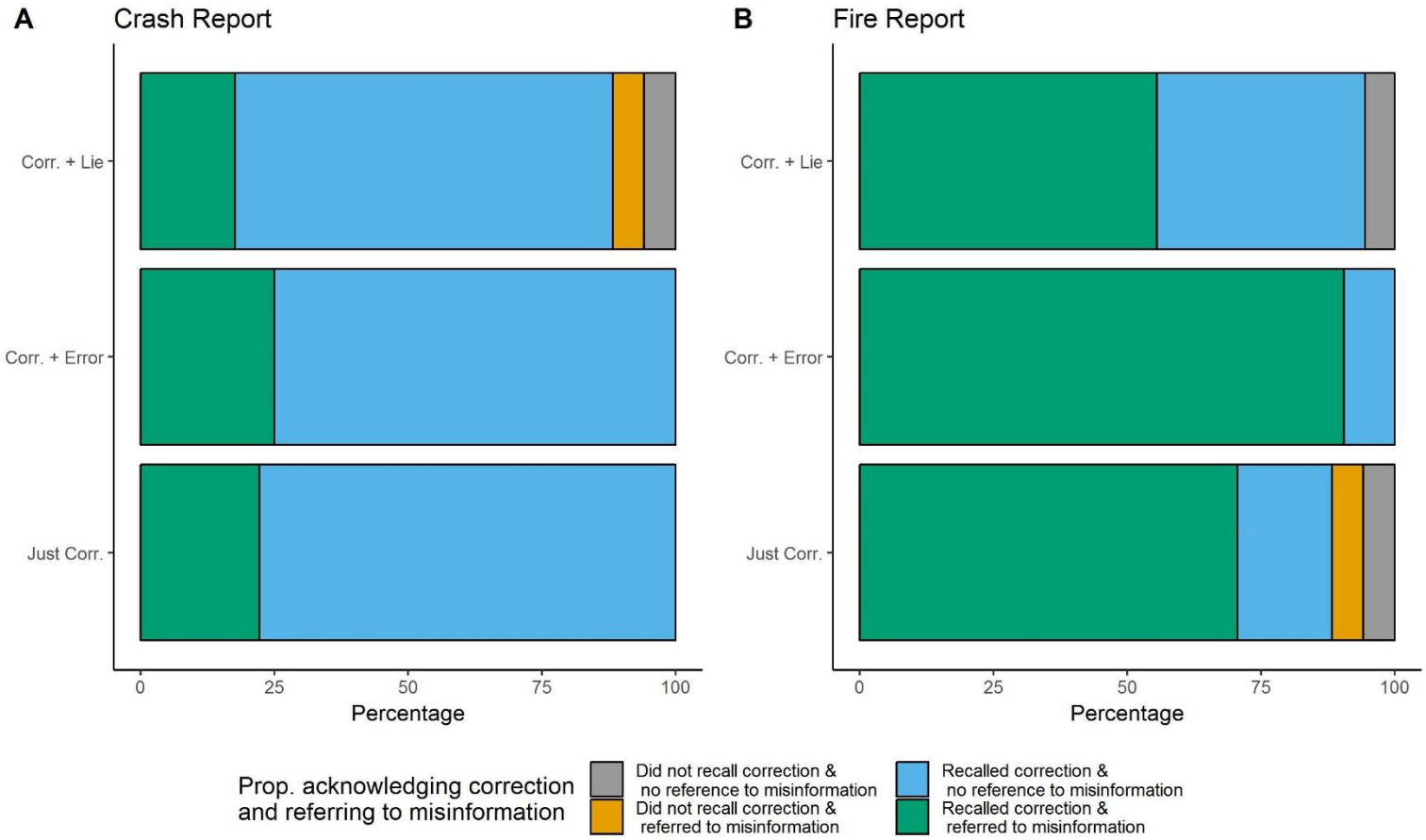

Figure 19 Proportions of participants who recalled or did not recall the correction, and either referred or did not refer to target (mis)information, by correction information condition and scenario in Experiment 5.

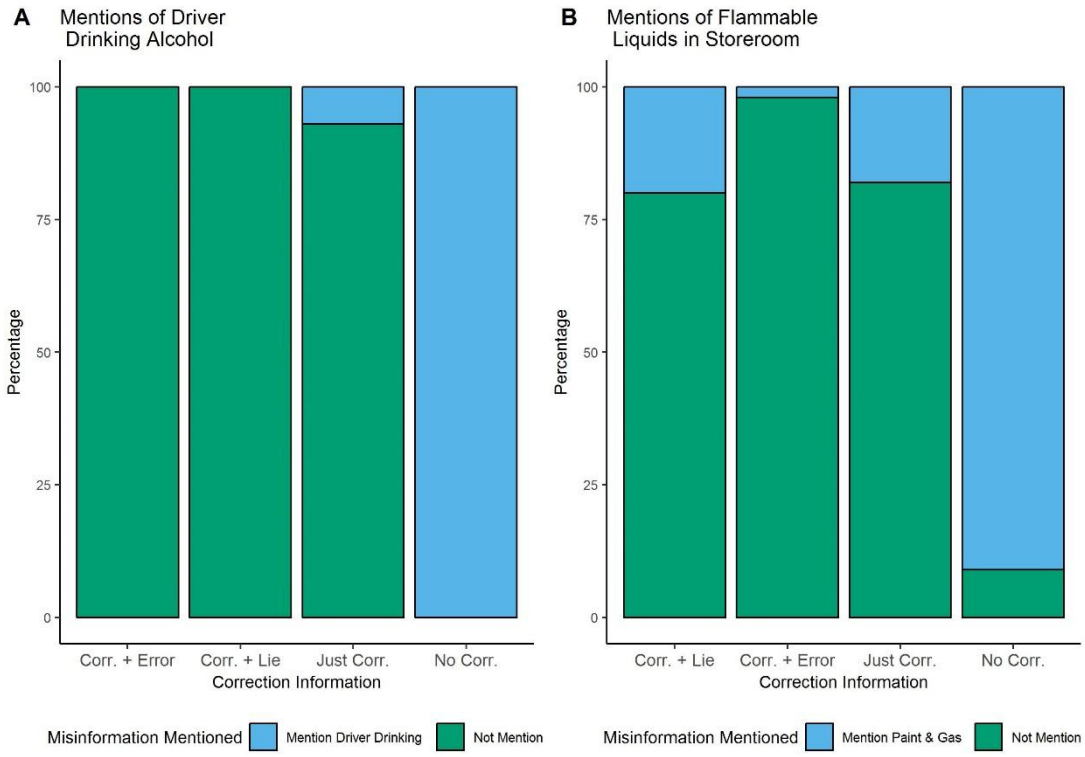

Figure 20 Proportion of references to misinformation on the recall question probing recall of target (mis)information as a function of correction information condition for the (A) crash scenario, and (B) warehouse fire scenario in Experiment 5. Bars for conditions featuring a correction represent responses after excluding participants who did not recall the

3.14.3. Recall Accuracy Scores

A null Poisson regression model was compared to the full Poisson regression model including an interaction term for correction information and scenario. The full model was not a significantly better fit for the data than the null model, χ^2 (7) = 84.18, p = .30. This indicates that the number of correctly recalled literal details did not differ between conditions

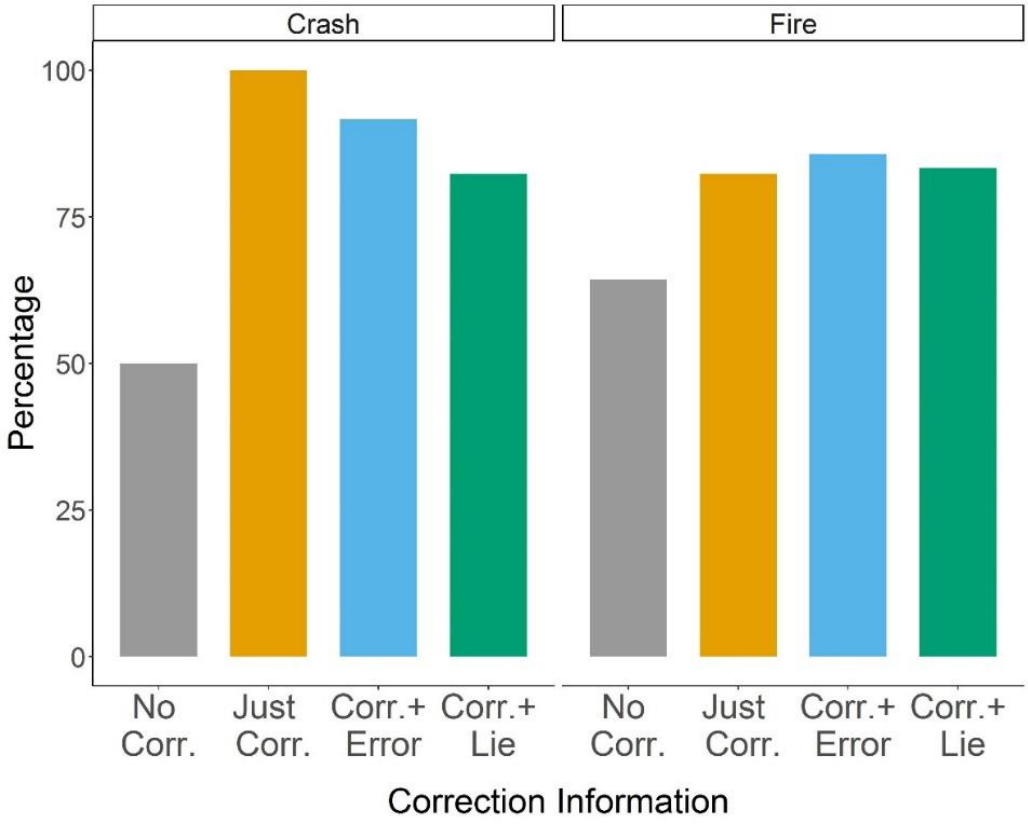

Figure 21 Critical information recall as a function of correction information. Shown separately for crash and warehouse fire reports in Experiment 5.

3.14.4. Critical Information Recall

Figure 21 shows that critical information recall was poorest for the groups who received the control message. This was true of both the crash and fire reports. Critical information recall was slightly poorer for participants reading the warehouse fire report. There was a significant association between correction information and accurate recall of the correction for the crash report, $\chi^2(3) = 93.45$, $p < .001$, when including the no correction condition. The association was still significant when excluding the no correction $\chi^2(2) = 19.72$, $p < .001$, suggesting correction recall performance was poorer in the two explanatory conditions. In contrast, there was a significant association between correction information condition and critical information recall when including the no correction condition, $\chi^2(3) = 85.89$, $p < .001$, but this disappeared when excluding the no correction group, $\chi^2(2) = 0.44$, $p = .80$.

3.15. Summary

Experiment 5's results provide further evidence that explanatory and non-explanatory corrections are equally effective at reducing reliance on misinformation. Participants were equally likely to refer to misinformation whether they received an explanatory correction or not; this was true of both the warehouse fire and crash reports. However, the CIE was substantially attenuated – or almost eliminated - in the crash report. The most probable inference score for the crash report was zero suggesting a complete correction. Experiment 5's results therefore showed corrections were far more effective at reducing reliance on misinformation for the crash report than the warehouse fire report, thereby replicating Experiment 4's findings. These results provide further evidence that the difference between the warehouse fire and crash reports lies in the type of 'base' correction presented to participants. Participants who received the warehouse fire and crash reports made an equivalent number of references to target (mis)information when this information was uncorrected. However, misinformation references were significantly lower when comparing differences between correction and correction + error groups across reports.

The fact that uncorrected groups did not differ but correction groups did, suggests that the difference was not related to the type of misinformation presented per se, but the way misinformation was corrected. The correction in the warehouse fire report negated the presence of flammable liquids in the building. However, the correction did not provide any evidence to counter the claim that that careless storage of oil paint and compressed gas caused the fire, thus leaving open the possibility that the gas cylinders and oil paints were somehow responsible for the fire. In contrast, the correction in the crash report ruled out the possibility that the driver's intoxicated state caused the van to crash. This is an important distinction between types of correction which deserves further investigation. The difference between reports could be related to the story content per se. However, this seems unlikely given that inference scores were similar in uncorrected conditions, but not the corrected conditions. Alternatively, there could be an interaction between story content and the characteristics of the correction.

Another interesting observation was that there was a discrepancy between inference and recall scores for participants who received the error correction in the fire story (and which also replicates the pattern of results found in Experiment 3). These participants were less likely to refer to oil paint and gas cylinders when asked what the contents of the storeroom were before the fire than the correction or correction + lie groups. Put another way, the correction + error group were more likely to answer this question correctly whilst also apparently maintaining the misinformation. One possible explanation for the discrepancy between the responses of the correction + error group could be that some participants made local but not global updates to their model of the event (Albrecht & O'Brien, 1993; McKoon & Ratcliff, 1992). The other two types of corrections did not provide information which replaced the oil paint and gas cylinders with other (non-flammable) materials. The error correction group may have made local change to contents of the storeroom whilst maintaining their global model of the event in which oil paint and gas cylinders caused the fire. The other two types of corrections did not provide information which replaced the oil paint and gas cylinders with other (non-flammable) materials so they were unable (or unwilling) to make a local or global change.

3.16. Conclusions

The aim of the three experiments reported in this chapter was to investigate whether providing an explanation where misinformation occurred can reduce the CIE. Legal decision-making, continued influence studies, and studies on text refutations, suggest that explaining why a given piece of information should be disregarded reduces its impact on later judgments and reasoning (e.g. Bush et al., 1994; Rapp & Kendeou, 2008; Steblay et al., 2006). The experiments reported in this chapter tested two types of explanatory corrections across multiple scenarios. The explanatory correction that appeared in all scenarios either described initial misinformation as an (unintentional) error or as an (intentional) deception. It was expected that explanatory corrections would be more successful at reducing the CIE than non-explanatory corrections. It was also (tentatively)

expected that correction which explained misinformation as a lie would be more effective than a correction that explained misinformation as an error.

These findings provide evidence that corrections that explain misinformation as an honest mistake, or as a lie, had no more impact than a negation of misinformation. This does not mean to say that corrections that explain how misinformation occurred are always equivalent to non-explanatory corrections. The explanations provided in the present set of experiments may not be satisfactory explanations for how misinformation occurred, and therefore have little impact over and above the correction which negated misinformation. Research on what makes a good explanation has shown that people prefer simple to more complex explanations (Lagnado, 1994; Lombrozo, 2007), and explanations that are broader in scope (or account for more information) than narrow scope explanations (Read & Marcus-Newhall, 1993).

It is also worth noting that causal explanations – the explanations used in the present set of experiments provided a *causal explanation* for why the interlocutor initially believed misinformation but now does not – are a form of social interaction and therefore subject to rules of conversation (Hilton, 1990). The social nature of the explanations offered in the present set of experiments could have constrained the effectiveness of the correction. As such, any pragmatic benefit gained from the explanation of the misinformation's invalidity might be nullified because it creates a whole new set of inferences about why an employee lied, or made an error.

These present findings are inconsistent with Bush et al.'s (1994) findings showing that explanatory corrections (which address the conversational implications of the contradiction) are more effective at reducing continued reliance on misinformation. One reason for the difference between the present findings and Bush et al.'s findings could be that corrections in their study did not explain the invalidity of misinformation in terms of human communication. Human communicators are supposed to conform to conversational norms such as providing true and relevant information. The fact that the person who initially conveyed the misinformation failed to provide true or relevant information could result in

people disregarding the explanation and focusing on the negation (i.e. that there were no flammable substances in the storeroom). Furthermore, some courtroom simulation studies suggest that people are more convinced by physical evidence than eyewitness evidence (Skolnick & Shaw, 2001), perhaps because people assume that human measuring devices (e.g. eyewitness) are inherently less reliable than physical ones (e.g. CCTV; see Lagnado et al., 2013). Explanatory corrections may have been more effective if the correction involved a physical explanation of why misinformation is incorrect – for example, if the correction had stated that oil paint and gas cylinders had been moved to a fire-secure room in a different building and therefore could not have caused the fire. This would of course need to be tested empirically to confirm whether this was the case.

Although this chapter's aim was to examine the effectiveness of explanatory corrections to misinformation the findings also indicated some potential constraints on the CIE. Published research on the CIE has tended to demonstrate effects under consideration using one scenario rather than comparing across scenarios.[33] Experiments 4 and 5 provide preliminary evidence that the CIE does not occur when the correction invalidates the explanation offered by the misinformation. These findings do not provide conclusive evidence that the CIE is constrained in these ways and is a post-hoc explanation of the findings. However, the findings do highlight the need for a more systematic examination of the CIE and the circumstances under which it is likely to occur and what types of misinformation, correction, or scenario give rise to the effect. These explanations also fit with the mental-model updating account of the CIE (Gordon et al., 2017; Johnson & Seifert, 1994; Lewandowsky et al., 2012; Wilkes & Leatherbarrow, 1988). Participants were able to abandon their initial incorrect model in favour of the correction one when the discrepancy between models was made more salient by providing evidence against the misinformation.

[33] Earlier research in this field examined misinformation effects across two reports and found some small differences in reliance on misinformation as a function of story (Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988)

There are several important implications that can be drawn from the experimental results reported in Chapter 3. First, the experiments provide additional evidence that similar results using the typical CIE are obtained online and in the lab. This suggests that the magnitude of the CIE is not increased in more educationally and experientially diverse populations such as those recruited through online labour markets. Experiment 4 was performed in the lab and produced very similar results to the experiments reported in Chapter 2, as well as in this chapter. The results of the present set of experiments are also consistent with previous lab-based studies (Ecker et al., 2011a; Ecker et al., 2011b; Guillory & Geraci, 2010). The second important observation is that the CIE observed here is smaller than in previous studies that have used the warehouse fire scenario (e.g. Johnson & Seifert, 1994; see also the experiments reported in Chapter 2 which used identical stimuli to Johnson & Seifert, 1994). The difference between studies could be due to the inherent ambiguity of the erroneous message about the contents of the storeroom. The misinformation and correction may raise questions about where the information about the oil paint and gas cylinders came from. There is also ambiguity about where the paint and gas are if they are not in the storeroom. The studies reported in this chapter made it unambiguous that there was no paint and gas on the premises, let alone in the storeroom.

4 The continued influence of implied and explicitly stated misinformation across multiple scenarios

4.1. Chapter Overview

The previous chapter considered different ways in which misinformation could be corrected. Findings showed that corrections that explained the origins of misinformation as an unintentional error or intentional lie were not more effective than corrections that negated misinformation. This chapter focuses primarily on the nature of the misinformation itself rather than the correction. An additional focus of this chapter is to move from using a single scenario with various manipulations, to using multiple scenarios.

Previous CIE research has consistently shown that causal misinformation – information that provides causal structure to a description of an event – is difficult to correct (Ecker et al., 2011; 2011a; 2011b; Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988). Findings from a recent CIE study suggests that misinformation which explicitly states a cause of an adverse outcome is more easily corrected than misinformation that implies the cause of an adverse outcome (Rich & Zaragoza, 2016). This finding has implications for the ways in which journalists and newsreaders infer causality when reporting about unfolding events. Erroneous information reported as events are still developing may have more of a continuing impact if it implies rather than explicitly states a likely the cause of an event.

One limitation of previous work showing that implied misinformation is more resistant to correction than explicitly stated misinformation is that it has only been demonstrated with a single scenario or news story. Chapter 4 extends existing research by examining the continuing influence of implied and explicitly stated misinformation in three different news stories (warehouse fire, crash, and head injury), across two experiments. The results of the experiments reported in this chapter showed no difference in the effectiveness of corrections to implied and explicitly stated

misinformation. Results also showed that a correction to misinformation almost eliminated reliance on misinformation for the crash scenario but resulted in a CIE for the warehouse fire and head injury scenarios. Finally, the inclusion of a control condition in which misinformation was only mentioned as it was being corrected showed that people refer to misinformation even if it does not form the basis of an initial mental model of the event.

4.2. Introduction

In the rush to break 'big news' stories, broadcasters often report incomplete, inaccurate, or mistaken information. Significant inaccuracies, and misleading or distorted information, must be promptly corrected and an apology published where appropriate (Independent Press Standards Organisation, 2016). One real-world example of this occurred during the coverage of the Westminster terror attack when Channel 4 news named the wrong man as the attacker (Sweney, 2017). Channel 4 news quickly rectified their mistake and issued a correction during the same news programme. Even when a correction is quickly issued, erroneous information (or misinformation) may still have had a lasting impact on how people interpret events and form impressions. For instance, the correction issued by Channel 4 news may not have been sufficient to counteract the reputational damage caused by naming the man as the attacker.

The journalistic code of conduct stipulates that the press must be careful not to publish inaccurate or misleading information (Independent Press Standards Organisation, 2016). In the event that this happens the information must be swiftly corrected or otherwise the news organization can face possible legal action. In the race to be the first to break a story the drive to report available information – even if its accuracy is uncertain – may be too tempting. When the likely cause of an unfolding event is unknown reporters may *imply* rather than *explicitly state* a likely cause in an attempt to circumvent legal ramifications. For example, when accounting for a celebrity's death, a reporter might suggest that the death was drug related by mentioning the famous person's history of drug use, or might directly assert

Figure 22 Example of news headlines that imply (top) and explicitly state (bottom) the cause of an adverse outcome

that the cause of death was a suspected overdose (a real example of this is depicted in Figure 22).

Although a fairly trivial example, it is not hard to see how misinformation that implies a likely cause of outcome could be problematic and lead to causal misunderstandings. There are many cases in which rumour and innuendo which implies likely causes of events or outcomes can circulate, particularly through social media, potentially affecting people's understanding of what occurred. For instance, Seifert (2002) describes a real case in which a news station reported that a family had been found dead in their home after having eaten at a Chinese restaurant. The news station reported a few days later that the deaths were caused by a faulty furnace. The conclusion of the news story was that the Chinese restaurant had closed; seemingly because the news story implied a relationship between the Chinese restaurant and the deaths. The news story never explicitly stated that the Chinese restaurant caused the deaths suggesting that this inference would have to be produced by the person hearing or reading the story.

Continued influence research has consistently shown that misinformation which provides a likely cause of an event is difficult to correct (Ecker, Lewandowsky, Cheung, & Maybery, 2015; Ecker et al., 2011a; Johnson & Seifert, 1994; Wilkes & Reynolds, 1999). Recent developments on this topic suggest that misinformation that implies rather than explicitly stating the cause of an adverse outcome might be more difficult to correct. In their study, Rich and Zaragoza (2016) examined the effectiveness of corrections to implied and explicitly stated misinformation in a scenario describing a theft of valuable jewellery from a couple's home while the couple was on vacation. The misinformation initially presented as a likely cause of the theft was that the couple's son had taken the jewellery from the house. The misinformation either implied the son's involvement by stating that the couple had asked their son to check in on the house while they were away, or explicitly stated that the police suspected the son had taken the jewellery from the house. In the implied case participants had to infer that the son had stolen the jewellery whereas this information was unambiguously provided in the explicit case. Later in the story, participants in the correction condition learned that the son had actually been out of town while the theft occurred, thereby invalidating the initial misinformation implicating the son's involvement in the theft.

Surprisingly, Rich and Zaragoza found that participants generated more post-correction references to implied misinformation than explicitly stated misinformation. The son's involvement in the crime was also rated as more likely when misinformation was implied rather than explicitly stated. There was no difference between implied and explicitly stated conditions when misinformation remained uncorrected. These findings showed that corrections to misinformation that explicitly stated the son's involvement in the crime were more effective than misinformation that implied the son's involvement in the crime. Furthermore, providing an alternative explanation informing participants that the actual thief had been caught led to an even larger correction effect for both explicit and implied misinformation conditions. There was also an interaction between correction and

misinformation, in that the effect of correction was larger in the explicit than the implied condition.

Rich and Zaragoza (2016) argued that a possible explanation for the findings is that people have to go beyond the available information to infer causal relations between story elements in the case where misinformation is implied. This in turn results in a more elaborate mental model of the described event. There is precedent for this explanation of Rich and Zaragoza's findings. For instance, it is well established that people generate causal inferences during narrative comprehension (Graesser, Singer, & Trabasso, 1994; Kintsch & van Dijk, 1978; Singer, Graesser, & Trabasso, 1994; Trabasso & van den Broek, 1985).

There is also evidence to suggest that readers generate inferences between story elements when causal relations between pieces of information are not explicitly stated. For example, Pennington and Hastie (1988; Exp 1) found that participants 'falsely' recognized probe sentences that were consistent with a verdict decision they had made on the basis of evidence items they had previously seen[34]. In this study, participants read trial materials and a series of sentences that represented evidence items. After reading information and evidence about the case, participants decided on a verdict and then completed a recognition test. The recognition probes used at test consisted of: target items that were consistent with a guilty or not guilty story (i.e. that participants in a previous study had mentioned as part of their explanation for their verdict choice), items from both stories, items from neither story, and critical lure items that were descriptions of plausible case-relevant events but had not been included as evidence. The results showed that verdict decisions predicted higher hit and false alarm (i.e. rates of responding 'yes' to non-presented items) for sentences in the stories corresponding to participants' verdict decisions. The rate of falsely recognizing critical lure sentences was higher for verdict story consistent

[34] The evidence items used for this study were gleaned from a previous study (Pennington & Hastie, 1986) in which participants watched a simulated murder trial and described their evaluations of the evidence and verdict through a 'talk-aloud' procedure. Item story membership (guilty vs. not guilty) was determined according to the verdict story it originally came from.

items than for the alternative verdict story items. These findings suggest that people go beyond the available information to infer causal links between pieces of story information when they are not explicitly provided.

Rich and Zaragoza (2016) argued that implied misinformation was more resistant to correction than explicitly stated misinformation because people construct a more elaborate representation of a described event when they have to self-generate causal links between story elements (Duffy, Shinjo, & Myers, 1990). Therefore, people have to infer the causal link between the misinformation (i.e. the son checking on the house) and the outcome (i.e. the jewellery box was missing) which makes it more difficult to correct. Consistent with this, Davies (1997; see also Mussweiler & Neumann, 2000 for a similar finding) found that people were more likely to discount discredited explanations when they were externally provided than when they were self-generated. Participants were presented with short summaries of fictional psychological experiments including methods, procedures, and findings. They were either asked to self-generate possible explanations for the findings or were provided with an explanation for the findings that had been generated by another participant. Participants were then told that the experiments were actually fictitious and that the case studies had actually been invented to illustrate some important methods and procedures in psychology. When asked to estimate the likelihood of the reported outcome if the experiment were actually to be carried out, participants who self-generated explanations for the findings exhibited significantly more belief persistence in the discredited findings than participants who received other-generated explanations. Implied misinformation may therefore be harder to correct because self-generated judgments or inferences require elaborate internal processing of information and internally generated information is less likely to be seen as contaminating (Mussweiler & Neumann, 2000).

Rich and Zaragoza's findings may not necessarily result from more 'elaborate' event representations constructed when self-generating causal links. It may be the case that the ambiguity of causal misinformation presented makes this information more easily accessible or salient. The evidence presented seems merely to suggest that inferences made by

participants had more of an effect than those externally presented, but this could occur for a number of reasons. Implied misinformation might also be difficult to correct because of the pragmatic inference people draw about why the information was supplied. Evidence for this comes from, Wegner, Wenzlaff, Kerker, and Beattie (1981) who gave participants different types of newspaper headlines which insinuated (e.g. “P is a criminal?”), or directly incriminated an individual. Source credibility affected the persuasiveness of directly incriminating assertions but had less of an impact on innuendo. Therefore the perceived credibility of sources offering implied and explicitly stated misinformation could mediate continued reliance on this information.

4.3. Experiment 6

Rich and Zaragoza (2016, p. 9) acknowledged that their findings were limited because they were only obtained with a single news story. This limitation is important because story content may interact with an individuals’ pre-existing knowledge and beliefs moderating the effects of implied and explicitly stated misinformation (as discussed in Chapters 1 and 3). For instance, Ecker et al. (2014) found that participants’ pre-existing attitudes (racial prejudice toward an ethnic minority group) influenced how they used race related information – although not processing of a correction. Experiment 6’s aim was, therefore, to validate Rich and Zaragoza’s (2016) findings via a conceptual replication of the finding with two different scenarios.

The reproducibility of experiments is considered a fundamental tenet of the scientific approach. This is particularly the case in psychological science which has recently faced a series of failed replications of findings from social psychology (e.g. Shanks et al., 2015; Shanks et al., 2013). The failure to replicate several findings has sparked controversy, highlighting ‘questionable research practices’ and triggering a methodological crisis in psychological science (Open Science Collaboration, 2012, 2015; Simons, 2014). The validity of claims based on psychological findings therefore hinges on their replication.

Conceptual replications may be particularly important for making policy recommendations because they test the rigour of the underlying hypothesis rather simply duplicating the sampling and experimental procedures. Conceptual replications have implications for the generalisability of findings whilst direct replications arguably test the reliability of the measure employed (Makel, Plucker, & Hegarty, 2012). The CIE appears to be a robust effect; however, effects of interest are typically only examined in a single scenario (as discussed in Chapter 1). Recommendations about the way corrections are designed, structured, and applied have been made on the basis of CIE research (e.g. Ecker et al., 2014; Lewandowsky et al., 2012), making it all the more important to establish the precise conditions under which the CIE occurs.

Experiment 6 used the same 'breaking news' format used in Chapter 3's experiments. The crash and warehouse fire reports (used in Experiments 4 and 5) were used in order to test whether explicitly stated misinformation is more easily corrected than implied misinformation. If story content is related to misinformation type, then these two factors should interact in some manner. The structure of the news reports used in Chapters 3 and 4 differs from previous CIE studies in two key ways (also discussed in Chapter 3). The first was that the statement provided immediately before initial misinformation included information about potential causes of the adverse outcome. This information was included in order to increase the possibility that participants would generate other likely causes of the outcome than the one provided in the misinformation. There were two major reasons for including this information. First, there is a reasonable pragmatic assumption that if you are asked questions about a story you have read, the answers to those questions will be in the story. Therefore, people could be mentioning the misinformation, knowing that it has been corrected, but because no other potential causes have been presented, participants feel they have to say something, even if though they know the causal information they have provided was corrected. Second, there is a potential that the CIE can be partially explained by the availability in memory of the causal explanation offered by misinformation (cf. Anderson, New, & Speer, 1985). That is,

participants do not encode the story as a causal whole; instead, when asked questions, they attempt to retrieve snippets of relevant information. If the only potential cause available in memory is the misinformation, and there is no causal model, then participants will tend to refer to it. Adding other potential causes goes some way towards addressing the issue.

Second, additional information about the event which was provided either side of the target (mis)information and critical correction information was not congruent with the explanation offered by the misinformation (i.e. it could not be interpreted as either supporting or refuting conclusions drawn from initial misinformation). This contrasts with the jewellery theft story used in Rich and Zaragoza (2016). The jewellery theft story included additional information that was congruent with the misinformation implying or directly stating that the son had stolen the jewellery (police are still attempting to determine whether other valuables are missing from the home, the television and home computer, however, had not been disturbed).. Providing additional event information that is congruent with the misinformation explanation offered (implied or explicitly stated) could result in participants placing more weight in the misinformation than the correction. The fact that the additional event information included in the present stories was designed not to be congruent misinformation could reduce reliance on misinformation relative to Rich and Zaragoza's (2016) results.

4.4. Method

4.4.1. Participants

Due to financial constraints no power calculation was performed to estimate the sample size. The aim was instead to recruit 20 participants per cell of the experimental design (N = 160). In total 168 (67 females and 101 males, aged 22 to 70, M_{age} = 38.18, *SD* = 11.11) US based participants were recruited from Amazon Mechanical Turk. Participants were paid $1.50 and took an average of 18 minutes to complete the experiment. In addition to the standard reward, participants were given the opportunity to earn an additional $10 based on high accuracy scores across instruction check and

fact recall questions[35]. One-hundred and fifteen (68%) of participants answered all three instructional attention check questions correctly.

4.4.2. Stimuli, Design & Procedure

The stimuli were generated in Qualtrics (Qualtrics, Provo, UT). Participants read one of 8 versions of a fictional news report that either described a warehouse fire or a crash, each consisting of 12 discrete messages. Figure 23 illustrates how message content was varied across experimental conditions, as well as the format used to present messages. The effect of correction information (No Correction, Correction), Scenario (Warehouse Fire, Crash), and Misinformation Type (Explicitly Stated, Implied) on reference to target (mis)information was assessed between groups; participants were randomly assigned to one of the 8 experimental groups (N = 16-25 in each group). The stimuli were identical to those used in Experiment 5 with the addition of the 'explicitly stated' misinformation conditions. The stimuli have been described in more detail in Chapter 3 so only the new conditions are described here (full details in Appendix J).

In the warehouse fire report the target (mis)information presented (Message 5) either implied (Fire Chief Lucas issues statement: "Cans of oil paint and pressurized gas cylinders were present in storeroom before fire") or explicitly stated (Fire Chief Lucas issues statement: "Investigation team suspect fire caused by carelessly stored flammable liquids. Cans of oil paint and pressurised gas cylinders were present in storeroom before fire") that oil paints and gas cylinders were a suspected cause of the fire. Each message was presented as the 'latest news' in a breaking news format, and was no longer than 280 characters. Messages in the same position within the sequence were matched for length across scenarios. Messages were presented individually for a minimum of 5 seconds; there was no maximum reading time. Participants clicked a button to proceed to the next message and were unable to return and view previous messages. After reading the

[35] The modified recognition test discussed in Chapter 3 was also included in Experiment 6. However, due to a programming error the wrong test was shown to participants. Responses to this question could therefore not be used to exclude participants.

report, participants completed a 16-part questionnaire consisting of 7 inference questions, 7 fact recall questions and 2 critical information recall questions (these are discussed in detail in Chapter 3; the questions can be found in Appendix I).

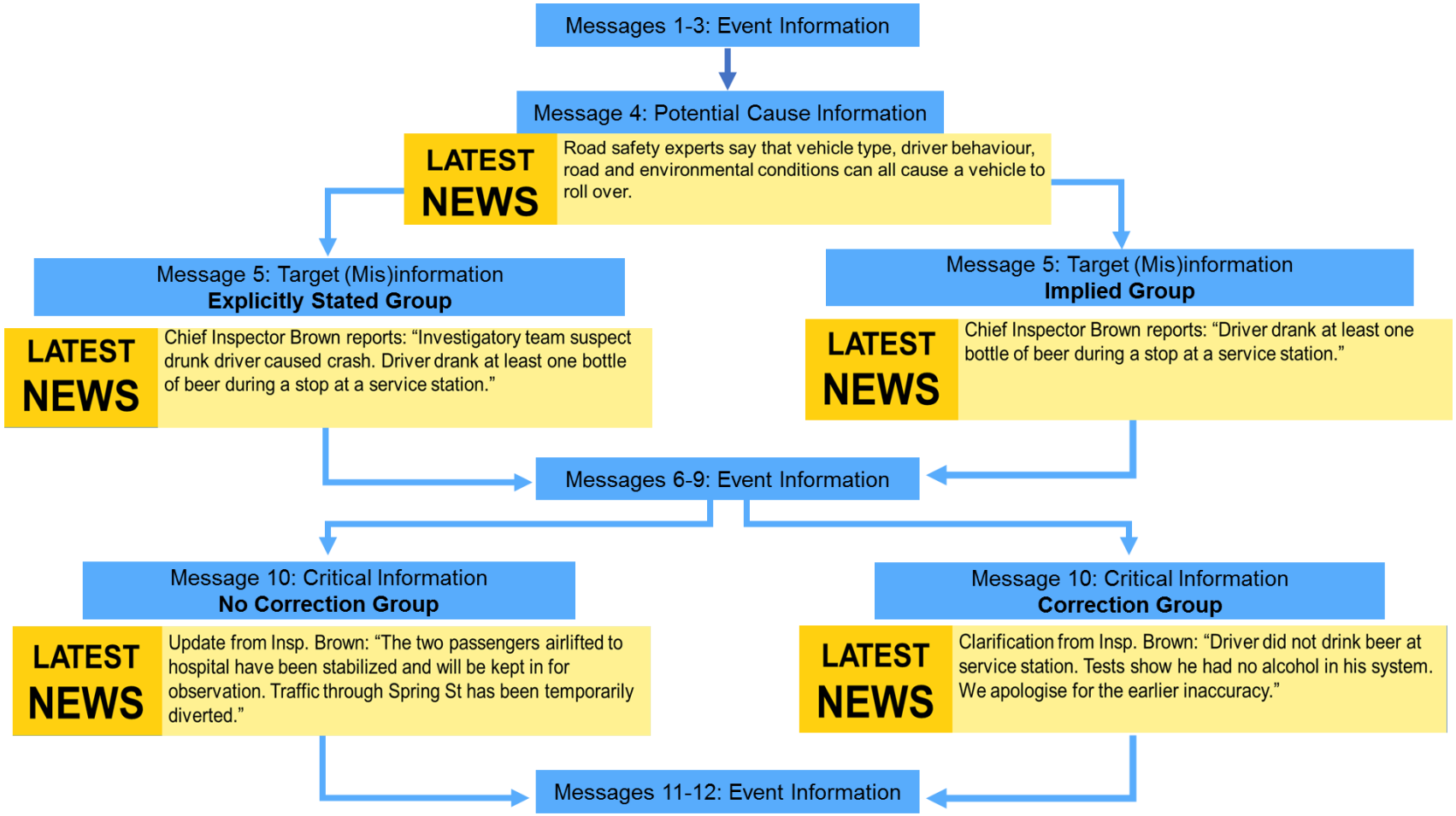

Figure 23 Schematic diagram depicting experimental manipulation of misinformation and correction information for the 'crash' report in Experiment 6.

4.5. Results

4.5.1. Coding of Responses

The criteria for coding responses are identical to those used in Chapter 3 (and can also be found in Appendix I). Table 8 shows examples of responses that was coded as a reference to target (mis)information and an example of a response that were not coded as such. References to flammable materials in the warehouse fire scenario which did not specifically mention storage (e.g. “It could have been avoided by keeping flammable objects or items in place”) were not treated as references to misinformation because there was no specific mention of gas, paint, liquids, substances, chemicals, or the fact they were (allegedly) kept in the storeroom. Similarly, in the crash scenario references to driver behaviour that did not mention intoxication or drunkenness were not counted as references to misinformation (e.g. “by having him be more alert drinking coffee”). The maximum individual score for inference questions was 7. Responses to factual questions were scored for accuracy; correct or partially correct responses were scored 1 and incorrect responses were scored 0; the maximum factual score was 7. Critical information recall scores were computed using the same criteria; the maximum individual critical information recall score was 2.

Inter-coder reliability. Responses were coded by a trained coder. A second, independent judge then coded 10% of responses from each narrative. Inter-rater agreement was 0.95 and Cohen’s K = 0.89±0.03, indicating a high level of agreement between coders, both of which are higher than the benchmark values of 0.7 and 0.6 (Krippendorff, 2012; Landis, & Koch, 1977), and there was no systematic bias between raters, $\chi^2 = 1.92$, $p = .17$.

Table 8 Example of inference response coding in Experiment 6

Scenario	Inference Question	Example of Response Scored 1	Example of Response Scored 0
Crash	How could this accident have been avoided?	If the driver had not been drinking.	He was in a court battle with his ex-wife.
Warehouse Fire	How could the fire at the warehouse have been avoided?	They should store flammable substances separately.	Pay more attention to the signs and smells of a fire. Not overworking workers.

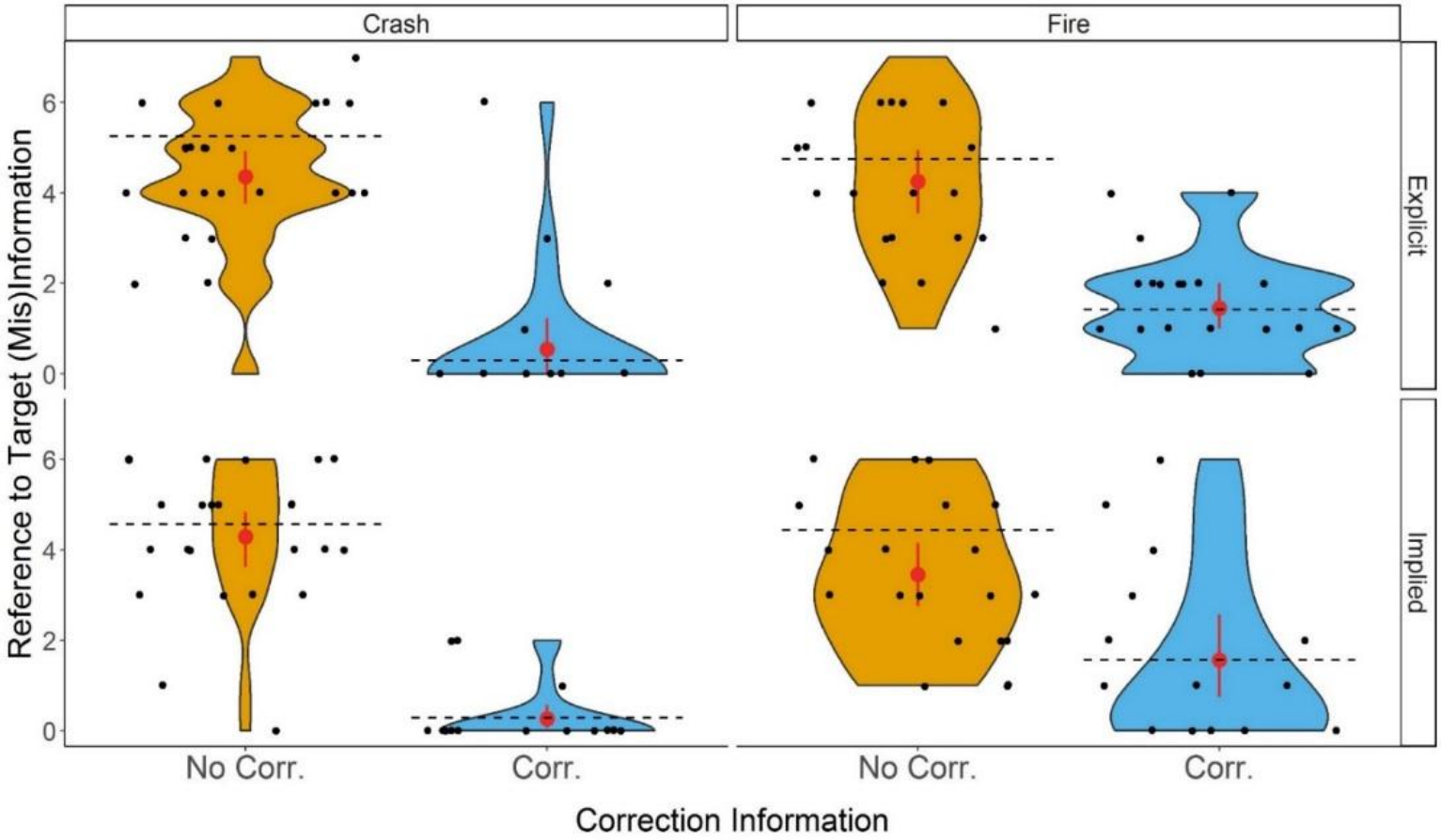

Figure 24 Distribution and probability density of inference scores as function of correction information, misinformation, and Report in Experiment 6. Red dots represent mean and lines error bars represent 95% confidence interval. Dashed lines represent means after excluding participants who answered the question about the point of the correction or control message correctly

4.5.2. Inference Scores

Figure 24 shows the distributional characteristics and means of inference scores across correction information, misinformation type, and scenario conditions. There is a clear pattern of results: exposure to a correction reduced references to target (mis)information relative to the uncorrected groups, irrespective of misinformation type. Figure 24 also shows that the distribution of misinformation references was more uniform for participants who were presented with a correction in the warehouse fire report whereas for the crash report references are skewed toward zero. Most importantly, there was no real difference in the number of misinformation references generated between the two misinformation type conditions.

A zero-inflated regression model was fit to the inference score data (see Chapter 3 for more extensive rationale for this analytic approach). Table 9 shows the results of the analysis of deviance test performed to establish the presence of interactions and main effects in the data. The analysis revealed a significant three-way interaction between correction information, misinformation type, and the scenario presented to participants.

Table 10 shows the marginal means and significance of post-hoc comparisons. The table of marginal means shows that correction information significantly reduced reference to target (mis)information for both the crash and fire reports. There was no difference between implied and explicitly stated corrected misinformation conditions. This finding held for both the warehouse fire and crash reports. The difference that appeared to drive the significant three-way interaction was between explicitly stated, corrected misinformation in the fire report, and implied corrected misinformation in the crash report. This difference was not predicted, and was not particularly informative with respect to replicating the difference between implied and explicitly stated misinformation, and is therefore not discussed further.

Table 9 Analysis of deviance test on model terms for inference scores in Experiment 6

Term	df	χ^2	p value
Correction Information	1	27.08	<.001***
Misinformation Type	1	2.62	.11
Scenario	1	0.12	.73
Correction Information*Scenario	1	0.40	.53
Correction Information*Misinformation Type	1	2.77	.10
Scenario *Misinformation Type	1	0.14	.71
Correction Information* Scenario *Misinformation Type	1	4.24	.04*

Table 10 Marginal inference score means and post-hoc comparisons Experiment 6

Correction Information	Misinformation Type	Scenario	Estimated marginal mean	Group
Corr.	Implied	Crash	0.26	a
Corr.	Explicitly Stated	Crash	0.55	ab
Corr.	Explicitly Stated	Warehouse Fire	1.45	b
Corr.	Implied	Warehouse Fire	1.56	ab
No Corr.	Implied	Warehouse Fire	3.45	c
No Corr.	Explicitly Stated	Fire	4.25	c
No Corr.	Implied	Crash	4.29	c
No Corr.	Explicitly Stated	Crash	4.36	c

Note: Confidence level used: 0.95. Confidence-level adjustment: sidak method for 8 estimates. P value adjustment: Tukey method for comparing a family of 8 estimates. Significance level used: alpha = 0.05. Comparisons that share the same letter group are not significantly different.

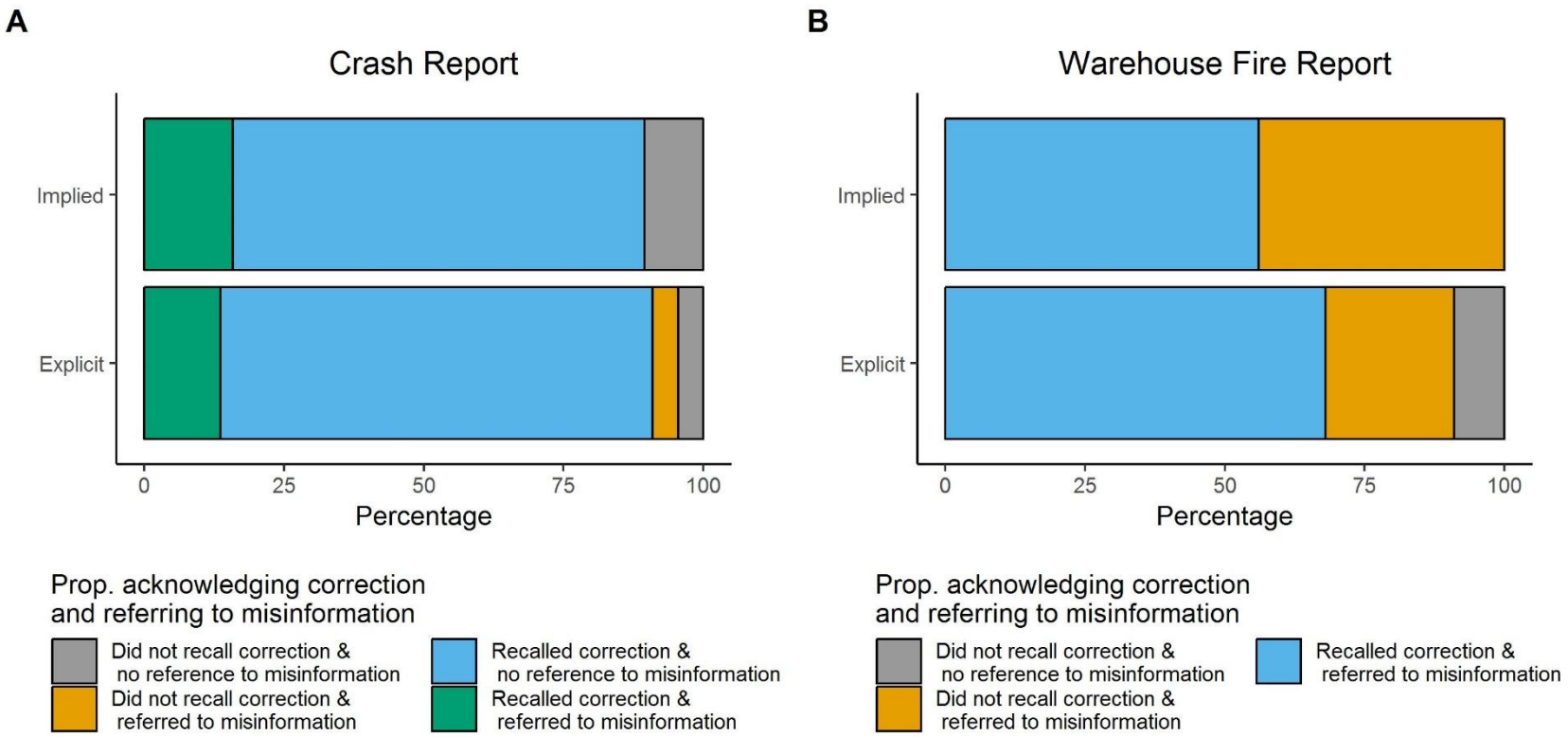
A
Crash Report
Implied
Explicit
0
25
50
75
100
Percentage
B
Warehouse Fire Report
Implied
Explicit
0
25
50
75
100
Percentage
Prop. acknowledging correction and referring to misinformation
Did not recall correction & no reference to misinformation
Did not recall correction & referred to misinformation
Recalled correction & no reference to misinformation
Recalled correction & referred to misinformation
Prop. acknowledging correction and referring to misinformation
Did not recall correction & no reference to misinformation
Did not recall correction & referred to misinformation
Recalled correction & referred to misinformation

Figure 25 Proportions of participants who recalled or did not recall the correction, and either referred or did not refer to target (mis)information, by correction information condition and scenario in Experiment 6.

Subset-analyses. Inference score analyses were also performed on two subsets of the data. The first subset only included participants who answered instruction attention check questions correctly (N = 115). This analysis was performed in order to examine the results obtained were attributable to participant inattention, as measured through their ability to correctly answer questions about the instructions. The final model obtained from this subset of participants retained the three-way interaction. Analysis of deviance tests showed that the three-way interaction was not quite significant, $p = .06$. Breakdown of the interaction through post-hoc tests similarly revealed that the differences between condition means were not meaningful with respect to replication of the original findings. The second subset included only participants who recalled the critical information (N = 100), in order to examine whether the results may be affected a failure to encode the critical information presented at Message 10 rather than general inattentiveness to the task instructions. The final model retained main effects of correction information and event report. However, the analysis of deviance tests indicated that only the main effect of correction information was significant, $p < .001$.

Correction acknowledgment. The proportion of participants who referred to misinformation even though they acknowledged that the information had been corrected for each is shown in Figure 25. The figure shows that the proportion of participants who made at least one uncontroverted reference to misinformation whilst also acknowledging that the information was corrected was substantially higher for the warehouse fire report (56-68%) than the crash report (16-14%). This was presumably because so few people referred to the misinformation in the crash report. This replicates findings reported in Chapter 3.

4.5.3. Recall Accuracy Scores

A zero-inflated regression model was fit to the recall accuracy scores to examine whether condition affected the number of details that were accurately recalled from the report. The results of the analysis of deviance test for recall scores can be found in Table 11. There was a significant effect of scenario on the number of accurate details recalled. When averaging across correction information and misinformation conditions participants recalled more accurate details from the crash (*M* = 5.52, *SD* = 1.55) and warehouse fire (*M* = 4.72, *SD* = 1.80), reports. The p-value for the three-way interaction term was significant (*p* = .046); the interaction was not broken down further[36].

Table 11 Analysis of deviance test on model terms for recall scores in Experiment 6

Term	df	χ^2	p-value
Correction Information	1	1.198	.16
Misinformation Type	1	4.76	.03*
Scenario	1	0.02	.89
Correction Information*Scenario	1	0.19	.67
Correction Information*Misinformation Type	1	0.32	.57
Scenario*Misinformation Type	1	0.03	.86
Correction Information*Scenario*Misinformation Type	1	3.97	.05*

Note: *. Effect is significant at the 0.05 level

[36] A stepwise approach comparing the AIC (Akaike Information Criterion) of each model confirmed that report was the only factor that should be included in the final model.

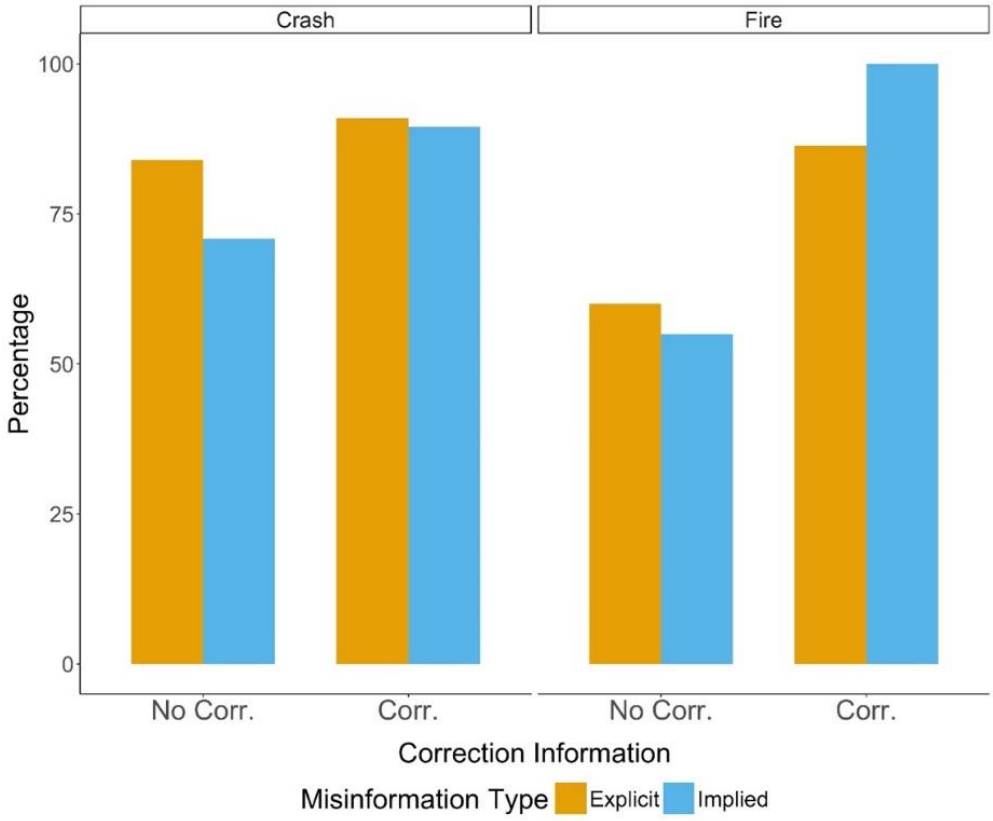

Figure 26 Proportion correctly recalling critical information (presented at Message) in Experiment 6

4.5.4. Critical Information Recall

Figure 26 shows the proportion of participants who accurately recalled critical information (either the correction or control message shown at Message 10). A chi-square test of independence on the crash report data revealed a significant association between condition and critical information recall, χ^2 (3) = 40.83, p < .001, and the warehouse fire report, χ^2 (3) = 22.47, p < .001. Participants who read the crash report were more likely to recall critical information when they received a correction (Explicitly Stated = 22.2%, Implied = 18.9%) than when misinformation was uncorrected (Explicitly Stated = 4.44%, Implied = 7.78%). The proportion of participants who read the warehouse fire report and correctly recalled the critical information was also higher when they received a correction (Explicitly Stated = 24.40%, Implied = 20.51%). The fact that participants were more likely to recall the correction

than the control critical information suggests that this information was more salient and therefore more available in memory. However, this could not be further confirmed by additional tests looking only at correction conditions as the data violated the chi-square assumptions (i.e. more than 20% of the cells had expected counts of less than 5).

4.6. Summary

Experiment 6 was designed to examine whether misinformation that implies a likely cause of an adverse outcome is more resistant to correction than explicitly stated misinformation in two scenarios. Corrections to misinformation were directly compared in a scenario where the CIE was previously eliminated (crash scenario) and one in which the CIE was present (warehouse fire scenario; see Experiment 4 and 5's results). This comparison made was in order to examine whether there was an interaction between correction information, misinformation type, and scenario.

Findings showed no evidence that implied misinformation was more resistant to correction than explicitly stated misinformation. That is, participants produced a similar number of post-correction references to implied and explicitly stated misinformation. These results are inconsistent with previous findings showing that implied misinformation is more resistant to correction than explicitly stated misinformation (Rich & Zaragoza, 2016). One possible reason Rich and Zaragoza's findings were not replicated could be the low sample size in Experiment 6 relative to prior work. Even after excluding participants who did not recall the correction there were at least 40 participants in each experimental condition whereas group sizes in the present experiment ranged from 16-25[37]. Although the experiment was not designed to be underpowered, financial constraints at the time the experiment was conducted limited the number of participants it was possible to recruit for the study. Another possible reason for the lack of replication could be that the implied versus explicitly stated manipulation was too subtle to exert any influence. Although the wording of misinformation used in the 'explicitly stated' misinformation condition was almost identical to that of prior work (i.e. "Police

[37] Group sizes were unequal due to an allocation error in Qualtrics.

suspect that …"), participants may not have interpreted this as an explicit statement of the likely cause of the outcome.

Experiment 6's results also showed further evidence that corrections almost eliminated reliance on misinformation for the crash report but not for the warehouse fire report. As noted in Chapter 3, this finding provides some possible constraints on the situations in which the CIE occurs. More specifically, the finding suggests that corrections which fully invalidate misinformation by completely ruling out its role in the outcome of the event do successfully eliminate reliance on misinformation. In contrast, corrections which leave some room for misinformation to be true result in a continuing influence of misinformation.

4.7. Experiment 7

Experiment 7 was designed address the limitations of Experiment 6 by increasing the sample size and using a different scenario in which the misinformation manipulation was more palpable (i.e. "Woman assaulted" rather than "Police suspect woman was assaulted"). It was also desirable to compare implied and explicitly stated misinformation in a scenario which had previously elicited the CIE (see Experiment 4's results).

Experiment 7 also included a novel control condition in which the correction is presented but the misinformation is not (this control condition was also included Experiments 2A and 2B reported in Chapter 2 of this thesis). This condition was included in order to further establish the extent to which misinformation is referred to simply because it has been mentioned or because it forms part of an initially constructed, coherent mental-model. Results presented in Chapter 2 showed that this control condition results in a similar number of references to misinformation as a condition in which misinformation is first presented and then corrected later. Therefore an additional aim of Experiment 7 was to replicate this with one of the new scenarios developed for this experimental programme.

The method, directional hypotheses, and analysis plan (including planned analyses, data stopping rule, and exclusion criteria) were pre-registered prior to data collection; this information can be found at: https://osf.io/ep2rs/.

Experiment 7 aimed to test the following hypotheses: A) a correction will reduce, but not eliminate, references to target (mis)information compared to no correction, B) implied misinformation will lead to a reduced effect of correction, and C) presenting a correction without initial misinformation will result in continued reliance on misinformation[38].

4.8. Method

4.8.1. Participants

[38] We also took this opportunity to include the Cognitive Reflection Test (CRT) which is a task designed to measure an individual's tendency to an incorrect but instinctive response and further reflect in order to find the correct answer (Frederick, 2005). The CRT has been correlated with judgmental biases (heuristics) (Toplak, West, & Stanovich, 2011). Analysis of the relationship between continued reliance on misinformation and cognitive reflection was exploratory. Previous CIE research has found that performance on the CIE does not correlate with personality measures reflecting the extent to which individuals are inclined to towards effortful cognitive activities (Rich, 2016). There was no evidence that CRT scores predicted continued reliance on misinformation so the results are not discussed further here.

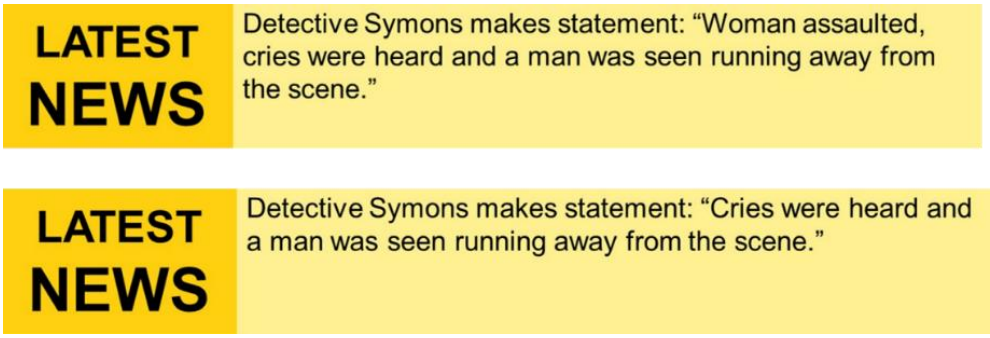

Figure 27 Explicitly stated (top) and implied (bottom) target (mis)information for the head injury scenario in Experiment 7

A power analysis using a medium effect size for the main effect of correction information suggested that a minimum of 210 participants would be required to achieve high statistical power (f = .25, α = .05, 1-β = .95). The 'stopping rule' was pre-registered (https://osf.io/ep2rs/) in order to constrain researcher degrees of freedom. The aim was over-recruit in order to replace participants that did not comply with the instructions (N=260). In total 268 U.S. based participants were recruited from Amazon Mechanical Turk. The target sample size was slightly exceeded because Amazon Mechanical Turk had been contaminated with a high number of responses from bots from or from another unidentified source, at the time of data collection. A reasonably high number of participants (N=20) were excluded prior to analysis because their responses were nonsensical, not relevant to the question (e.g. "Many countries around the world are installing closed-circuit television (CCTV) surveillance camera systems as an additional tool in fighting crime and making their streets safe. Based on many studies, the very presence of camera surveillance systems has discouraged criminals, thus preventing crimes from happening") because they repeated the information in the question, or because they exceeded the time restriction of 60 minutes. Following exclusions, there were 248 (96 female and 152 male, M_{age} = 35.58, SD = 11.09) participants. Participants were paid $1.80 and took 17 minutes on average to complete the study.

4.8.2. Stimuli, Design & Procedure

The stimuli used in Experiment 7 were almost identical to those used for the head injury scenario in Experiment 4 (see Appendix I). The only exception was that the present study included two different types of misinformation: one which implied that a woman found unconscious in the street had been attacked by a man seen running away and the other in which this information was explicitly stated. Figure 27 shows the wording and presentation format of the implied and explicitly stated 'breaking news' statements used in Experiment 7. One group of participants also received a control message at the same point others read target (mis)information. The control message was as follows: "Detective Symons makes statement: "We ask the public to stay away from the scene while we investigate." The critical (correction) information presented at Message 10 stated: Det. Symons revises earlier statement: "A man seen running away was not involved in the incident. Injuries could not have come from physical assault", instead of referring to the 'the man'. The wording was changed in order to avoid participants thinking they had missed earlier information.

The effect of correction information (No Correction, Correction), and Misinformation Type (Explicitly Stated, Implied) on reference to target (mis)information was assessed between groups (k = 4). The present study also included an additional control condition in which misinformation was only mentioned as it was being corrected. Participants were therefore randomly allocated to one of 5 experimental conditions[39]: 1) Implied misinformation + No Correction, 2) Explicitly Stated misinformation + No Correction, 3) Implied misinformation + Correction, 4) Explicitly Stated misinformation + Correction, 5) No misinformation + Correction. The two 'no correction' conditions served as an empirical baseline to establish the effectiveness of a correction. The 'no misinformation + correction' control condition made it possible to establish whether some of the CIE can be explained by the availability of the causal explanation posited by the misinformation.

[39] N = 49-52 in each group

The experimental procedure was much the same as Experiment 6: participants read the report, answered intermixed inference and factual recall questions, and then answered critical information question. There were two small differences to the procedure in Experiment 6. First, the instructional attention check included at the beginning of the study required participants to answer all three questions correctly before proceeding to the study (as suggested by Crump et al. 2013). Second, the modified recognition test was not included after answering the questionnaire. Instead, participants completed the Cognitive Reflection Test (CRT) after they answered the critical information recall questions.

4.9. Results

4.9.1. Coding of Responses

The criteria for coding responses were identical to those used in Chapter 3 (and can also be found in Appendix H), and so are not fully reiterated here. Table 12 shows examples of responses that were coded as a reference to target (mis)information and an example of a response that was not coded as such. The maximum individual score for inference questions was 7. Responses to factual questions were scored for accuracy; correct or partially correct responses were scored 1 and incorrect responses were scored 0; the maximum factual score was 7. Critical information recall scores were computed using the same criteria; the maximum individual critical information recall score was 2[40].

Table 12 Example of inference response coding in Experiment 7

Inference Question	Example of Response Scored 1	Example of Response Scored 0

[40] Due to time and resource constraints it was not possible to perform inter-coder reliability analysis for Experiment 7. This will be done at a later stage.

What do you think is a likely explanation for what happened to the injured woman?	A person followed her after she left her assignment, followed her to this location and then, thinking that he was in a secluded location, hit her on the head from behind. Perhaps he stole her handbag or committed another form of assault before leaving, but from the screams and the attention of the locals he probably fled.	It sounded to me like she was in some sort accident possibly a motor vehicle one.

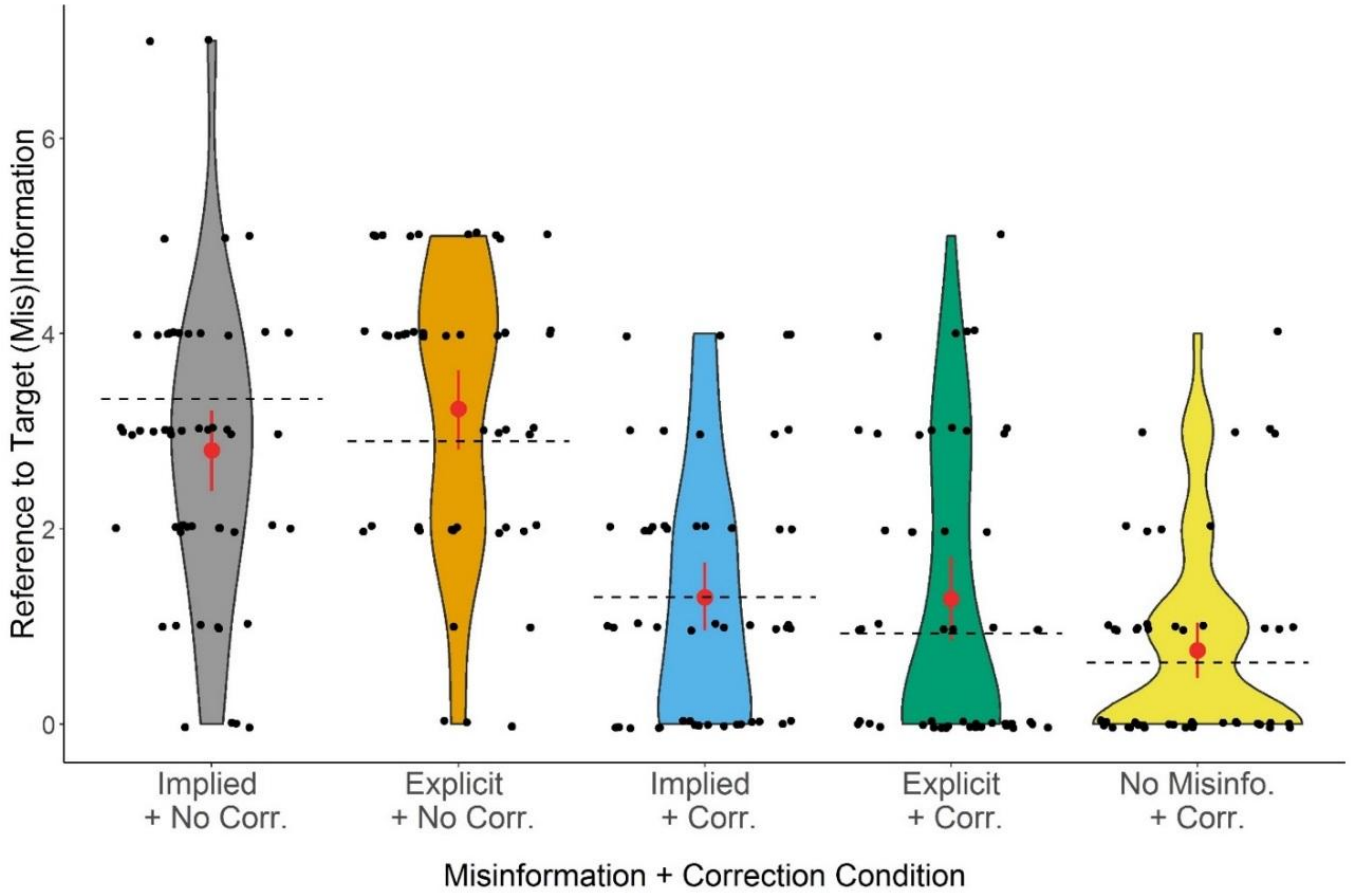

Figure 28 Distribution and probability density of inference scores as function of correction information and misinformation type in Experiment 7. Red dots represent mean and lines error bars represent 95% confidence interval. Dashed lines represent means after excluding participants who answered the question about the point of the correction or control message correctly

4.9.2. Inference Scores

Figure 28 shows the distributional characteristics of inference scores across misinformation and correction information conditions. The mean number of references to target (mis)information are also shown here. The pattern of results is consistent with previous findings reported in this chapter: the number of post-correction target (mis)information references produced were equivalent for both implied and explicitly stated misinformation. The number of references to implied and explicitly stated target (mis)information produced in the uncorrected conditions were roughly similar. Figure 28 also shows that the number of references to target (mis)information were reduced – but not eliminated completely - for the groups who received a correction. Interestingly, participants for whom misinformation was only mentioned during the correction still referred to the misinformation despite the fact that it was only mentioned in the correction.

A zero-inflated regression model was fit to inference scores with misinformation and correction information group as the predictor. An analysis of deviance test on inference scores showed that misinformation and correction information condition significantly predicted the number of references to misinformation produced, χ^2 (4) = 30.93, p < .001. Table 13 shows the marginal means and post-hoc comparisons with tukey adjustment for multiple comparisons. The table shows that the corrections to both implied and explicitly stated misinformation significantly reduced the number of references to target (mis)information compared to uncorrected misinformation. The differences between implied and explicitly stated misinformation did not differ between uncorrected or corrected conditions. The condition featuring no initial misinformation and a correction significantly differed from the uncorrected conditions but not from corrected conditions.

Table 13 Marginal inference score means and post-hoc comparisons Experiment 7

Misinformation Type	Correction Information	Estimated marginal means	Group
Implied	No Correction	2.81	b
Explicitly Stated	No Correction	3.23	b
Implied	Correction	1.30	a
Explicitly Stated	Correction	1.29	a
No Misinformation	Correction	0.76	a

Note: Confidence level used: 0.95. Confidence-level adjustment: sidak method for 8 estimates. P value adjustment: Tukey method for comparing a family of 8 estimates significance level used: alpha = 0.05. Comparisons that share the same letter group are not significantly different.

Question analysis. The proportion of references to target (mis)information was examined separately for each inference question and condition. Figure 29 shows that participants were most likely to refer to the misinformation in response to the questions that asked who, if anyone, was responsible for the woman's injuries and what the most likely explanation for the woman's injuries was. There did not appear to be any tangible pattern of responding with respect to type of misinformation presented (i.e. implied or explicitly stated). Interestingly, participants were more likely to refer to misinformation when asked what a likely explanation for the women's injuries was than they were when asked what the most likely cause of the woman's injuries was.

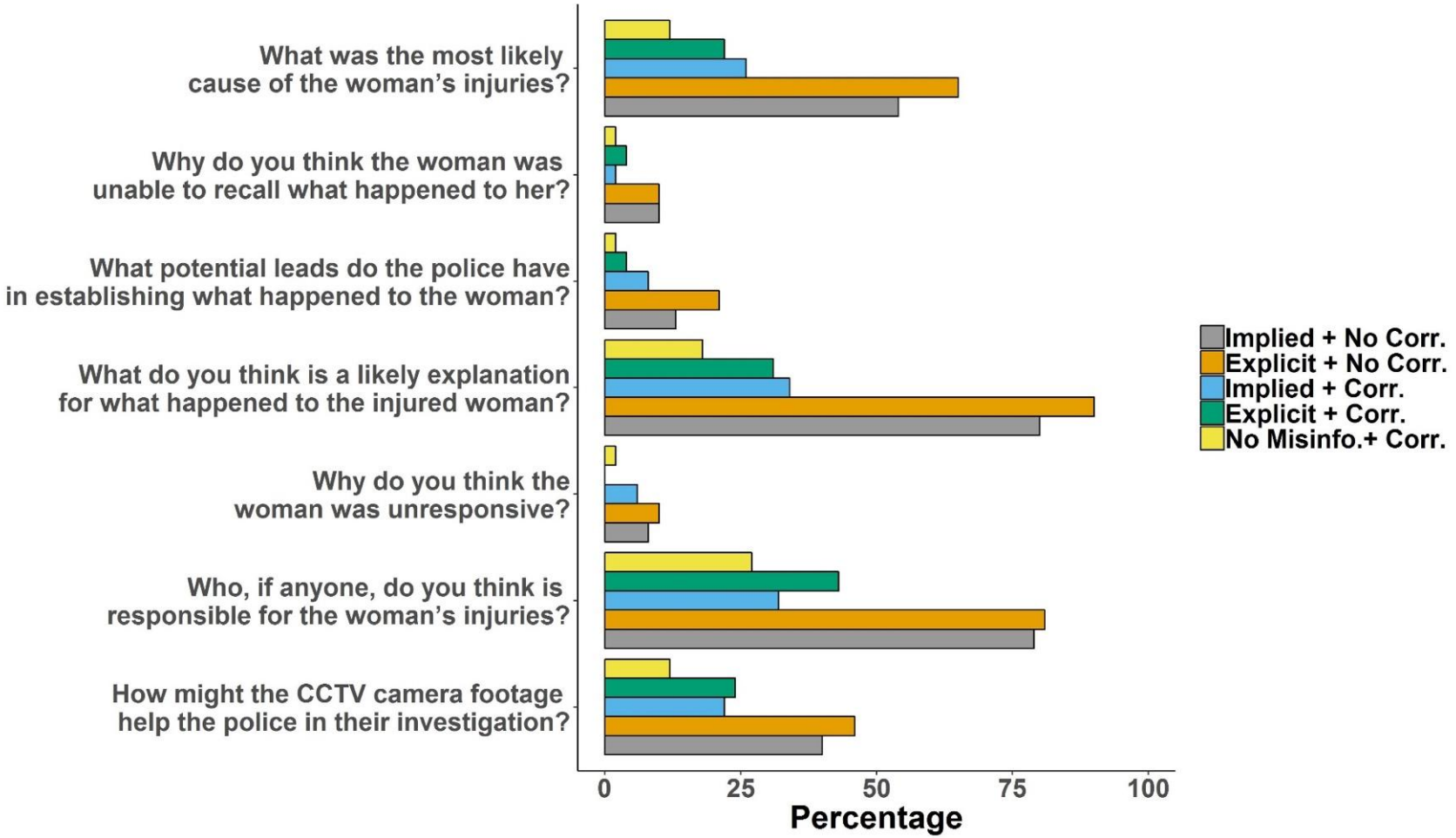

Figure 29 Proportion of references to target (mis)information by inference question and condition in Experiment 7

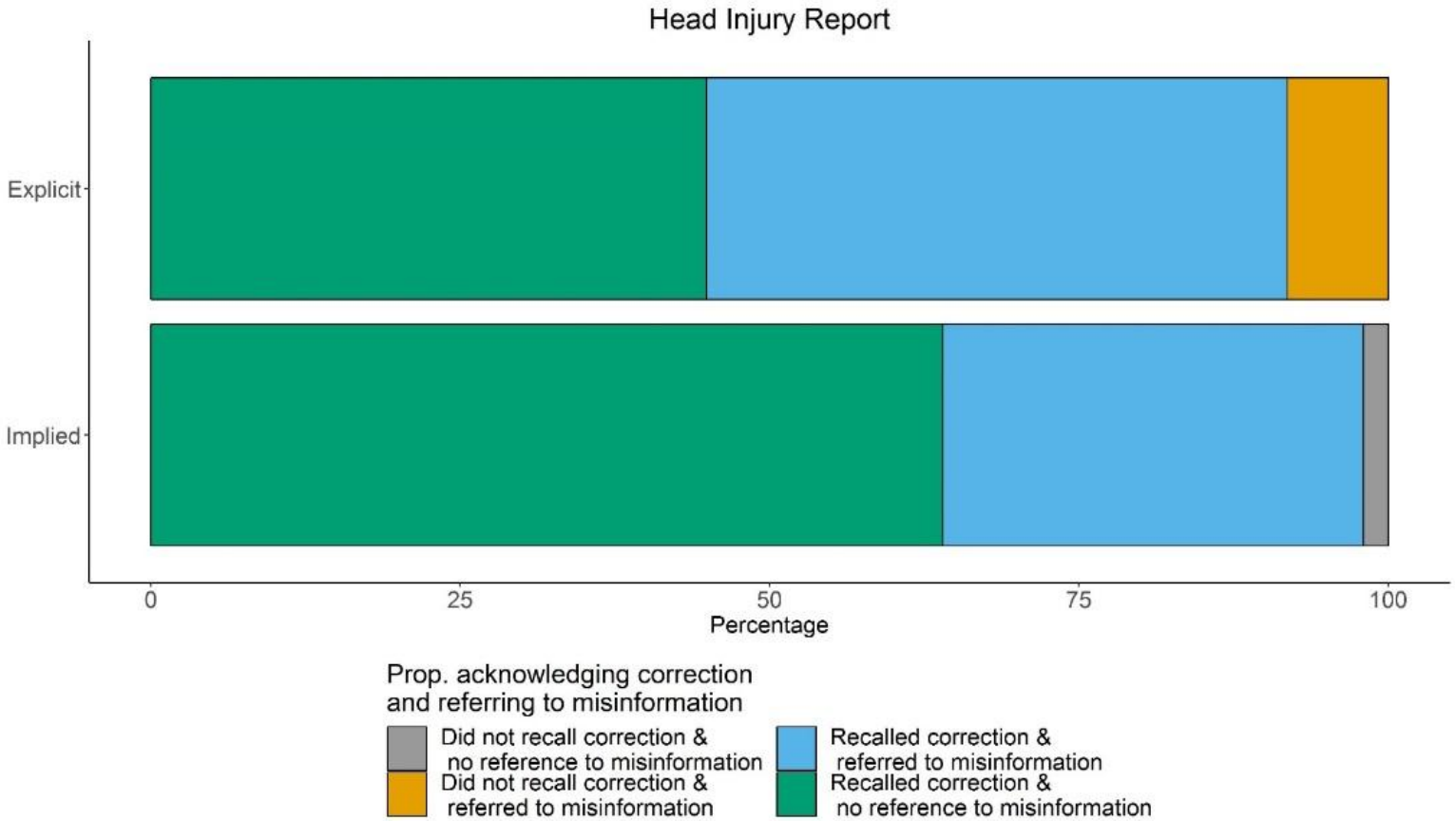

Figure 30 Proportions of participants from the correction groups who recalled or did not recall the correction, and either referred or did not refer to target (mis)information, by misinformation type in Experiment 7.

Correction acknowledgment. The proportion of participants who referred to misinformation even though they acknowledged that the information had been corrected is shown for each condition in Figure 30. The figure shows that the proportion of participants who made at least one reference to misinformation whilst also acknowledging that misinformation was corrected was higher for the group who received a correction to implied misinformation (64%) than for explicitly stated misinformation (45%). This somewhat contradicts the inference scores which show no difference between the conditions in terms of the average number of misinformation references produced. One participant who acknowledged the correction and referred to the misinformation in response to the inference questions wrote “Even though there was a man deemed not responsible that ran away from the incident, I still think he was the one responsible”. The same participant later wrote that “He stated that the man that was seen running away from the incident was not involved” in response to question asking what the implication of the second report from Detective Symons was. It is not entirely clear why participants who received explicitly stated misinformation were more likely to refer to misinformation whilst also acknowledging that it was corrected***.*** This could be due to the salience, and therefore availability of this information in memory, when answering inference questions.

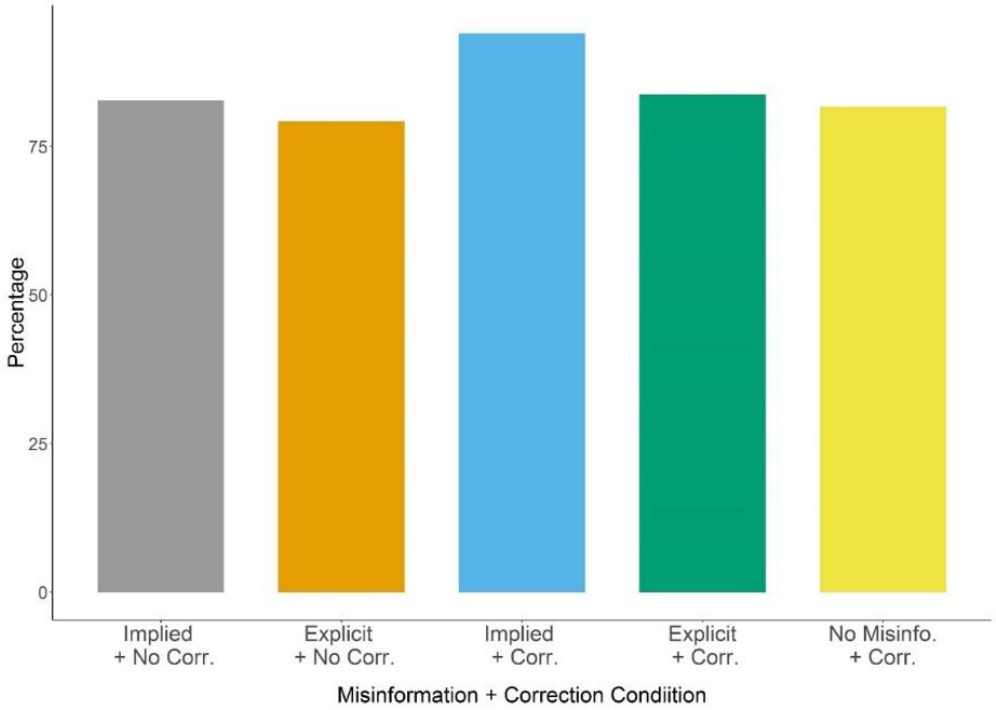

Figure 31 Proportion correctly recalling critical information (presented at Message 10) in Experiment 7.

4.9.3. Recall Accuracy Scores

The Poisson regression model fit to recall accuracy scores indicated combination of misinformation and correction information did not predict the number of details that were accurately recalled from the report, χ^2 (4) = 0.90, p = .92.

4.9.4. Critical Information Recall

A chi-square test of independence on these data revealed a significant association between condition and critical information recall, χ^2 (4) = 114.53, p < .001. This was primarily due to the low proportion of the no correction group (Implied = 17.3%, Explicitly Stated = 20.8%) who accurately recalled the critical information (i.e. that police had cordoned off the area and were appealing for witnesses). In contrast over half of the participants in the groups featuring a correction recalled the critical correction information: implied +

correction (94%), explicit + correction (83.6%) and no misinformation + correction (81.6%). The difference in conditions may be explained by the difference in the salience of the critical information (i.e. correction of misinformation vs. inconsequential update from police). The test was not significant when looking just at the conditions featuring a correction, χ^2 (2) = 3.73, p = .15.

4.10. Summary

Experiment 7 was designed to test the effectiveness of corrections to misinformation that implies or explicitly states the likely cause of an adverse outcome. Experiment 7 differed from Experiment 6 in that it tested the effectiveness of corrections in a different scenario in which explicitly stated misinformation was more obvious (i.e. “Woman assaulted, man seen running away” vs. “Man seen running away”). The number of participants recruited in Experiment 7 was also doubled to increase statistical power, and there was also a more balanced allocation to groups. An additional control condition in which misinformation was not initially presented and was only referred to whilst being corrected, was also included.

Experiment 7’s results provide further evidence that misinformation implying or explicitly stating a likely cause of an adverse outcome are equally resistant to correction. The number of post-correction references to target (mis)information did not significantly differ between the implied and explicitly stated misinformation conditions. The uncorrected misinformation conditions also showed no difference between implied and explicitly stated misinformation. Interestingly, the condition in which misinformation was only presented as part of the correction was not significantly lower than the two conditions featuring a correction. This may suggest that a large part of the CIE occurs because of the availability of the causal explanation offered by misinformation rather than a preference for a coherent mental model. This result is also consistent with Experiment 2A and 2B’s results showing that presenting misinformation only as part of the correction did not significantly reduce reference to misinformation from the conditions featuring both misinformation and its correction.

4.11. Conclusion

The two experiments reported in this chapter were designed to examine whether misinformation which implies a likely cause of an adverse outcome is more resistant to correction than misinformation which explicitly states a likely cause. An additional consideration was the generalisability of this effect to different scenarios. One limitation of previous CIE studies is that they typically employ one scenario to test effects of interest. One of the scenarios used in the present study was a modified version of the warehouse fire scenario used in previous research (Guillory & Geraci, 2010; Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988). The other two scenarios (crash and head injury) were newly developed for the purposes of this experimental programme.

The results of experiments reported in this chapter provide evidence that implied and explicitly stated misinformation are equally resistant to correction when a CIE is present. There was no difference in the number of misinformation references produced following a correction to implied or explicitly stated misinformation in three different scenarios. These findings are inconsistent with previous studies showing that implied misinformation is more resistant to correction than explicitly stated misinformation (Rich & Zaragoza, 2016).

Although Experiment 6 found a significant interaction between correction information, misinformation type, and scenario, the interaction was driven by an uninformative difference between conditions; specifically, the number of references to corrected implied and explicitly stated target (mis)information significantly differed between the warehouse fire and crash reports. Participants in conditions featuring a correction referred to target (mis)information significantly less than uncorrected groups, regardless of whether misinformation was implied or explicitly stated, or the scenario that participants read. Experiment 7's results similarly showed that the number of post-correction target (mis)information references produced did not differ between implied and explicitly stated conditions. Uncorrected implied and

explicitly misinformation conditions, also produced a similar number of references to target (mis)information.

One potential explanation for the lack of replication in Experiment 6 was eliminated by increasing the sample size in Experiment 7. The lack of replication in Experiment 6 did not appear to be the result of the subtleness of the explicitly stated misinformation either. Explicitly stated misinformation in Experiment 7 made the report of the woman being assaulted abundantly clear.

Another possible explanation for the lack of replication could be due to the fact that the scenarios used in Experiments 6 and 7 (and in Chapter 3) were specifically designed such that additional event information, that was presented either side of misinformation and of the correction, was not congruent with the misinformation. Previous studies that have used the warehouse fire scenario have included information that is congruent with the causal explanation offered by misinformation. For example, the warehouse fire scenario includes reports that that there were 'explosions' and that 'firefighters had inhaled toxic fumes', which can be interpreted as effects of burning gas cylinders and oil paint. This information could therefore be interpreted as evidence that the misinformation is in fact true, and that the correction is in fact false. The jewellery theft scenario used in Rich and Zaragoza (2016) included additional information that was congruent with the misinformation explanation. For instance, participants were told that items such as the TV were not missing from the house. Knowing that these items were not stolen increases the possibility that the culprit was only interested in jewellery. This in turn makes it more likely that the person who stole the jewellery knew where it was located.

The reports used in Experiments 6 and 7 only included one piece of information that could be interpreted as congruent with the explanation offered by misinformation. Namely, that 'thick and oily smoke had hindered firefighters' efforts', that the driver of the van that crashed was 'a recent divorcee who had a prolonged legal battle with his ex-wife', or that 'An initial medical report suggests head injuries are consistent with impact from a blow to the head'. The inclusion of these 'additional information' statements may

only slightly increase the possibility that the misinformation caused the outcome. One reason that there was no difference between implied and explicitly stated misinformation conditions could be that the stimuli used did not allow participants to construct a particularly elaborate mental model of the event. If participants did not construct particularly elaborate mental event models in the first place then whether misinformation implied or explicitly stated a likely cause of the fire might not make much of a difference to generating misinformation consistent causal inferences.

After considering the different possibilities it is still not entirely clear why the difference between implied and explicitly stated misinformation was not replicated. In order to further investigate this, it would be necessary to test different scenarios that vary in terms of their structure. Perhaps one way of doing this would be to experimentally manipulate whether the report includes additional information that is congruent, incongruent, or neutral with respect to the explanation offered by misinformation. Experiments 6 and 7's results may suggest that story structure, rather than story content per se, interact with the effectiveness of corrections to implied or explicitly stated misinformation. However, to verify this idea it would need to be tested experimentally.

In addition to the main findings, Experiment 6's results also showed further evidence that the CIE was substantially attenuated in the crash report compared to the warehouse fire and head injury reports. More than half of the participants who read the warehouse fire (56-68%) and head injury (45-64%) reports referred to corrected misinformation at least whereas only a small proportion of participants who read the crash report did so (14-16%). These results are consistent with the experimental results reported in Chapter 3 of this thesis. One possible explanation for this difference could be related to the degree of inference required for a given scenario, which was not experimentally manipulated here. As has been noted previously, the correction used in some scenarios (e.g. warehouse fire, head injury) leaves open the possibility that the explanation offered by misinformation might be in some way true, whereas the correction in the crash scenario rules out the explanation offered by misinformation. The latter case reduces the degree of

inference required to understand what has happened whereas the former cases require a higher degree of inference.

To conclude, the present set of experiments showed no evidence that corrections to explicitly stated misinformation are more effective than corrections to implied misinformation. The fact that this result was not replicated across three different scenarios suggests that is not due to the scenarios used. Future studies should also measure participants' recall of target (mis)information to establish how well this information is preserved in memory as participants may not have been able to recall the details that distinguish misinformation conditions. Despite these limitations, the present study's results highlight the importance of testing the boundary conditions of the CIE by replicating previous findings with different scenarios. Furthermore, these results provide further evidence that the CIE is not guaranteed to emerge under all circumstances which provides some possible constraints on the generalisability of continued influence findings.

5 General Discussion

5.1. Thesis Aims

Existing research on the continued influence effect (CIE) of misinformation has established that the CIE can be observed under some circumstances, but not whether it would always be expected. A review of CIE studies suggests that methodological factors may affect the extent to which a correction reduces reliance on misinformation independently of experimental manipulations. In particular, the limited number of scenarios used in CIE studies, variability in sample sizes, and the student population typically used in CIE studies, could have an impact on the validity and reliability findings obtained using the CIE approach. The latter can pose significant challenges to understanding the prevalence of the CIE as a phenomenon in the world. This thesis' primary aim was to advance understanding of the continued influence effect, and the conditions under which it occurs, and to overcome existing methodological problems of validity and reliability allowing for more effective testing of the CIE in the future. This was achieved through a series of methodological steps. The first step was to develop and validate a methodology that allows for web-based testing of the CIE to facilitate and streamline data collection from larger and more representative samples. The second step was to use this methodology to examine the prevalence of the CIE across different scenarios, in more realistic settings (i.e. report of a potential assault or motor accident), and with larger samples. This methodology was then used to investigate the robustness of two claims from the CIE literature, and also whether continued reliance on misinformation extends to a novel control condition in which misinformation is only mentioned in the correction.

The first claim from the CIE literature examined was that corrections which explain how misinformation occurred (e.g. from unintentional error or intentional lie) can help people understand the contradiction between misinformation and the correction, and reduce continued reliance on misinformation (as suggested by Bush, Johnson, & Seifert, 1994; Johnson & Seifert, 1994; Seifert, 2002). Corrections that explain where misinformation originated (and is therefore erroneous) can be likened to real-world situations in which jurors are asked to ignore inadmissible information because it is unreliable, a scientific article is retracted because data has been fabricated, or information in the news is corrected because initial information was relayed in error. The second claim investigated was that misinformation, which implies a likely cause of an adverse outcome, is more resistant to correction than misinformation explicitly stating a likely cause (Rich & Zaragoza, 2016). Misinformation that implies a likely cause of an adverse outcome may be likened to situations in which the media or politicians insinuate or use innuendo based on false or inaccurate information. The robustness of these claims was examined to establish the circumstances under which the CIE is likely to occur.

5.2. Chapter Summaries

5.2.1. Chapter 1

Chapter 1 reviewed existing CIE research, showing that people often continue to rely on misinformation despite clear and credible corrections (Lewandowsky et al., 2012; Seifert, 2002; 2014). The review showed that corrections usually halve the number of times people refer to misinformation in comparison to a situation in which misinformation remains uncorrected (e.g. Ecker et al., 2010; Ecker et al., 2011b). However, corrections rarely eliminate reliance on misinformation completely when using the CIE approach. This basic finding has been replicated for different scenarios (e.g. warehouse fire, plane crash, jewellery theft), types of correction (Bush, Johnson, & Seifert, 1994; Guillory & Geraci, 2013), and of misinformation (Ecker, Lewandowsky, & Apai, 2011; Rich & Zaragoza, 2016). The CIE has been explained either as

a failure to appropriately update a mental-model of an event (e.g. Johnson & Seifert, 1994), or as a failure to engage strategic memory processes during retrieval of the correction (e.g. Ecker et al., 2011b). The precise mechanism by which the CIE occurs remains unresolved in the literature.

Several factors have been shown to moderate the CIE. For instance, corrections are more effective when they provide an alternative causal explanation (Ecker et al., 2010; Ecker et al., 2011a; Johnson & Seifert, 1994; Rich & Zaragoza, 2016), are preceded by pre-exposure warnings about the persistence of misinformation (Ecker et al., 2010), originate from a highly trustworthy source (Guillory & Geraci, 2013), and address the conversational implications of contradictory statements (Bush et al., 1994). Although these factors have been shown to substantially reduce the CIE, even combination of strategies (e.g. pre-exposure warnings and alternative explanation) has failed to eliminate continued reliance on misinformation completely (Ecker et al., 2010).

Despite consensus on the basic finding, there has been considerable variability in the magnitude of the reduction in reliance on misinformation produced by a correction (see Table 1 in Chapter 1). This suggests that there are factors that affect the magnitude of the CIE over and above the specific variables being experimentally manipulated. This has implications for the validity reliability, and generalisability of findings obtained via the CIE methodology.

One factor that potentially moderates the presence and magnitude of CIE the scenario used. Studies on the CIE have focused on a limited number of scenarios and effects are rarely examined across different scenarios. It is not clear whether these scenarios are representative of the variety of situations in which correction misinformation might be encountered, or whether they are primarily used because they reliably produce the CIE. This raises questions about the representativeness and ecological validity of the scenarios used in CIE studies. By extension, the limited number of scenarios may be indicative a ‘file drawer’ problem; some scenarios just do not elicit the CIE and are therefore never published (Ioannidis et al., 2014; Rosenthal,

1979; Spellman, 2012). Such a 'file drawer' problem could lead to a misrepresentation of the set of circumstances that bring about the CIE as well as the mechanism/s by which it occurs.

Another factor that could affect the validity and generalisability of results is that CIE studies have mainly been conducted in the lab with university students. University students are not representative of the general adult population as their demographics are inherently biased with respect to age, experience, intellectual ability, ethnicity, and socioeconomic status. Focusing attention on population with narrow demographics may undermine the generalisability of results to more diverse populations but also lead to misestimating the prevalence of the CIE as a phenomenon in the world. More pointedly, recruiting through university subject pools often entails small samples sizes which can result in low statistical power. This may be particularly the case for the independent group experimental designs that are often necessary in CIE studies. Low power can result in reduced chance of detecting a true effect, but also reduces the likelihood that a statistically significant result is a true effect (Button et al., 2013; Ioannidis, 2005; Ioannidis, Munafo, Fusar-Poli, Nosek, & David, 2014). Accordingly, some of the variability observed in CIE findings (i.e. correction either reduces but does not eliminate or does not reduce reliance on misinformation) could be explained by the low sample sizes in some studies.

5.2.2. Chapter 2

Chapter 2 reported four experiments examining the feasibility of collecting data on the CIE online and comparing open-ended and closed-ended inference measures. To compare online results to those obtained a lab-based experiment, experimental manipulations were tested using the exact same warehouse fire scenario used in Johnson and Seifert (1994; Exp 3A). Experiments 1A and 1B also employed a common CIE experimental design (Ecker et al., 2010; Ecker et al., 2011a; Johnson & Seifert, 1994; Rich & Zaragoza, 2016), in which the number of references to target (mis)information produced when misinformation was corrected, corrected and

an alternative explanation provided, or remained uncorrected, were compared. Experiments 2A and 2B used a similar design but substituted the alternative explanation condition for a novel control condition in which misinformation was only mentioned as part of the correction.

Findings of all four experiments showed that both open and closed elicitation procedures resulted in a clear CIE. Participants continued to refer to the misinformation despite a clear correction even though a majority recalled the correction. These findings provide clear evidence that both open-ended and closed-ended questions can be used in online experiments. Overall, there was limited evidence for an effect of correction for open-ended questions but substantial evidence for an effect of a correction using closed-ended question. The effect of a correction on misinformation references was relatively small for open-ended questions when using the warehouse fire scenario. Effects could therefore be hard to consistently detect using small sample sizes, which may explain variability in the CIE findings. It may also be the case that response format can exaggerate or minimise the difference between corrected and uncorrected groups. These findings provide further evidence for the variability of a correction's effectiveness across studies and emphasise the influence sample size on detection of effects. They also highlight the need for a more systematic investigation of the CIE and suggest different measures can affect the strength of a correction in reducing the CIE.

Experiments 2A and 2B's findings showed that participants continued to use misinformation to answer inference questions, whether the misinformation was only mentioned in the correction, or was presented early in a series of statements and corrected later. Theoretically, this suggests that the CIE may not (always) arise from participants' reluctance to part with an existing mental model without an alternative explanation (Ecker et al., 2010; Ecker et al., 2011a; Johnson & Seifert, 1994; Rich & Zaragoza, 2016). Rather, this finding suggests that participants search their memory for possible causes when asked inferential questions but fail to retrieve the information correcting the misinformation or disregard it. More generally, these findings have implications for web-based cognitive psychology experiments using

open-ended questions to elicit responses. Participants recruited via Amazon Mechanical Turk provided high-quality data in response to open-ended questions.

5.2.3. Chapter 3

The second set of experiments, reported in Chapter 3, were designed to examine whether corrections that explained why misinformation occurred more successfully reduced the CIE than corrections which negate misinformation (as found by Bush et al., 1994). Johnson and Seifert (1994; see also Seifert, 2002) proposed that corrections which explain where misinformation originated from (i.e. honest mistake, deceit) – and therefore address the conversation implications of the contradiction - might be more effective than corrections which negate misinformation and only address literal implications. Accordingly, two types of explanation were examined in Chapter 3: one in which the correction explained that misinformation occurred because of a lie and one in which misinformation occurred because of an error. There was a tentative prediction that due to our adroit detection of and general disbelief in lies that a correction which explained misinformation as a lie would be more effective than the error counterpart at reducing misinformation reliance. Experiment 3 tested the effect of explanatory corrections in a modified version of the warehouse fire scenario, in which additional information either side of the misinformation was not biased in favour the causal explanation offered by the misinformation. Experiments 4 and 5 moved to examining the effect explanatory corrections more broadly across different scenarios that upheld the same underlying scenario structure.

Chapter 3's findings showed that explanatory corrections did not reduce the number of misinformation references below a non-explanatory correction. Explanatory corrections were therefore no more effective than a negation of earlier misinformation. There was also no evidence that a correction which explained misinformation as a lie was any more effective at reducing the CIE than a correction that explained misinformation as an error.

All three types of correction resulted in comparable levels of continued reliance on misinformation - and reduced, but did not eliminate, the CIE.

These findings are inconsistent with previous work showing that corrections that account for conversational implications of the contradiction were more effective at reducing use of misinformation than corrections that account only for logic (Bush et al., 1994). The use of novel and carefully constructed scenarios in Experiments 4 and 5 provided further evidence that explanatory and non-explanatory corrections did not differ. The difference between past and present findings could be due to the scenario structure, low sample size, or possibly the ambiguousness of the correction used in prior work.

Another possible explanation could be that explanatory corrections which describe the origins of misinformation as an honest mistake or a lie are distrusted because these explanations generate further pragmatic inferences about why the misinformation was introduced. In such an example, a participant might ask themselves why no one checked to see what the person said was accurate or genuine. This might limit the effectiveness of an explanatory correction to the level of a negation. Experiments 4 and 5 also provided evidence that the CIE occurs for some scenarios but not for others. The CIE was eliminated in the crash scenario, in which the correction made it explicitly clear the likely cause offered by misinformation could not have brought about the outcome (i.e. a test showed the driver had not been drinking).

5.2.4. Chapter 4

The final two experiments, reported in Chapter 4, were designed to examine the claim that misinformation, which implies a likely cause of an adverse outcome, is more resistant to correction than explicitly stated misinformation (as demonstrated by Rich and Zaragoza, 2016). These experiments addressed limitations of prior work by examining the effect of implied versus explicitly stated causal misinformation, across three scenarios that were developed for this programme of research.

Findings showed no evidence that corrections to implied misinformation were less effective than those that corrected explicitly stated misinformation. This was true of all the three scenarios examined in Chapter 4. These findings are inconsistent with previous studies showing that implied misinformation is more resistant to correction than explicitly stated misinformation (Rich & Zaragoza, 2016). One potential explanation for the lack of replication in Experiment 6 was lack of power to detect an effect due to small sample size. This explanation in was ruled out in Experiment 7 wherein the number of participants assigned to each condition was doubled. Structural differences in the scenarios used in previous and the present work may also explain why the difference between implied and explicitly stated misinformation did not replicate. However, additional experiments would be required to confirm this empirically.

Experiment 6 also showed further evidence that the CIE was substantially attenuated for the crash scenario. One possible explanation for why the CIE occurs in some scenarios but not others could be related to the degree of inference required for a given scenario, which was not experimentally manipulated in this thesis. In addition to this, Experiment 7's results provided further evidence that the CIE is partly driven by the availability of the causal explanation offered by misinformation. Participants for whom misinformation was only mentioned during its correction referred to misinformation almost as often as a group who received misinformation first and the correction later. This is an important finding as it suggests that the CIE may in part be driven by the availability of the causal explanation offered by misinformation, rather than a mental-model updating failure.

5.2.5. Conclusions

I proposed that the specific scenario in which misinformation appears, and variability in sample sizes across studies, and restricted demographic, moderate whether the CIE occurs, and the strength of the correction in reducing the CIE. My findings show that the CIE is by no means guaranteed to arise under all circumstances. I have also tentatively argued that the

presence of the CIE could depend on the degree of inference required in each scenario. More precisely, that the CIE is substantially attenuated when the correction provides evidence ruling out the explanation offered by misinformation. Providing evidence that rules out the explanation offered by misinformation reduces the amount of inference required to resolve the contradiction between misinformation and its correction. Corrections that rule out the explanation offered by misinformation are likely to interact with other elements of the presented scenario. I have argued that the CIE may be partially explained by the availability of the causal explanation when answering inference questions rather than a reluctance to give up ones mental-model without an alternative explanation. Finally, I conclude that the conditions that give rise to the CIE are ill-specified, and that focussing on a limited number of scenarios, that are unrepresentative of the types of situations in which misinformation naturally occurs, limits the validity and reliability of CIE findings.

5.3. Limitations and Future Research

Given the pervasiveness and potential impact of misinformation in society, and the potential for CIE research to be used in policy recommendations, further investigation using different variants of the CIE paradigm is crucial for understanding reasoning and judgements about misinformation outside of the lab (or online testing environment). There were three main findings from my research. First, I showed that corrections which explain how misinformation has occurred (i.e. from deceit or an honest mistake) reduce continued reliance on misinformation. Second, I provided evidence against the claim that misinformation which implies a likely cause of an adverse outcome is more impervious to correction than explicitly stated causal misinformation. Third, I found that the CIE does not occur in situations in which misinformation is sufficiently invalidated. Finally, I showed that the CIE may be partially driven by the availability of the causal explanation offered by misinformation. These findings show that the CIE is not guaranteed to occur and raises important questions about the specific set of circumstances that bring about the CIE. Misinformation, and corrections, can manifest in a

variety of ways and the types of situations misinformation can appear in vary widely. My research highlights the need for a more systematic approach to investigation of the CIE and shows that there is much more that can be done in this field of research.

Another area that warrants further research is the different types of explanation for how misinformation occurred and why it is erroneous. My research has focused on two types of explanations for how misinformation originated: from dishonesty or an honest mistake. There are many ways that misinformation can be disseminated as well as explanations for how it occurred that rely on the intentionality distinction. For instance, misinformation can be initially disseminated accidentally – such as from fact checking errors in breaking news stories. Misinformation can be disseminated be more deliberately as a means of persuading people of a viewpoint. For instance, misinformation can originate from a misapplication of statistics, such as occurred in the Vote Leave's EU referendum campaign. Future research should examine whether explanatory corrections that appeal to accidental versus deliberate reasons for dissemination of misinformation are always equally effective.

There may be further differences in the types of explanation people find acceptable or 'good' explanations for where, how, or why misinformation occurred. For instance, research on evaluating explanations has shown that simpler explanations are often judged as better and more likely to be true, but that more complex explanations are preferred if they are more probable than simpler explanations (Lagnado, 1994; Lombrozo, 2007). Furthermore, Hilton (1990) has also argued that explanations are constrained by the rules of conversation (cf. Grice, 1975), and therefore must identify the factor that makes the difference between the target event (misinformation) and a counterfactual contrast case. There are several factors that could therefore moderate acceptance of an explanatory correction to misinformation that are worthy of further investigation.

Future research should also focus on whether the correction provides concrete evidence to discredit misinformation and render it irrelevant. My

research provides preliminary evidence that corrections which provide evidence to counter the causal explanation offered by the misinformation (e.g. by providing a physical test showing that the crash was not caused by the driver being drunk), substantially attenuate the CIE. Previous work looking at correcting inaccurate information has shown that detailed refutations (e.g. by highlighting misrepresentation of scientific findings, and so on) affect the likelihood of a correction being accepted (Swire et al., 2017). To date, research on the CIE has not investigated the extent to which providing evidence to refute the hypothesis posited by the misinformation moderates the effectiveness of a correction. For instance, empirical studies on legal decision-making suggest that people value direct evidence – such as eyewitness testimony and confessions – over circumstantial evidence like DNA evidence or fingerprints (see Heller, 2006 for review of studies on how mock jurors evaluate direct and circumstantial evidence). Investigating the ways in which different types of evidence presented for the correction influence the CIE, may be particularly important in courtrooms where jurors may be faced with counter-evidence, or in situations in which scientific evidence is used to counter misinformation about an issue.

The relationship between scenario structure, content, and the magnitude of the CIE, also warrants further exploration. My research has shown that some CIE findings do not extend to different scenarios (Bush et al., 1994; Rich & Zaragoza, 2016). One explanation for why previous findings were not replicated in the scenarios developed for this programme of research, could be that the congruency of additional event information with the causal explanation offered by misinformation was reduced relative to prior studies. Some scenarios may require a higher degree of inference to establish a causal explanation for the outcome described. This was not experimentally manipulated in the experiments reported in this thesis but is a worthy avenue of further investigation. Scenario factors such as structure, temporal sequencing, content, ambiguousness, or description of spatial elements, and prior knowledge or experience of the situations similar to the scenario, could be systematically varied to establish their impact on the CIE. This in turn may help establish the precise mechanisms that bring about the CIE.

Finally, the roles of incentivisation and the valence of the outcome described in the scenario may also moderate the CIE, and therefore deserve further investigation. Studies on the CIE typically use neutral scenarios in which there is no inherent reason to believe the correction over the misinformation presented. In the real-world people are often incentivised to believe certain things over others: either because we are motivated to confirm our pre-existing beliefs (or prejudices) or because our social status/job is preserving a particular belief. One reason that people continue to rely on misinformation in CIE studies could be that people are simply motivated to believe (or respond with) the less cognitively demanding piece of information, if there is no incentive to believe the correction over the misinformation. However, if there are personal repercussions involved in the continued use of misinformation then people might be less inclined to do so. This could be experimentally implemented by creating a situation in which continued use of misinformation comes with a monetary or social cost – such as in experimental game theory studies.

Regarding the valence of the described outcome, it may seem trivial to point out that all CIE studies have used scenarios with negative outcomes. Prior work on the CIE has examined how the valence of misinformation (i.e. positive, negative, or neutral) affects its continuing impact on reasoning and found that negatively valenced information is more likely to be used to answer inferential questions than positively valenced, or neutral, misinformation (Guillory & Geraci, 2016). Misinformation may therefore have less of an impact in scenarios with positive outcomes and this may have implications for the types of situations that the CIE is likely to occur.

5.4. Theoretical Insights and Implications

The findings reported in this thesis provide insight into the theoretical foundations of the CIE. Theoretically, the present work raises questions about whether observing the CIE in experimental settings reflects a mental-model updating failure, strategic memory process failure, pragmatic demand, or simple recall (or perhaps something else entirely). The patterns of findings imply that the CIE is less frequent, smaller in magnitude, and more fragile

than generally assumed. This is because there are several factors which are not usually discussed in the CIE literature (e.g. the scenario of context that misinformation appears in, the ambiguity of the scenario, and the scope of the correction to invalidate the misinformation), but which nonetheless can affect whether and how the CIE occurs.

5.4.1. Mechanistic accounts

The findings reported in this thesis have implications for the mechanisms by which the CIE occurs. In particular, this thesis has implications for the idea that the CIE is driven by a failure, or unwillingness, to update a mental-model of an event unless an alternative is provided (Johnson-Laird, 1980; Johnson & Seifert, 1994; Seifert, 2002). First, my findings suggest that a coherent but incorrect model can be abandoned in favour a correct but incoherent model when the correction provides evidence to refute the explanation offered by misinformation. Therefore, it does not matter if a correction creates a gap in the model so long as it sufficiently invalidates the explanation offered by the misinformation (e.g. a test showed the driver had no alcohol in his system vs. no oil paint and gas cylinders ever in the warehouse).

These findings could, however, still be interpreted as facilitating mental-model updating of an event. For instance, Putnam, Wahlheim and Jacoby (2014) have argued that factors which enhance detection of a conflict between competing event interpretations enable updating. Similarly, Kendeou, Walsh, Smith, and O'Brien (2014) have proposed that for effective knowledge revision to occur both invalidated and correction event interpretations must be activated. Empirical evidence for this comes from Ecker et al. (2017) who found that *explicitly* repeating the misinformation during correction was more effective at reducing misinformation reliance than avoiding repetition of the misinformation or explaining why misinformation was incorrect without repeating it. Findings from experiments reported in this thesis that used the crash scenario are concordant with this explanation. More specifically, stating that 'a test showed the driver had no alcohol in his system' could arguably

make the conflict between the competing interpretations of the event more apparent.

The results of experiments reported in this thesis also suggest that the CIE does not always arise from a failure to update one's mental model of an event. More specifically, the results obtained by introducing of a novel control condition suggest that – at least under some circumstances - the CIE may occur because participants search their memory for possible causes when asked inferential questions but fail to retrieve the information correcting the misinformation. The finding that participants still referred to misinformation even when it was only mentioned during the correction suggest that it is the availability of the causal explanation offered by misinformation rather than its role in the mental-model which drives some instances of the CIE.

This finding could be interpreted as supporting a retrieval-failure account of the CIE. More specifically, when strategic memory processes are not engaged (either through lack of motivation or because of some detriment), participants may fail to retrieve the source and validity of the correction, and therefore rely on the explanation which was automatically activated by the inference question. If automatic processes are employed the "negation tag" linked to misinformation (e.g. a man seen running away did NOT assault the woman), may be lost because strategic processes are necessary to retrieve the source and veracity of the information. The findings are also consistent with research suggesting that belief perseverance – the tendency to cling to newly created beliefs when they are discredited – is mediated by the availability of causal arguments supporting initial beliefs (cf. Anderson et al. 1985).

The fact that the findings reported in this thesis, as well as in the CIE literature more generally, can be readily interpreted as supporting either the retrieval failure or mental-model updating accounts emphasises the need for a more systematic approach to studying the CIE. Furthermore, because there are significant issues surrounding the limited use of scenarios, different mechanisms may bring about the CIE depending on the specifics of the experimental method and scenario presented to participants. This highlights

the need to carefully examine whether specifics of the scenario moderate the CIE.

5.4.2. Demand effects

An alternative – and perhaps more prosaic - interpretation of CIE findings, posits that the CIE reflects a demand effect of the methodology itself. Such an interpretation would see that a participant who exhibits the CIE may ask themselves why the report (or the experimenter) only provided one potential cause, which was then refuted, and then asked to answer a series of questions which seemingly require the corrected cause to be answered. The problem could also lie in the Gricean pragmatics of the experimental situation whereby participants contemplate why the experimenter would present information about a potential cause, say that it was not true, and then ask questions about the potential cause information. Thus, participants in CIE studies use the causal information presented in the story because it is the only relevant information available. Support for the idea that the CIE arises from a pragmatic demand in experimental settings comes from the fact that some participants did mention potential alternative causes when potential cause information was provided immediately before the misinformation statement (e.g. stating that the woman's head injuries were sustained when a car hit her rather than from an assault).

Assumptions about the "communicative intention" of the information provided by the experimenter can inform how participants determine the relevance of the information provided to them. These assumptions can result in what appear to be judgemental errors that do not conform to the normative model the experimenter has in mind which only considers the literal meaning of the statement and not the communicative intention (Bless et al., 1993). If the CIE is the result of pragmatic demand in research settings, then findings obtained through this methodology are of limited value for understanding the cognitive mechanisms involved in processing corrections to misinformation. Furthermore, findings obtained through this methodology may be uninformative with respect to developing counter-misinformation strategies.

5.4.3. Gricean pragmatics of contradictory statements

The findings also have implications for the Gricean pragmatics (Grice, 1975) of presenting contradictory information. Seifert (2002; see also Johnson & Seifert, 1994) argued that corrections pose a problem for comprehension because people expect human generated information to be relevant and truthful. Prior CIE work suggests that corrections which explain how misinformation occurred reduce the CIE more than corrections which address only the literal implications of the contradiction (Bush et al., 1994). The results of the experiments reported in this thesis do not support this distinction. In fact, corrections which addressed the literal implications of the contradiction were as effective as those that addressed the pragmatic implications (i.e. by explaining that misinformation originated from an error or a lie). Furthermore, the correction information (without the lie or error explanation) included in the crash scenario addressed only the literal content of the contradiction by providing evidence that the driver had not been drinking, yet this correction almost eliminated the CIE. These findings suggest that Gricean pragmatics may only play a minor role in comprehending contradictions – or at least that the pragmatic inferences people make when faced with a contradiction are complex and not well understood.

5.4.4. Formal approaches to modelling the CIE

The CIE represents a descriptive model of how people process corrections to misinformation which is often assumed to depend on having a coherent, causally related account in which a single or minimal correction has a significant impact on the construal of meaning (Johnson & Seifert, 1994). The continued reliance on misinformation after a correction is often depicted as a bias – or systematic deviation from a normative standard - and therefore irrational (e.g. Lewandowsky et al., 2012). This perspective assumes two things; first, that the optimal solution is always to disregard initially prior information in favour of new information, and second that the 'true' value of the correction is known.

To date, the literature on the CIE has provided no normative account of how people should "optimally" process corrections to misinformation. The lack of formulism is important because there may be situations in which continued reliance on misinformation is rational given the sparsity of information and inherent uncertainty of the situation at hand. In uncertain situations with sparse information, people may use cues, such as source credibility/reliability, to assess the validity of misinformation and its correction, and decide how much to incorporate these pieces of information into their beliefs. Indeed, research on the CIE suggests that the reliability - or credibility - of the sources providing the correction moderates the continued impact of misinformation (Guillory & Geraci, 2010; 2013).

One common normative standard of inference is Bayesian belief revision. Bayes' Theorem provides a normative rule for updating beliefs considering new evidence and is therefore valuable for studying human reasoning. Normative predictions derived from Bayes' Theorem can be compared to participant's responses in experiments. Bayesian probability has been used to study various aspects of human reasoning. For instance, Bayesian probability has been used to study judgement (Tversky & Kahneman, 1983), conditional reasoning (Oaksford & Chater, 2003, 2007; Over, 2009), argumentation (Hahn & Oaksford, 2007), as well as other areas of cognition (Chater, Oaksford, Hahn, & Heit, 2010).

The Bayesian network (BN) framework, in particular, is ideal for examining whether the CIE is rational in some circumstances, because it provides the means to test people's causal models of scenarios - including their models of the reliability and credibility of the sources providing information - and compare inferences to a normative standard (Fenton, Neil, & Lagnado, 2013; Lagnado, Fenton, & Neil, 2013; Pearl, 1988, 2001). Bayesian networks (BN) use graph structures to represent the probabilistic relationships between hypotheses and evidence (including reliability), using conditional probabilities to represent the strength of relations, and show what inferences are rationally permitted from a given model of the available information. The BN approach has been used to model inferences about the convincingness of

arguments from experts in terms of their access to (expertise), and capacity to convey (trustworthiness), accurate information (Harris, Hahn, Madsen, & Hsu, 2016), and shown that participants' quantitative judgements are broadly consistent with Bayesian model predictions.

As noted in the previous section, factors such as congruency of additional information with the misinformation explanation, and the reliability (or credibility) of sources providing the misinformation or the correction, are potential moderators of the CIE. The BN framework could be used to test some basic elements of the CIE, such as a source who initially presents information but then later contradicts themselves, to examine what inferences are rationally permitted for a given causal model of available information.

A basic model of the CIE that incorporates the reliability of sources would include a hypothesis (e.g. carelessly stored flammable liquids caused the fire), which is confirmed by the source of misinformation, but then contradicted by the same source later. This could be used to examine how much the reliability of a single source who contradicts themselves should be penalised for the contradiction, and how much to update belief in the hypothesis when this occurs. This could be compared to a situation in which the misinformation and correction (contradiction) come from different sources to examine participant's judgements about the probability of the hypothesis given the contradictory reports. Comparing actual judgements to predictions from a Bayesian model might reveal whether there are situations in which retaining belief in misinformation after a correction is rational. Formally modelling the causal relations between information included in a scenario would make it possible to test participants' causal models of scenarios. This would therefore provide better understanding the cognitive mechanisms involved in the CIE, and the strategies that might be useful for improving reasoning about misinformation.

5.5. Practical Implications

Throughout this thesis I have shown that the CIE occurs despite clear corrections that explain how misinformation originated, and irrespective of

whether misinformation implies or explicitly states a likely cause of the adverse outcome described in the scenario. Through this research I have also shown that the CIE is not guaranteed to arise under all circumstances and is substantially attenuated when a correction provided information that invalidates the causal explanation offered by the misinformation. In addition to this, by introducing a novel control condition in which misinformation is only presented as part of a correction, I showed that at least part of the CIE can be explained by the availability of causal explanations when answering causal inference questions.

These findings have important implications in the domains such as the media, science, law, healthcare, education, and politics. Explanations are intended to clarify the causes, context, and consequences of the set of facts they describe. Experts usually make use of explanations when attempting to argue the case for or against a set of facts. Explaining that a piece of erroneous information originated from a deception or honest mistake may have little impact in reducing misinformation reliance over and above simply stating that the information is incorrect, as observed in Chapter 3. Furthermore, news reports, or blogs, that make use of innuendo and speculation when the full facts are unknown, may be as difficult to correct as erroneous information that is directly asserts a cause of an outcome, as evidenced in Chapter 4. Furthermore, the findings reported in this thesis suggest that the extent to which people continue to rely on misinformation might depend on the degree of inference required by the correction, as shown in Chapters 3 and 4. This suggests that, where possible, corrections to misinformation should make use of clear and incontrovertible evidence that sufficiently invalidates erroneous information rather simply negating it. Finally, the mere mention of misinformation during its correction could be sufficient to instigate continued reliance on misinformation, if this explanation for an event is available during retrieval, as evidenced in Chapters 2 and 4. This finding suggests that the mere mention of misinformation in a correction could be sufficient to trigger a continued reliance on misinformation even if people were not initially exposed to misinformation.

There are also important practical implications in terms of terms of using the CIE make policy recommendations (e.g. Lewandowsky et al., 2012). There is still a lack of resolution about the precise set of circumstances that bring about the CIE. Furthermore, it is unclear whether observing the CIE is limited to artificial scenarios that do not reflect how people encounter misinformation the real world. Thus, it is still unclear whether it is possible to use findings from CIE studies to infer anything about the prevalence of the CIE in society. This means that researchers should be careful about making policy recommendations from the findings that could be easily influenced by the type of task and stimuli used in experiments.

References

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(5), 1061-1070.

Anderson, C. A. (1982). Inoculation and counterexplanation: Debiasing techniques in the perseverance of social theories. *Social Cognition*, *1*(2), 126–139.

Anderson, C. A. (1983). Abstract and Concrete Data in the Perseverance of Social Theories: When Weak Data Lead to Unshakeable Beliefs. *Journal of Experlmental Social Psychology*, *19*, 93–108.

Anderson, C. A., & Kellam, K. L. (1992). Belief Perseverance, Biased Assimilation, and Covariation Detection: The Effects of Hypothetical Social Theories and New Data. *Personality and Social Psychology Bulletin*, *18*(5), 555–565. http://doi.org/10.1177/0146167292185005

Anderson, C. A., New, B. L., & Speer, J. R. (1985). Argument availability as a mediator of social theory perseverance. *Social Cognition*, *3*(3), 235–249. http://doi.org/10.1521/soco.1985.3.3.235

Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review*, *5*(1), 1–21. http://doi.org/10.3758/BF03209454

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 603.

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods*, *43*(3), 800-813.

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. http://doi.org/10.1111/ajps.12081

Berman, G. L., & Cutler, B. L. (1996). Effects of inconsistencies in eyewitness testimony on mock-juror decision making. *Journal of Applied Psychology*, *81*(2), 170–177. http://doi.org/10.1037/0021-9010.81.2.170

Berman, G. L., Narby, D. J., & Cutler, B. L. (1995). Effects of inconsistent eyewitness statements on mock-jurors' evaluations of the eyewitness, perceptions of defendant culpability and verdicts. *Law and Human Behavior*, *19*(1), 79–88. http://doi.org/10.1007/BF01499074

Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology*, *37*(1), 48–74. http://doi.org/10.1037/0022-3514.37.1.48

Birnbaum, M. H., Wong, R., & Wong, L. K. (1976). *Combining information from sources that vary in credibility*. *Memory & Cognition* (Vol. 4). Retrieved from https://link.springer.com/content/pdf/10.3758/BF03213185.pdf

Bjork, E. L., & Bjork, R. A. (1996). Continuing influences of to-be-forgotten information. *Consciousness and Cognition*, *5*(1–2), 176–196.

Bjork, R. A., Bjork, E. L., & Anderson, M. C. (1998). Varieties of goal-directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), Intentional forgetting: Interdisciplinary approaches (pp. 103–137). Mahwah, NJ: Lawrence ErlbaumAssociates.

Bjork, R. A., & Woodward, A. E. (1973). Directed forgetting of individual words in free recall. *Journal of Experimental Psychology*, *99*(1), 22–27. Retrieved from https://bjorklab.psych.ucla.edu/wp-content/uploads/sites/13/2016/07/RBjork_Woodward_1973_JEPdforgetting.pdf

Bless, H., Strack, F., & Schwarz, N. (1993). The informative functions of research procedures: Bias and the logic of conversation. *European Journal of Social Psychology*, *23*(2), 149–165. http://doi.org/10.1002/ejsp.2420230204

Borckardt, J. J., Sprohge, E., & Nash, M. (2003). Effects of the Inclusion and Refutation of Peripheral Details on Eyewitness Credibility. *Journal of Applied Social Psychology*, *33*(10), 2187–2197. http://doi.org/10.1111/j.1559-1816.2003.tb01880.x

Brydges, C. R., Gignac, G. E., & Ecker, U. K. (2018). Working memory capacity, short-term memory capacity, and the continued influence effect: A latent-variable analysis. *Intelligence*, *69*, 117–122.

Bush, J. G., Johnson, H. M., & Seifert, C. M. (1994). The implications of corrections: Then why did you mention it. In Proceedings of the 16th Annual Conference of the Cognitive Science Society, pp. 112-117.

Capella, J. N., Ophir, Y., & Sutton, J. (2018). The Importance of Measuring Knowledge in the Age of Misinformation and Challenges in the Tobacco Domain. In B. . Southwell, E. Thorson, & L. Sheble (Eds.), *Misinformation and Mass Audiences* (pp. 51–70). University of Texas Press.

Carretta, T. R., & Moreland, R. L. (1983). The Direct and Indirect Effects of Inadmissible Evidence 1. *Journal of Applied Social Psychology*, *13*(4), 291–309.

Chambers, K. K. L., & Zaragoza, M. M. S. (2001). Intended and unintended effects of explicit warnings on eyewitness suggestibility: Evidence from source identification tests. *Memory & Cognition*, *29*(8), 1120–1129. http://doi.org/10.3758/BF03206381

Chater, N., & Oaksford, M. (2013). Programs as causal models: Speculations

on mental programs and mental representation. *Cognitive Science*, *37*(6), 1171–1191.

Connor Desai, S., Reimers, S & Lagnado, D. (2016). Consistency and credibility in legal reasoning: A Bayesian network approach. In Proceedings of the 38th Annual Conference of the Cognitive Science Society, pp. 626-631.

Open Science Collaboration (2012). An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. *Perspectives on Psychological Science*, *7*(6), 657–660. http://doi.org/10.1177/1745691612462588

Cook, J., Ecker, U., & Lewandowsky, S. (2015). Misinformation and How to Correct It. In *Emerging Trends in the Social and Behavioral Sciences* http://doi.org/10.1002/9781118900772.etrds0222.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, *163*, 163–228.

Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*(2), 159-180.

Craik, F. I., & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(3), 474–479. http://doi.org/10.1037/0278-7393.13.3.474

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLoS ONE*, *8*(3), e57410. http://doi.org/10.1371/journal.pone.0057410

Davies, M. F. (1997). Belief persistence after evidential discrediting: The impact of generated versus provided explanations on the likelihood of discredited outcomes. *Journal of Experimental Social Psychology*, *33*(6), 561–578. http://doi.org/10.1006/jesp.1997.1336

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., … Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*(3). http://doi.org/10.1073/pnas.1517441113

Duffy, S. A., Shinjo, M., & Myers, J. L. (1990). The effect of encoding task on memory for sentence pairs varying in causal relatedness. *Journal of Memory and Language*, *29*(1), 27–42. http://doi.org/10.1016/0749-596X(90)90008-N

Ecker, U. K. H., Lewandowsky, S., Fenton, O., & Martin, K. (2014). Do people keep believing because they want to? Preexisting attitudes and the continued influence of misinformation. *Memory & Cognition*, *42*(2), 292–304. http://doi.org/10.3758/s13421-013-0358-x

Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, *38*(8), 1087–1100.

http://doi.org/10.3758/MC.38.8.1087

Ecker, U. K., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition*, *6*(2), 185–192.

Ecker, U. K., Lewandowsky, S., & Apai, J. (2011). Terrorists brought down the plane!—No, actually it was a technical fault: Processing corrections of emotive information. *The Quarterly Journal of Experimental Psychology*, *64*(2), 283–310.

Ecker, U. K., Lewandowsky, S., Cheung, C. S., & Maybery, M. T. (2015). He did it! She did it! No, she did not! Multiple causal explanations and the continued influence of misinformation. *Journal of Memory and Language*, *85*, 101–115.

Ecker, U. K., Lewandowsky, S., Swire, B., & Chang, D. (2011). Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction. *Psychonomic Bulletin & Review*, *18*(3), 570–578.

Ecker, U. K., Swire, B., Lewandowsky, S.(2014). Correcting Misinformation—A Challenge for Education and Cognitive Science. In Rapp D. N. & Braasch J. L. G. (Eds.), *Processing Inaccurate Information*: *Theoretical and Applied Perspectives from Cognitive Science and the Educational Sciences* (pp. 13–38). 2014; Cambridge, MA: MIT Press

Ecker, U., Lewandowsky, S., Cheung, C., & Maybery, M. (2015). He did it! She did it! No, she did not! Multiple causal explanations and the continued influence of misinformation. *Journal of Memory and Language*, *85*(October), 101–115. http://doi.org/10.1016/j.jml.2015.09.002

Elliott, R., Farrington, B., & Manheimer, H. (1988). Eyewitnesses credible and discredible. *Journal of Applied Social Psychology*, *18*(16), 1411–1422.

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge Does Not Protect Against Illusory Truth, *144*(5), 993–1002. Retrieved from http://dx.doi.org/10.1037/xge0000098.supp

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*(11), 1215–1226. http://doi.org/10.1177/01461672972311008

Fisher, R. P., Brewer, N., & Mitchell, G. (2009). The Relation between Consistency and Accuracy of Eyewitness Testimony: Legal versus Cognitive Explanations. *Handbook of Psychology of Investigative Interviewing: Current Developments and Future Directions*, (January), 121–136. http://doi.org/10.1002/9780470747599.ch8

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25–42. http://doi.org/10.1257/089533005775196732

Frew, E. J., Whynes, D. K., & Wolstenholme, J. L. (2003). Eliciting Willingness to Pay: Comparing Closed-Ended with Open-Ended and Payment Scale

Formats. *Medical Decision Making*, *23*(2), 150–159. http://doi.org/10.1177/0272989X03251245

Frew, E. J., Wolstenholme, J. L., & Whynes, D. K. (2004). Comparing willingness-to-pay: bidding game format versus open-ended and payment scale formats. *Health Policy*, *68*(3), 289–298.

Full Fact. (2017). £350 million EU claim "a clear misuse of official statistics." Retrieved August 8, 2018, from https://fullfact.org/europe/350-million-week-boris-johnson-statistics-authority-misuse/

Gallo, D. A., Roediger, H. L. H., & McDermott, K. K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review*, *8*(3), 579–586. http://doi.org/10.3758/BF03196194

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. http://doi.org/10.3758/s13423-012-0296-9

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the Unbelievable: Some Problems in the Rejection of False Information. *Journal of Personality and Social Psychology*, *59*(4), 601–613. http://doi.org/10.1037/0022-3514.59.4.601

Gordon, A., Brooks, J. C., Quadflieg, S., Ecker, U. K., & Lewandowsky, S. (2017). Exploring the neural substrates of misinformation processing. *Neuropsychologia*, *106*, 216–224.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371.

Graesser, A., Ozuru, Y., & Sullins, J. (2010). What is a good question? In M McKeown & G. Kucan (Eds.), *Bringing reading research to life* (pp. 112–141). New York: Guilford.

Greitemeyer, T. (2014). Article retracted, but the message lives on. *Psychonomic Bulletin & Review*, *21*(2), 557–561. http://doi.org/10.3758/s13423-013-0500-6

Grice, H.P. (1975). Logic and conversation, in P. Cole and J. Morgan (eds) Syntax and Semantics 3: Speech Acts, New York: Academic Press

Grysman, A. (2015). Collecting narrative data on Amazon's Mechanical Turk. *Applied Cognitive Psychology*, *29*(4), 573–583.

Guillory, J. J., & Geraci, L. (2010). The persistence of inferences in memory for younger and older adults: remembering facts and believing inferences. *Psychonomic Bulletin & Review*, *17*(1), 73–81. http://doi.org/10.3758/PBR.17.1.73

Guillory, J. J., & Geraci, L. (2013). Correcting erroneous inferences in memory: The role of source credibility. *Journal of Applied Research in Memory and Cognition*, *2*(4), 201–209. http://doi.org/10.1016/j.jarmac.2013.10.001

Guillory, J. J., & Geraci, L. (2016). The Persistence of Erroneous Information

in Memory: The Effect of Valence on the Acceptance of Corrected Information. *Applied Cognitive Psychology*, *30*(2), 282–288. http://doi.org/10.1002/acp.3183

Hardwicke, T. E. (2016). *Persistence and plasticity in the human memory system: An empirical investigation of the overwriting hypothesis*. University College London, London.

Harris, A. J., & Hahn, U. (2009). Bayesian rationality in evaluating multiple testimonies: Incorporating the role of coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1366.

Harris, A. J. L. L., Hahn, U., Madsen, J. K., & Hsu, A. S. (2016). The Appeal to Expert Opinion: Quantitative Support for a Bayesian Network Approach. *Cognitive Science*, *40*(6), 1496–1533. http://doi.org/10.1111/cogs.12276

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

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, *39*(12), 1372-1388.

Hatvany, N., & Strack, F. (1980). The Impact of A Discredited Key Witness. *Journal of Applied Social Psychology*, *10*(6), 490–509. http://doi.org/10.1111/j.1559-1816.1980.tb00728.x

Heller, K. J. (2006). The Cognitive Psychology of Circumstantial Evidence. Michigan Law Review, 105(2), 241-305.

Herron, J. E., & Rugg, M. D. (2003). Retrieval Orientation and the Control of Recollection. *Journal of Cognitive Neuroscience*, *15*(6), 843–854. Retrieved from http://orca.cf.ac.uk/52205/1/Herron 2003.pdf

Hilton, D. J. (1990). Conversational processes and causal explanation. *Psychological Bulletin*, *107*(1), 65-81.

Horne, Z., Powell, D., Hummel, J. E., & Holyoak, K. J. (2015). Countering antivaccination attitudes. *Proceedings of the National Academy of Sciences*, *112*(33), 10321–10324. http://doi.org/10.1073/pnas.1504019112

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, *15*(4), 635–650.

Igou, E. R., & Bless, H. (2003). Inferring the importance of arguments: Order effects and conversational rules. *Journal of Experimental Social Psychology*, *39*(1), 91–99.

Igou, E. R., & Bless, H. (2005). The conversational basis for the dilution effect. *Journal of Language and Social Psychology*, *24*(1), 25–35.

Independent Press Standards Organisation. (2016). Editors' Code of Practice. Retrieved August 18, 2018, from https://www.ipso.co.uk/editors-code-of-practice/#Accuracy

Ipsos Mori. (2016). How Britain voted in the 2016 EU referendum.

Jacoby, L. L. (1991). A Process Dissociation Framework: Separating Automatic from Intentional Uses of Memory. *Journal of Memory and Language*, *30*, 513–541.

Jacoby, L. L. (1996). Dissociating automatic and consciously controlled effects of study/test compatibility. *Journal of Memory and Language*, *35*(1), 32–52.

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, *4*(1), 71–115.

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness*. Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, *107*(43), 18243–18250.

Johnson, H. M. (1994). Processes of successful intentional forgetting. *Psychological Bulletin*, *116*(2), 274-292.

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

Johnson, H. M., & Seifert, C. M. (1999). Modifying mental representations: Comprehending corrections. *The Construction of Mental Representations during Reading*, 303–318.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*. 114(1), 3. http://doi.org/10.1037/0033-2909.114.1.3

Kassin, S. M., & Sommers, S. R. (1997). Inadmissible Testimony, Instructions to Disregard, and the Jury: Substantive Versus Procedural Considerations. *Personality and Social Psychology Bulletin*. http://doi.org/10.1177/01461672972310005

Kassin, S. M., & Sukel, H. (1997). Coerced confessions and the jury: An experimental test of the “harmless error” rule. *Law and Human Behavior*, *21*(1), 27–46.

Khemlani, S. S., Sussman, A. B., & Oppenheimer, D. M. (2010). Harry Potter and the sorcerer’s scope: latent scope biases in explanatory reasoning. *Memory & Cognition*, *39*, 527–535. http://doi.org/10.3758/s13421-010-0028-1

Khoe, W., Kroll, N. E., Yonelinas, A. P., Dobbins, I. G., & Knight, R. T. (2000). The contribution of recollection and familiarity to yes–no and forced-choice recognition tests in healthy subjects and amnesics. *Neuropsychologia*, *38*(10), 1333–1341.

Kintsch, W., & van Dijk, T. a. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. http://doi.org/10.1037/0033-295X.85.5.363

Knowlton, B. J., & Squire, L. R. (1995). Remembering and Knowing: Two Different Expressions of Declarative Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*. Retrieved from https://s3.amazonaws.com/academia.edu.documents/41045675/Knowlton_Squire.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1528389124&Signature=ES09b0DTMkDrxjja026ieAl58cY%3D&response-content-disposition=inline%3B filename%3DRemembering_and_knowing_Two_different_ex.pdf

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480-498.

Kunda, Z., & Oleson, K. C. (1995). Maintaining stereotypes in the face of disconfirmation: Constructing grounds for subtyping deviants. *Journal of Personality and Social Psychology*, *68*(4), 565-579.

Lagnado, D. (1994). The psychology of explanation: A Bayesian approach. *Unpublished Master's Thesis, University of Birmingham, Birmingham, England*.

Lagnado, D. A. (2011). Causal thinking. In P. M. Illari, F. Russo, & J. Williamson (Eds.), Causality in the sciences (pp. 129–149). Oxford: Oxford University Press.

Lagnado, D. A., Fenton, N., & Neil, M. (2013). Legal idioms: a framework for evidential reasoning. *Argument & Computation*, *4*(1), 46–63.

Lagnado, D. a, & Harvey, N. (2008). The impact of discredited evidence. *Psychonomic Bulletin & Review*, *15*(6), 1166–1173. http://doi.org/10.3758/PBR.15.6.1166

Landau, J. D., Glahn, N. Von, & von Glahn, N. (2004). Warnings Reduce the Magnitude of the Imagination Inflation Effect. *The American Journal of Psychology*, *117*(4), 579-593. http://doi.org/10.2307/4148993

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.

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

Lewandowsky, S., Ecker, U. K., & Cook, J. (2017). Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era. *Journal of Applied Research in Memory and Cognition*, *6*(4), 353–369.

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

Light, L. L., & Anderson, P. A. (1985). Working-memory capacity, age, and memory for discourse. *Journal of Gerontology*, *40*(6), 737–747.

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, *7*(4), 560–572.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology*, *55*, 232–257.

http://doi.org/10.1016/j.cogpsych.2006.09.006

Lombrozo, T. (2016). Explanation. In J. Sytsma & W. Buckwalter (Eds.),

Blackwell Companion to Experimental Philosophy (pp. 491–503).

Blackwell. doi:10.1002/9781118661666.ch34.

Macleod, C. M. (1989). Directed forgetting affects both direct and indirect tests of memory. *Journal of Experimental Psychology Learning Memory and Cognition*, *15*(1), 13–21. http://doi.org/10.1037/0278-7393.15.1.13

MacLeod, C. M. (1999). The item and list methods of directed forgetting: Test differences and the role of demand characteristics. *Psychonomic Bulletin and Review*, *6*(1), 123–129. http://doi.org/10.3758/BF03210819

Madsen, J. K., Bailey, R. M., & Pilditch, T. D. (2018). Large networks of rational agents form persistent echo chambers. *Scientific Reports*, *8*(1), 12391.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, *7*(6), 537–542.

Marsh, E. J., Meade, M. L., & Roediger III, H. L. (2003). Learning facts from fiction. *Journal of Memory and Language*, *49*(4), 519–536.

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, *114*(1), 1-16.

McDermott, K. B., & Chan, J. C. (2006). Effects of repetition on memory for pragmatic inferences. *Memory & Cognition*, *34*(6), 1273–1284.

McGinnies, E., & Ward, C. D. (1980). Better Liked than Right: Trustworthiness and expertise as factors in credibility. *Personality and Social Psychology Bulletin*, *6*(3), 467–472. http://doi.org/10.1177/014616728063023

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*(3), 440–466. http://doi.org/10.1037/0033-295X.99.3.440

Moons, W. G., Mackie, D. M., & Garcia-Marques, T. (2009). The impact of repetition-induced familiarity on agreement with weak and strong arguments. *Journal of Personality and Social Psychology*, *96*(1), 32–44. http://doi.org/10.1037/a0013461

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of Bayes factors for common designs.*

Mussweiler, T., & Neumann, R. (2000). Sources of mental contamination: Comparing the effects of self-generated versus externally provided primes. *Journal of Experimental Social Psychology*, *36*(2), 194–206. http://doi.org/10.1006/jesp.1999.1415

Nyhan, B., & Reifler, J. (2010). When Corrections Fail : The Persistence of Political Misperceptions, *32*(2), 303–330.

Nyhan, B., & Reifler, J. (2015). Displacing Misinformation about Events: An Experimental Test of Causal Corrections. *Journal of Experimental Political Science*, *2*(01), 81–93. http://doi.org/10.1017/XPS.2014.22

Nyhan, B., Reifler, J., & Ubel, P. a. (2013). The hazards of correcting myths about health care reform. *Medical Care*, *51*(2). http://doi.org/10.1097/MLR.0b013e318279486b

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716-aac4716. http://doi.org/10.1126/science.aac4716

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. http://doi.org/10.1016/j.jesp.2009.03.009

Ozuru, Y., Briner, S., Kurby, C. A., & McNamara, D. S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *67*(3), 215–227. http://doi.org/10.1037/a0032918

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163.

Peer, E., Vosgerau, J., & Acquisti, A. (2013). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, *46*(4), 1023–1031. http://doi.org/10.3758/s13428-013-0434-y

Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 521–533. http://doi.org/http://dx.doi.org/10.1037/0278-7393.14.3.521

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the Story Model for juror decision making. *Journal of Personality and Social Psychology*, *62*(2), 189–206. http://doi.org/10.1037/0022-3514.62.2.189

Pennington, N., & Hastie, R. (1993). Reasoning in explanation-based decision making. *Cognition*, *49*(1–2), 123–163.

Pennycook, G., Cannon, T. D., & Rand, D. G. (2017). Prior Exposure Increases Perceived Accuracy of Fake News. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2958246

Pornpitakpan, C. (2004). The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence. *Journal of Applied Social Psychology*, *34*(2), 243–281. http://doi.org/10.1111/j.1559-1816.2004.tb02547.x

Prull, M. W., Dawes, L. L. C., Martin III, A. M., Rosenberg, H. F., & Light, L. L. (2006). Recollection and familiarity in recognition memory: adult age differences and neuropsychological test correlates. *Psychology and Aging*, *21*(1), 107-118.

Putnam, A. L., Wahlheim, C. N., & Jacoby, L. L. (2014). Memory for flip-flopping: Detection and recollection of political contradictions. *Memory & Cognition*, *42*, 1198–1210. http://doi.org/10.3758/s13421-014-0419-9

Rapp, D. N., & Kendeou, P. (2007). Revising what readers know: Updating text representations during narrative comprehension. *Memory & Cognition*, *35*(8), 2019–2032.

Read, S. J., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology*, *65*(3), 429-477.

Reimers, S., & Stewart, N. (2007). Adobe Flash as a medium for online experimentation: A test of reaction time measurement capabilities. *Behavior Research Methods*, *39*(3), 365–370.

Reimers, S., & Stewart, N. (2015). Presentation and response timing accuracy in Adobe Flash and HTML5/JavaScript Web experiments. *Behavior Research Methods*, *47*(2), 309–327. http://doi.org/10.3758/s13428-014-0471-1

Reja, U., Manfreda, K. L., Hlebec, V., & Vehovar, V. (2003). Open-ended vs. Close-ended Questions in Web Questionnaires. *Developments in Applied Statistics*, *19*(August 2014), 159–177. Retrieved from http://www.websm.org/uploadi/editor/Reja_2003_open_vs_close-ended_questions.pdf

Rich, P. R., & Zaragoza, M. S. (2016). The continued influence of implied and explicitly stated misinformation in news reports. *Journal of Experimental Psychology. Learning, Memory & Cognition*, *42*(1), 62–74. http://doi.org/http://dx.doi.org/10.1037/xlm0000155

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*(3), 385–407.

Ross, L., & Anderson, C. A. 1982. Shortcomings in the attribution process: On the origins and maintenance of erroneous social assessments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases: 129-152. Cambridge, England: Cambridge University Press.

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in Self-Perception and Social Perception: Biased Attributional Processes in the Debriefing Paradigm. *Journal Oj Personality and Social Psychology*, *32*(5), 880–802. Retrieved from https://s3.amazonaws.com/academia.edu.documents/39942959/Perseverance_in_selfperception_and_social_perception.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1526991886&Signature=3W0LYbqevIK%2BCb5sZMe3%2FxSN7Hg%3D&response-content-disposition=inline%3B f

Schul, Y., & Manzury, F. (1990). The effects of type of encoding and strength of discounting appeal on the success of ignoring an invalid testimony. *European Journal of Social Psychology*, *20*(4), 337–349. http://doi.org/10.1002/ejsp.2420200405

Schul, Y., & Mayo, R. (2014). Discounting Information: When False Information is Preserved and When it is Not. In *Processing Inaccurate Information: Theoretical and Applied Perspectives from Cognitive*

Science and the Educational Sciences (pp. 203–221). MIT Press.

Schul, Y., Mayo, R., & Burnstein, E. (2004). Encoding under trust and distrust: the spontaneous activation of incongruent cognitions. *Journal of Personality and Social Psychology*, *86*(5), 668-679.

Schul, Y., & Mazursky, D. (1990). Conditions facilitating successful discounting in consumer decision making. *Journal of Consumer Research*, *16*(4), 442–451.

Schwarz, N. (1996). Cognition and communication: Judgment biases, research methods, and the logic of conversation. Hillsdale, N.J.: Lawrence Erlbaum.

Schwarz, N., Hippler, H. J., Deutsch, B., & Strack, F. (1985). Response Scales - Effects of Category Range on Reported Behavior and Comparative Judgments. *Public Opinion Quarterly*, *49*(3), 388–395. http://doi.org/10.1086/268936

Schwarz, N., Knäuper, B., Hippler, H.-J., Noelle-Neumann, E., & Clark, L. (1990). Response scales: Effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly*. 49:388–95.

Schwarz, N., Strack, F., Müller, G., & Chassein, B. (1988). The range of response alternatives may determine the meaning of the question: Further evidence on informative functions of response alternatives. *Social Cognition*, *6*(2), 107–117.

Seifert, C. M. (2002). The continued influence of misinformation in memory: What makes a correction effective? *Psychology of Learning and Motivation*, *41*(26), 265–294.

Seifert, C. M. (2014). *The Continued Influence Effect: The Persistence of Misinformation in Memory and Reasoning Following Correction*. (J. L. G. Rapp, David N., Braasch, Ed.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences*. Cambridge, Massachusetts: MIT Press.

Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., Cenac, Z., … Moore, C. (2013). Priming Intelligent Behavior: An Elusive Phenomenon. *PLoS ONE*, *8*(4), e56515. http://doi.org/10.1371/journal.pone.0056515

Shanks, D., Vadillo, M., Riedel, B., Clymo, A., Govind, S., Hickin, N., … Puhlmann, L. (2015). Romance, risk, and replication: Can consumer choices and risk-taking be primed by mating motives? *Journal of Experimental Psychology: General*, *144*(6), 142–158. Retrieved from http://psycnet.apa.org/record/2015-48744-001

Shapiro, L. R. (2006). The effects of question type and eyewitness temperament on accuracy and quantity of recall for a simulated misdemeanor crime. *Emporia State Research Studies*, *43*(1), 1–7.

Simcox, T., & Fiez, J. A. (2014). Collecting response times using amazon mechanical turk and adobe flash. *Behavior Research Methods*, *46*(1), 95–111.

Singer, M., Graesser, A. C., & Trabasso, T. (1994). Minimal or global inference during reading. *Journal of Memory and Language*, *33*(4), 421–441.

Skolnick, P., & Shaw, J. I. (2001). A Comparison of Eyewitness and Physical Evidence on Mock-Juror Decision Making. *Criminal Justice and Behavior*, *28*(5), 614–630.

Sloman, S. A., & Lagnado, D. (2015). Causality in thought. Annual Review of Psychology, 66, 223-247.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*(2), 108–131.

Smith, M. J., Ellenberg, S. S., Bell, L. M., & Rubin, D. M. (2008). Media coverage of the measles-mumps-rubella vaccine and autism controversy and its relationship to MMR immunization rates in the United States. *Pediatrics*, *121*(4), 836–843.

Sommers, S. R., & Kassin, S. M. (2001). On the many impacts of inadmissible testimony: Selective compliance, need for cognition, and the overcorrection bias. *Personality and Social Psychology Bulletin*, *27*(10), 1368–1377.

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. http://doi.org/10.1007/s10979-006-9039-7

Stewart, N., Chandler, J., & Paolacci, G. (2017). Crowdsourcing Samples in Cognitive Science. *Trends in Cognitive Sciences*, *21*(10), 736–748. http://doi.org/10.1016/j.tics.2017.06.007

Sweney, M. (2017). Ofcom condemns Channel 4 News for naming wrong man as Westminster attacker | Media | The Guardian. Retrieved August 18, 2018, from https://www.theguardian.com/uk-news/2017/sep/11/channel-4-news-condemned-by-ofcom-for-westminster-attack-blunder

Tenney, E. R., Cleary, H. M. D., & Spellman, B. A. (2009). Unpacking the doubt in “beyond a reasonable doubt”: Plausible alternative stories increase not guilty verdicts. *Basic and Applied Social Psychology*, *31*(1), 1–8. http://doi.org/10.1080/01973530802659687

Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility: Research report. *Psychological Science*, *18*(1), 46–50. http://doi.org/10.1111/j.1467-9280.2007.01847.x

Tenney, E. R., Spellman, B. A., & MacCoun, R. J. (2008). The benefits of knowing what you know (and what you don’t): How calibration affects credibility. *Journal of Experimental Social Psychology*, *44*(5), 1368–1375. http://doi.org/10.1016/j.jesp.2008.04.006

Tetlock, P. E., Lerner, J. S., & Boettger, R. (1996). The dilution effect:

judgmental bias, conversational convention, or a bit of both? *European Journal of Social Psychology*, *26*(6), 915–934.

Thompson, W. C., Fong, G. T., & Rosenhan, D. L. (1981). Inadmissible evidence and juror verdicts. *Journal of Personality and Social Psychology*, *40*(3), 453–463. http://doi.org/10.1037/0022-3514.40.3.453

Thorson, E. (2016). *Belief echoes: The persistent effects of corrected misinformation*. *Political Communication*. http://doi.org/10.1080/10584609.2015.1102187

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275–1289. http://doi.org/10.3758/s13421-011-0104-1

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*(5), 612–630. http://doi.org/10.1016/0749-596X(85)90049-X

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne*, *26*(1), 1-12.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. *Progress in Social Psychology*, *1*, 49–72.

Van Boekel, M., Lassonde, K. A., O'Brien, E. J., & Kendeou, P. (2017). Source credibility and the processing of refutation texts. *Memory & Cognition*, *45*(1), 168–181.

Wais, P. E., Mickes, L., & Wixted, J. T. (2008). Remember/know judgments probe degrees of recollection. *Journal of Cognitive Neuroscience*, *20*(3), 400–405. http://doi.org/10.1162/jocn.2008.20041

Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, 1–19.

Wegner, D. M., Wenzlaff, R., Kerker, R. M., & Beattie, A. E. (1981). Incrimination through innuendo: Can media questions become public answers? *Journal of Personality and Social Psychology*, *40*(5), 822–832. http://doi.org/10.1037/0022-3514.40.5.822

Weinberg, H. I., & Baron, R. S. (1982). The discredible eyewitness. *Personality and Social Psychology Bulletin*, *8*(1), 60–67.

Whitley Jr, B. E. (1987). The effects of discredited eyewitness testimony: A meta-analysis. *The Journal of Social Psychology*, *127*(2), 209–214.

Wilkes, A. L., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *The Quarterly Journal of Experimental Psychology Section A*, *40*(2), 361–387. http://doi.org/10.1080/02724988843000168

Wilkes, A. L., & Reynolds, D. J. (1999). On certain limitations accompanying readers' interpretations of corrections in episodic text. *The Quarterly Journal of Experimental Psychology: Section A*, *52*(1), 165–183.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection

interpretation of remember/know judgments. *Psychonomic Bulletin & Review*, *11*(4), 616–641. http://doi.org/10.3758/BF03196616

Wolfe, C. R. (2017). Twenty years of Internet-based research at SCiP: A discussion of surviving concepts and new methodologies. *Behavior Research Methods*, 1615–1620. http://doi.org/10.3758/s13428-017-0858-x

Woodward, A. E., & Bjork, R. A. (1971). Forgetting and remembering in free recall: Intentional and unintentional. *Journal of Experimental Psychology*, *89*(1),109–116. Retrieved from https://pdfs.semanticscholar.org/87bb/60ecc4c6cdb3781c995ec7192e141caf78be.pdf

World Economic Forum. (2013). *Global Risks 2013 - Reports - World Economic Forum*.

Wyer, R. S., & Budesheim, T. L. (1987). Person memory and judgments: The impact of information that one is told to disregard. *Journal of Personality and Social Psychology*, *53*(1), 14.

Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language*, *46*(3), 441–517. http://doi.org/10.1006/jmla.2002.2864

Yonelinas, A. P., Aly, M., Wang, W.-C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus*, *20*(11), 1178–1194.

Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., … Dong, Q. (2010). Individual differences in false memory from misinformation: Cognitive factors. *MEMORY*, *18*(5), 543–555. http://doi.org/10.1080/09658211.2010.487051org/10.1080/09658211.2010.487051

Zwaan, R. a., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. Journal of experimental psychology: Learning, memory, and cognition, 21(2), 386.

Zwaan, R. A., Pecher, D., Paolacci, G., Bouwmeester, S., Verkoeijen, P., Dijkstra, K., & Zeelenberg, R. (2017). Participant nonnaiveté and the reproducibility of cognitive psychology. *Psychonomic Bulletin & Review*, 1-5. http://doi.org/10.3758/s13423-017-1348-y

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162-185.

Appendices

Appendix A

Table A.1. Experimental Stimuli used in Experiments 1A, 1B, 2A, and 2B

Message #	Content
Message 1	Jan. 25th 8:58 p.m. Alarm call received from premises of a wholesale stationery warehouse. Premises consist of offices, display room, and storage hall.
Message 2	A serious fire was reported in the storage hall, already out of control and requiring instant response. Fire engine dispatched at 9:00 p.m.
Message 3	The alarm was raised by the night security guard, who had smelled smoke and gone to investigate.
Message 4	Jan. 26th 4:00 a.m. Attending fire captain suggests that the fire was started by a short circuit in the wiring of a closet off the main storage hall. Police now investigating.
Message 5	The fire officer had recorded several fire code violations on the premises at a surprise inspection two months earlier.
Message 6 [**Target (Mis)information**]	4:30 a.m. Message received from Police Investigator Lucas saying that they have reports that cans of oil paint and pressurized gas cylinders had been present in the closet before the fire.
Message 6 [**Control – No Misinformation**][41]	4:30 a.m. Message received from Police Investigator Lucas saying that they have that they have urged local residents to keep their windows and doors shut.
Message 7	The display room was reported to contain display cases, catalogues, and the sales staffs' desks. It was only staffed from 11 a.m. to 2 p.m., due to diminishing sales.
Message 8	Firefighters attending the scene report thick, oily smoke and sheets of flames hampering their efforts, and an intense heat that made the fire particularly difficult to bring under control.
Message 9	It has been learned that a number of explosions occurred during the blaze, which endangered firefighters in the vicinity. No fatalities were reported.

[41] This condition only appeared in Experiments 2A and 2B.

Message 10	Two firefighters are reported to have been taken to the hospital as a result of breathing toxic fumes that built up in the area in which they were working.
Message 11	A small fire had been discovered on the same premises, six months previously. It had been successfully tackled by the workers themselves.
Message 12	10:00 a.m. The owner of the affected premises estimates that total damage will amount to hundreds of thousands of dollars, although the premises were insured.
Message 13 [No Correction]	10:40 a.m. A second message received from Police Investigator Lucas regarding the investigation into the fire. It stated that the two firefighters taken to the hospital had been released.
Message 13 [Correction & Alternative Explanation Conditions]	10:40 a.m. A second message received from Police Investigator Lucas regarding the investigation into the fire. It stated that the closet reportedly containing cans of oil paint and gas cylinders had actually been empty before the fire.
Message 13 [No Misinformation/Correction Condition][42]	10:40 a.m. A second message received from Police Investigator Lucas regarding the investigation into the fire. It stated that a closet reportedly containing cans of oil paint and gas cylinders had actually been empty before the fire.
Message 14 [No Correction/Correction/No Misinformation]	The shipping supervisor has disclosed that the storage hall contained bales of paper; mailing and legal-size envelopes; scissors, pencils, and other school supplies; and a large number of photocopying machines.
Message 14 [Alternative Explanation][43]	11:08 a.m. Firefighters have found evidence of gasoline-soaked rags near where the bales of paper had been stored in the storage hall, as well as several emptied steel drums of suspicious nature. The owner denies any knowledge of these materials.
Message 15	11:30 a.m. Attending fire captain reports that the fire is now out and that the storage hall has been completely gutted.

[42] This condition only appeared in Experiments 2A and 2B.

[43] This condition only appeared in Experiments 1A and 1B.

Appendix B

Questions and used in Experiments 1A, 1B, 2A, and 2B and response options used in Experiments 1A and 2B

Inference Questions

1. Why did the fire spread so quickly?
    a. Burning paint may have spilled over a large area
    b. Flammable materials could have been deliberately soaked in gasoline
    c. There could have been large amounts of paper throughout the building
    d. The kitchen door may have been left open

2. What was the possible cause of the fumes?
    a. Oil-based paint
    b. Gasoline
    c. Paper and cardboard
    d. Cooking oil

3. What aspect of the fire might the police want to continue investigating?
    a. Dangerously flammable materials were stored carelessly
    b. The presence of items of a suspicious nature
    c. Unaddressed fire code violations
    d. Fire not adequately prevented by open fire door

4. What could have caused the explosions?
    a. Fire came in contact with compressed gas cylinders
    b. Steel drums filled with liquid accelerants
    c. Volatile compounds in photocopiers caught on fire
    d. Cooking equipment caught on fire

5. Where was the probable location of the explosions?
    a. The storage closet.
    b. The storage hall
    c. The display room.
    d. The kitchen

6. What was the most likely overall cause of the fire?
    a. Flammable liquids and gases not stored properly
    b. Someone deliberately set fire to the property
    c. The owner had allowed paper and cardboard to be left lying around
    d. The cooker in the kitchen was left on

Factual Questions

1. Where on the premises was the fire located?
    a. In a closet off the main storage hall
    b. In the storage hall
    c. In the owner's office
    d. In a supply room next to the storage hall

2. What features of the fire were noted by the security guard?
    a. The smell of smoke
    b. The smell of gasoline
    c. The triggering of the alarm system
    d. The sight of flames through the window

3. What business was the firm in?
    a. Wholesale stationery
    b. Toy manufacturer
    c. Electrical supplies producer
    d. Book printing services

4. What was present in the closet before the fire?
    a. Cans of oil paint and pressurised gas cylinders
    b. The storage closet was empty before the fire
    c. Printer cartridges and toners
    d. The worker's uniforms

5. What was the cost of the damage done?
    a. Hundreds of thousands of dollars
    b. Millions of dollars
    c. Hundreds of dollars
    d. Tens of thousands of dollars

6. When was the fire eventually put out?
    a. 11.30 a.m.
    b. 11.08 a.m.
    c. 6.30 a.m.
    d. 12.00 p.m.

Appendix C

Table C.1. Coding criteria for warehouse fire story open-ended inference questions in Experiments 1A, 1B, 2A, and 2B

No.	Question	Example response to receive score of 1 on reference to target (mis)information measure
1	Why did the fire spread so quickly?	The fire spread due to oil in storage.
2	What was the possible cause of the fumes?	Oil cans and gas explosions
3	What aspect of the fire might the police want to continue investigating?	Why the cylinders were there
4	What could have caused the explosions?	Pressurised containers of aerosols.
5	Where was the probable location of the explosions?	In the closet and possibly in the offices
6	What was the most likely overall cause of the fire?	Carelessness with stocking flammable liquids and paper near an electrical supply.

Table C.2. Coding criteria for factual recall questions in 1A, 1B, 2A, and 2B

No.	Question	Correct Answer
1	Where on the premises was the fire located?	In a closet off the main storage hall
2	What features of the fire were noted by the security guard?	The smell of smoke
3	What business was the firm in?	Wholesale stationery
4	What was present in the closet before the fire?	Cans of oil paint and pressurised gas cylinders [No Correction Condition].
		The storage closet was empty before the fire. [Correction and No Misinformation/Correction Condition]
5	What was the cost of the damage done?	Hundreds of thousands of dollars

6	When was the fire eventually put out?	11.30 a.m.

Table C.3. Coding criteria for critical information recall questions 1A, 1B, 2A, and 2B

Question	**Example response to receive score of 1**
What was the point of the second message from Fire Chief Lucas?	Yes, the news reports were unclear about whether or not there were inflammable substances.
Were you aware of any corrections or contradictions in the messages that you read?	Yes, I was aware of the correction by the officer.

Appendix D

In order to make comparisons between conditions, responses to the question probing recall of critical information that appeared at Message 13 (i.e., either a correction or control message) were analysed. This analysis differed from the preregistered confirmatory analysis. The second question was only relevant to the conditions featuring initial misinformation and its correction so was not analysed. Chi-square tests tested dependence between correction information condition and recall of critical information.

Effect of correction information condition on critical information recall responses (Experiment 1A)

Relative frequencies did not significantly differ, χ^2 (2) = 3.12, p = .21. Accurate recall of critical information occurred at rate of 50% for the no correction group, 48% for the correction group, and 28% for the alternative explanation group.

Effect of correction information condition on critical information recall responses (Experiment 1B)

Relative frequencies did not significantly differ, χ^2 (2) = 0.67, p = .72. Accurate recall of critical information occurred at rate of 33% for the no correction group, 36% for the correction group, and 25% for the alternative explanation group.

Effect of correction information condition on critical information recall responses (Experiment 2A)

Relative frequencies were significantly different, χ^2 (2) = 13.73, p = .001. The no correction group recalled critical information at a rate of 56%, the correction group accurately recalled critical information at a rate of 42% and the no misinformation group at a rate of 21%.

Effect of correction information condition on critical information recall responses (Experiment 2B)

Relative critical information recall frequencies were significantly different, χ^2 (2) = 21.09, p < .001. The no correction group recalled critical information at a rate of 50%, the correction group accurately recalled critical information at a rate of 66% and the no misinformation group at a rate of 22%.

Appendix E

Table E.1. Experimental Stimuli used in Experiment 3

Message #	Content
Message 1	Large blaze reported at wholesale stationery warehouse in Fern Hill industrial park. The fire broke out around 9pm.
Message 2	The alarm was first raised by night security guard who smelled smoke and went to investigate.
Message 3	More than 60 firefighters are battling to contain the huge blaze. Workers at nearby warehouses are being evacuated.
Message 4	Three warehouse workers, suffering from smoke inhalation, have been taken to St Columbus Hospital.
Message 5 [Target (Mis)information; All Conditions]	Fire Chief Lucas issues statement: “Cans of oil paint and pressurized gas cylinders were present in the storeroom before the fire.”
Message 6	Witness Greg Burns said “A large number of emergency services arrived very quickly, so it was clearly a major fire.”
Message 7	The fire officer reported several fire code violations had been recorded on the premises at surprise inspection two months earlier.
Message 8 [Causal Detail]	Thick, oily smoke & sheets of flames hinder fire-fighters efforts, intense heat has made the fire difficult to bring under control.
Message 9	Firefighters are using an aerial ladder platform in their attempts to extinguish the flames.
Message 10 [Control - No Correction]	Update from Fire Chief Lucas: “The warehouse employees taken to hospital have been released.”
Message 10 [Just Correction]	Correction from Fire Chief Lucas: “No cans of oil paint and pressurized gas cylinders had ever been present in the warehouse.”
Message 10 [Correction + Error Explanation]	Correction from Fire Chief Lucas: “An employee confused soda-stream canisters & coffee cans in the storeroom, for paint & gas cylinders”

Message 10 [Correction + Lie Explanation]	Correction from Fire Chief Lucas: “An employee admitted to lying about presence of paint and gas in the storeroom.”
Message 11	The fire was finally brought under control around 12pm.
Message 12	An attending fire captain reports that the fire is now out and that the storage hall has been completely gutted.

Appendix F

Table F.1. Inference questions and response coding criteria in Experiment 3

No.	Question	Example response to receive score of 1 on false information measure
1	What evidence of careless management is there in relation to this fire?	The rumors of dangerous materials could be evidence of careless management.
2	How could the fire at the warehouse have been avoided?	The fire could have been avoided if the surprise inspection had led to not storing the oil paint and the pressurized gas cylinders together in a dangerous manner.
3	What precautions could be taken in the future to ensure this doesn't happen again?	Maybe don't use oil paint. And separate the paint from the pressurized gas cylinders.
4	What aspect of the fire should the police focus on in their investigation?	I think they should focus if there were any hazardous material stored in the warehouse.
5	Does any aspect of the fire deserve further investigation?	Presence of flammable material and root cause of the fire
6	Do you think any workers should be disciplined for their role in the fire?	If a manager or superior instructed anyone to store flammable materials in a way that caused the blaze they should be disciplined.
7	What was the most likely cause of the fire? [cause question]	The cause of the fire was likely the oil and pressurized cans in the storage.

Table F.2. Coding criteria for factual recall questions in Experiment 3

Question	Example response to receive score of 1
What was the point of the second message from Fire Chief Lucas?	Yes, the news reports were unclear about whether or not there were inflammable substances.
Were you aware of any corrections or contradictions in the messages that you read?	Yes, I was aware of the correction by the officer.

Table F.3. Coding criteria for critical information recall questions

No.	Question	Correct Answer
1	Where was the warehouse located?	Fern Hill Industrial Park
2	What features of the fire were noted by the security guard	The smell of smoke
3	Approximately how many firefighters battled to contain the fire?	60 firefighters
4	Which hospital were the workers taken to?	St Columbus
5	What was present in the storeroom before the fire?	[Condition dependent: No correction = Cans of oil paint and gas cylinders, Correction / Correction + Lie Explanation = Nothing / Do not know, Correction + Error Explanation = Soda-canisters and coffee cans
6	What did firefighters use to try and extinguish the flames?	Aerial ladder platform.
7	At what time was the fire eventually brought under control?	4am.

Appendix G

Instructions presented before answering inference and fact questions in Experiment 3

The same instructions were used in Experiments 4, 5, 6, and 7 except that the

> You will now see 16 questions about the incident you just read about. You will not necessarily be able to answer all the questions directly from the information you have just seen. You may need to make a judgment about what you think happened, as well as use your wider knowledge that is relevant to the questions here.
>
> Please answer each of the following questions on the basis of your understanding of the reports and your understanding about industrial fires in general, writing your answer in the text box provided and giving as much detail as necessary.
>
> As a guide, you need to type at least 25 characters to proceed to the next question. It may be possible to respond to some questions with single word answers - please answer these questions using full sentences.

word was changed to reflect the particular scenario.

Appendix H

Table H.1. Warehouse Fire Stimuli used in Experiments 4 and 5

Message #	Content
Message 1	Large blaze reported at wholesale stationery warehouse in Fern Hill industrial park. The fire broke out around 9pm.
Message 2	Night security guard, who smelled smoke and went to investigate, first raised the alarm.
Message 3	More than 60 firefighters are battling to contain the huge blaze. Fire dept. investigators trying to establish cause of the blaze. Nearby residential building evacuated over fears of damage due to fire.
Message 4 [Potential Cause Information]	Recent report from fire department indicates most industrial fires are due to equipment & machinery, flammable substances, hot work, & electrical hazards.
Message 5 [Target (Mis)information]	Fire Chief Lucas issues statement: "Cans of oil paint and pressurized gas cylinders were present in storeroom before fire."
Message 6	Three warehouse workers working overtime have been taken to St Columbus Hospital, due to smoke inhalation.
Message 7	Warehouse fire safety officer reports surprise inspection at the premises two months earlier. Full report has not been published yet.
Message 8 [Causal Detail]	Thick, oily smoke + sheets of flames hinder firefighters' efforts; intense heat has made the fire difficult to bring under control.
Message 9	Firefighters have been using an aerial ladder platform in their attempts to extinguish the flames. The owner is concerned about damage to stock. Shocked neighbours posted videos

	online.
Message 10 [No Correction]	Update from Chief Lucas: “The warehouse employees taken to hospital were treated for smoke inhalation and have now been released. Temporary accommodation is available for evacuated residents.”
Message 10 [Correction]	Correction from Chief Lucas: “No flammable items actually in storeroom. No paint or gas had ever been present in the warehouse. We apologise for the earlier error.”
Message 10 [Correction + Error Explanation]	Correction from Chief Lucas: “No flammable items actually in storeroom. No paint or gas had ever been present in the warehouse. An employee confused soda canisters and coffee cans for paint and gas.”
Message 10 [Correction + Lie Explanation]	Correction from Chief Lucas: “No flammable items actually in storeroom. No paint or gas had ever been present in the warehouse. Unhappy employee admitted lying about presence of paint & gas in storeroom.”
Message 11	The fire was finally brought under control around 4am early the following morning. A couple of firefighters were seen high-fiving each other.
Message 12	Warehouse fire is now out and the storage hall has been completely gutted. Owner expects substantial fire damage costs.

Table H.2. Crash stimuli used in Experiments 4 and 5

5 Message #	6 Content
Message 1	7 Serious accident involving van reported on Spring St. around 4pm today. Van was carrying 12 people, including the driver.

Message 2	8 Passing driver reported the accident after noticing the van had crashed into a steep embankment & rolled on its side.
Message 3	9 A rescue crew was immediately dispatched to the scene upon report of the accident, arriving at the scene within 10 minutes. Police are interviewing those involved in the crash.
Message 4 [Potential Cause Information]	10 Road safety experts say that vehicle type, driver behaviour, road and environmental conditions can all cause a vehicle to roll over.
Message 5 [Target (Mis)information]	11 Chief Inspector Brown reports: "Driver drank at least one bottle of beer during a stop at a service station."
Message 6	12 Serious damage caused to side of the van. Three passengers incurred injuries and have been airlifted to hospital.
Message 7	13 The charter van company released a statement that the vehicle had passed a recent inspection with minor faults.
Message 8 [Causal Detail]	14 The driver of the van, a recent divorcee, had been involved in a prolonged legal battle with his ex-wife.
Message 9	15 Rescue workers are using special cutting equipment to free two of the passengers. Passengers who have been freed from the van appear visibly distressed.
Message 10 [No Correction]	16 Update from Insp. Brown: "The two passengers airlifted to hospital have been stabilized and will be kept in for observation. Traffic through Spring St has been temporarily diverted."
Message 10 [Correction]	17 Clarification from Insp. Brown: "Driver did not drink beer at service station. Tests show he had no alcohol in his system. We apologise for the earlier inaccuracy."
Message 10 [Correction + Error Explanation]	18 Clarification from Insp. Brown: "Driver did not drink beer at service station. Tests show he had no alcohol in his system. The bottle he was seen drinking actually contained non-alcoholic ginger beer."

Message 10 [Correction + Lie Explanation]	19 Clarification from Insp. Brown: “Driver was not drinking alcohol. Tests show driver had no alcohol in his system. Passenger who made allegation admitted lying because of earlier argument with driver.”
Message 11	Van was transporting passengers back home from the Beat Bunker music festival when the crash occurred. 20
Message 12	21 Passengers have now been discharged from hospital. Police will continue investigating the accident.

Table H.3. Injury Stimuli used in Experiment 4

Message #	22 Content
Message 1	23 At about 8.30pm last night officers responded to call regarding an injured woman lying in a street in downtown San Luis.
Message 2	24 A resident heard shouts and looked out his window to see a woman collapsed; emergency services were called immediately.
Message 3	25 Officers arriving on the scene, found the middle-aged woman, unresponsive, with a head injury. Police are working to determine to circumstances surrounding her injuries.
Message 4 [Potential Cause Information]	26 Officer says that common reasons for head injuries are blows to the head sustained by falling, physical assault, or motor vehicle accidents.
Message 5 [Target (Mis) information]	Detective Symons makes statement: "Cries were heard and a man was seen running away from the scene." 27
Message 6	28 Paramedics treated the injured woman on the scene. She was then rushed to hospital where she received further treatment.
Message 7	29 Police have been examining CCTV camera footage within a mile radius of the incident. Obstructions mean footage is inconclusive.
Message 8 [Causal Detail]	An initial medical report suggests head injuries are consistent with impact from a blow to the head. 30
Message 9	31 The area on Maddox St where the incident occurred has been cordoned off whilst the police continue their investigation. Police have appealed for witnesses or anyone with information to come forward.
Message 10 [No Correction]	32 Update from Det. Symons: "Injured woman has been identified and we have been in contact with her family. Please respect cordon boundaries while investigation in continuing."

Message 10 [Correction]	33 Det. Symons revises earlier statement: “Man seen running away not involved in incident. Injuries could not have come from physical assault. We are sorry for our earlier mistake.”
Message 10 [Correction + Error Explanation]	Det. Symons revises earlier statement: “Man seen running away not involved in incident. Injuries could not have come from physical assault. Man was in fact running to call an ambulance.”
	34
Message 10 [Correction + Lie Explanation]	35 Det. Symons revises earlier statement: “Man seen running away not involved in incident. Injuries could not have come from physical assault. Notorious attention seeker lied about seeing man running away.”
Message 11	36 Injured woman is believed to be a housekeeper who had been working in the area, and was returning to her car parked on the street.
Message 12	Injured woman has been stabilized but has a fractured skull. She is not yet in a fit state to be interviewed.
	37

Table H.4. Missing person stimuli used in Experiment 4

Message #	38 Content
Message 1	39 Edge Park police are looking for 19-year old, Joe Pryce, missing since Wednesday morning.
Message 2	40 Joe missed work and failed to pick up his girlfriend from her soccer practice, after which his parents reported him missing.
Message 3	41 Police sent out a search team after completion of a risk assessment. Police are interviewing Joe's colleagues at Butler's pharmacy.
Message 4 [Potential Cause Information]	42 Edge Park has seen several disappearances in recent years; reasons range from financial difficulties to adventure hiking, and mental health issues.
Message 5 [Target (Mis) information]	43 Lieutenant Lopez releases report: "Joe's car was seen leaving town through toll road."
Message 6	44 Donna Pryce, Joe's mother, heard Joe leave for work early that morning but didn't notice anything unusual.
Message 7	45 Joe's parents first used the 'Find My Friends' app to check his last location. He last checked in at work on Tuesday PM.
Message 8 [Causal Detail]	46 Joe has been known to frequent Lake Fairmount and other off-road locations, from time to time.
Message 9	47 Police are also checking for any activity on Joe's phone, bank and social media accounts. Joe's family and friends have launched social media appeal for help locating him.
Message 10 [No Correction]	48 Update from Lt. Lopez: "K9 and helicopter search and rescue teams have been deployed. We ask anyone independently searching to allow our personnel to conduct their search efforts."
Message 10 [Correction]	49 Lt. Lopez withdraws initial report: "Joe's car did not leave town: It had been in auto repair shop since before he went missing. We regret our earlier error."

Message 10 [Correction + Error Explanation]	50 Lt. Lopez withdraws initial report: "Joe's car did not leave town: It had been in auto repair shop since before he went missing. Car color was misidentified in low light."
Message 10 [Correction + Lie Explanation]	51 Lt. Lopez withdraws initial report: "Joe's car did not leave town: It had been in auto repair shop since before he went missing. Alleged witness admitted lying to police in hope of getting reward."
Message 11	52 Joe is described as white male, 5 '11''with brown hair & blue eyes. He was wearing black jacket and blue jeans, when he was last seen.
Message 12	53 Joe's parents are holding a press conference. Police spokesperson thanked media, and members of public for assistance.

Appendix I

Coding criteria for the warehouse fire scenario questions are identical to Experiment 3 and are therefore not reported again here.

Table I.1. Inference questions and response coding criteria for crash scenario in Experiments 4 and 5

No.	Question	Example response to receive score of 1 on target (mis)information measure
1	What evidence is there of negligent driving in relation to this accident?	The driver had been drinking alcohol whilst driving.
2	How could this accident have been avoided?	Probably if the driver had not been drinking alcohol whilst driving.
3	Were any of the people in the vehicle particularly responsible for the crash?	Probably the driver since he was drinking during the service stop.
4	What measures could the charter van company take prevent future accidents?	Make sure they vet their drivers better so they don't hire drunks.
5	What aspects of the accident should further investigations be focused on?	Finding out whether there was any alcohol on the bus.
6	For what reasons might the passengers want to take legal action against the charter van company?	Because the driver had been drinking and he should not have been.
7	What do you think the most likely cause of the crash was? [cause question]	Careless driving because the driver was drunk.

Table I.2. Factual recall questions and response coding for crash scenario in Experiments 4 and 5

No.	Question	Correct Answer
1	How many people was the van carrying?	Twelve
2	Where did the accident occur?	Spring Street
3	What did the van crash into?	A steep embankment
4	What method of transport was used to take three of the passengers to the hospital?	They were airlifted
5	What was the driver drinking during the service stop?	[Condition dependent] No correction =Alcohol, Explanatory Correction (Error) = Non-alcoholic beer, Correction only / Explanatory correction (Lie) = Unknown / Not alcohol
6	What event was the van transporting people from?	Beat Bunker
7	What was the van driver's marital status?	Recent divorcee

Table H.3. Critical information recall questions and response coding criteria for crash scenario in Experiments 4 and 5

No.	Question	Example response to receive score of 1
1	What was the purpose of the second statement from Inspector Brown?	To tell everyone that the driver was not drinking alcohol.
2	Were you aware of any modifications or amendments to the messages you read?	Yes, the passenger's statement was found not to be true because there was no alcohol in the bottle.

Table I.4. Inference questions and response coding criteria for injury scenario used in Experiment 4

No.	Question	Example response to receive score of 1 on target (mis)information measure
1	How might the CCTV camera footage help the police in their investigation?	They would have evidence that the man seen running away assaulted the man
2	Who, if anyone, do you think is responsible for the woman's injuries?	The man seen running away hit him [with something], then hid it in his house.
3	Why do you think the woman was unresponsive?	She was knocked unconscious by the woman seen running away
4	What do you think is a likely explanation for what happened to the injured woman?	She was assaulted by the man who was seen running away
5	What potential leads do the police have in establishing what happened to the woman?	Well someone ran away from the woman – suggests he was probably assaulted
6	Why do you think the woman was unable to recall what happened to her?	Because the man seen running away hit her around the head with something or punched her.
7	What was the most likely cause of the woman's injuries? [cause question]	She was assaulted by someone.

Table I.5. Factual recall questions and response coding criteria for injury scenario used in Experiment 4

No.	Question	Correct Answer
1	What time did officers initially respond to the call?	Around 8.30pm
2	Roughly how old was the injured woman?	Middle-aged
3	What did the injured woman receive treatment for?	A head injury / fractured skull
4	What was the name of the street the police cordoned off?	Maddox Street
5	How were the woman's injuries sustained?	[Condition dependent: No correction = Assault, Correction conditions = Unclear / Not assault / A fall]
6	What was the injured woman's profession?	Housekeeper
7	What head injury did the woman sustain?	Fractured skull

Table I.6. Critical information recall questions and response coding for injury scenario used in Experiment 4

Question	Example response to receive score of 1
What was the implication of Detective Symons second report?	Evidence was inconsistent with assault
What facts about the incident did the police change their minds about, based on information they discovered later?	That the woman was assaulted.

Table I.7. Inference questions and response coding criteria for missing scenario used in Experiment 4

No.	Question	Example response to receive score of 1 on target (mis)information measure
1	What reasons are there for the police to be concerned about Joe's disappearance?	Yes, someone saw his car leaving town, and he likes to do off-road hiking so he might have had an accident.
2	What leads might the police have in locating Joe?	Someone spotted his car leaving town
3	Where do you think Joe is?	He went off somewhere in his car – probably went hiking or swimming in the lake
4	Precisely how might the traffic camera footage relate to Joe's disappearance?	The footage showed that Joe left Edge Park and went somewhere out of town
5	What do you think the risk assessment conducted by the police might have shown?	That Joe was experiencing some difficulties and may have driven off somewhere
6	Which aspects of Joe's disappearance do you believe deserve further investigation?	Probably the fact that his car was seen leaving town. They should cast their search net wider.
7	What do you think the most likely reason for Joe's disappearance is? [cause question]	He drove off somewhere in his car and got injured or something.

Table I.8. Factual recall questions and response coding criteria for missing person scenario used in Experiment 4

No.	Question	Correction Answer
1	How old is Joe?	19
2	Where does Joe work?	Butler's pharmacy
3	What was the name of Joe's mother?	Donna Pryce
4	What form of technology did Joe's parents use to find out his last location?	'Find My Friends' App
5	What evidence did the police find in relation to the disappearance?	[Condition dependent] No correction = Someone saw his car leaving town, All correction conditions = Nothing
6	Where did Joe's family and friends launch an appeal for help locating him?	Social media
7	What was Joe wearing when he was last seen?	Black jacket and blue jeans

Table I.8. Critical information recall questions and response coding criteria for missing person report used in Experiment 4

No.	Question	Example response to receive score of 1
1	What details did Lt. Lopez' second report provide?	Lopez said that the cameras showed the witness was wrong about Joe's car
2	Did you notice any inconsistencies between the messages that you read?	There was a witness who apparently saw Joe's car but then it turned out they were wrong.

Appendix J

Experiment 6 used the warehouse fire and crash scenarios. These were almost identical to previous experiments except that two different types of misinformation were presented to different groups of participants. Only the messages from the new misinformation conditions are reported here to avoid repetition.

Table J.1. Explicitly stated target (mis)information for warehouse fire and crash scenarios in Experiment 6

Warehouse Fire	Crash
Fire Chief Lucas issues statement: "Investigation team suspect fire caused by carelessly stored flammable liquids. Cans of oil paint and pressurized gas cylinders were present in storeroom before fire."	Chief Inspector Brown reports: "Investigatory team suspect drunk driver caused crash. Driver drank at least one bottle of beer during a stop at a service station."