WHY DOES MISINFORMATION PERSIST?
COGNITIVE EXPLANATIONS OF THE IMPLICIT MESSAGE EFFECT

By

Reed Miller Reynolds

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Communication—Doctor of Philosophy

2020

ABSTRACT

WHY DOES MISINFORMATION PERSIST?
COGNITIVE EXPLANATIONS OF THE IMPLICIT MESSAGE EFFECT

By

Reed Miller Reynolds

Recent controversies have emerged regarding false information in contemporary discourse. Research suggests that misinformation communicated implicitly is harder to correct than explicitly stated misinformation (the implicitness effect), but the mechanism has remained speculative. Prior research has proposed the failure to monitor (FTM) hypothesis, including the prediction that inadequate information retrieval may explain the implicitness effect. This study experimentally varied misinformation implicitness and correction strength, measuring outcomes including misinformation persistence (MP), attribution accessibility, and mental representations generated by participants. Results indicate the accessibility of misinformation-consistent attributions is associated with increased MP, but accessibility does not mediate the implicitness effect. In contrast, *misintegration*, a cognitive process that makes the misinformation consistent with corrections, moderates the implicitness effect. Analyses reveal several distinct mechanisms that predict misinformation persistence, including message characteristics, receiver ability to retrieve critical information, and the quality of receiver-generated inferences. Theoretical implications are discussed.

ACKNOWLEDGEMENTS

This project would not have been possible without guidance and inspiration from many people along the way. In particular, I would like to acknowledge my advisor, John Sherry, for his perspectives on communication research. He has showed me the importance of re-evaluating assumptions and being willing to take risks in research. I remember many discussions with him where seemingly simply concepts turned out to be more complex than they appeared. Gaining these insights takes people like John asking probing questions. I have also been lucky to benefit from John's humor and wit and extensive knowledge of alternative-music history—not to mention John's prowess as the best bass player of all mid-Michigan Arabic jazz bands. My relationship with John has been rewarding intellectually, creatively, and personally.

I would also like to acknowledge Frank Boster for his guidance on many fronts. I am constantly amazed by his commitment to mentorship, his ability to explain difficult topics, and his empathy. I remember Frank taking many opportunities to work through research issues with me and other students. He prioritized more than getting the best answer, he would teach the process of discovery to his students. Frank is one of those people I learn something new from literally every time I see them.

Mandy Holmstrom has also been hugely influential on my development as a scholar. Her lab was the first I joined at MSU and I worked with her the whole way through my program. She showed me the value of theoretical focus, to have projects build one upon another, to drill down further and further on a specific question. I am also grateful for her giving me the opportunity to lead projects and learn through trial-and-error.

So many people have contributed to my development as a scholar I struggle to articulate them all. Erik Altmann has been a great resource while I have worked through challenging topics in cognitive science. This dissertation began as a class project for him. His thoroughness and uncompromising nature have pushed me to improve my thinking and my methods. Likewise, all my research collaborators have helped me grow as a researcher and a writer, including Gary Bente, Jingbo Meng, Elizabeth Dorrance Hall, Daniel Bergan, David Ewoldsen, Nancy Rhodes, and Kenneth Levine. I would also like to acknowledge my master's adviser, Brian Spitzberg, who helped inspire me to pursue a doctorate.

I have also benefited tremendously from the support of my fellow students, support staff, and culture of productivity in the Department of Communication at MSU. In particular, Thomi Chrisinske and Marge Barkman have been beyond helpful; I would be lost without them trying to navigate the ins and outs of graduate school.

Finally, I want to thank my family for being supportive and interested in my academic pursuits. My mom, my dad, my brothers, and extended family have given me the inspiration and confidence to earn my doctorate. I remember my mom in particular peaking my interest in literature from an early age, and my dad cultivating extreme curiosity about the world and how it works. From an early age to the present moment, I feel fortunate to have opportunities. I love you all. Thank you!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Misinformation persistence (MP)[1] has received ample scholarly attention in recent years (Chan et al., 2017; Walter & Murphy, 2018; Walter & Tukachinsky, 2020). Defined as the resilience of false beliefs to corrective messages (Nyhan & Reifler, 2010), the phenomenon has produced controversies in contemporary discourse, generating concern about its ability to spread (Maheshwari, 2016), and impact vital issues at every social level (Ecker et al., 2010; Ecker et al., 2014; Lewandowsky et al., 2012). Researchers have identified several causes of MP, but some mechanisms remain unclear, particularly when biasing motivations do not explain the phenomenon; for example, research has shown that MP can occur when corrective messages are remembered and believed by receivers (Rich & Zaragoza, 2016).

This puzzle raises questions about the communicative and cognitive causes of misinformation persistence. Rich and Zaragoza (2016) found that message implicitness increases MP, specifically that mere suggestion is harder to correct than directly stated misinformation. To explain this effect, Rich and Zaragoza (2016) proposed the *failure to monitor inconsistencies* hypothesis (FTM), which states that people have limited awareness of contradictions among their own beliefs, and that people may accept new information without fully considering its implication or revising prior beliefs. Rich and Zaragoza (2016) argue that implicit messages make monitoring failure more likely, leading to greater MP. Direct evidence of the FTM hypothesis has not been established, however, and mechanisms of monitoring failure have yet to be explored.

This study proposes that belief accessibility (Roskos-Ewoldsen, 1997; Roskos-Ewoldsen & Fazio, 1997) mediates the implicitness effect, serving as a mechanism of monitoring failure. Specifically, we argue that implicit messages cause receivers to perform additional cognitive

[1] Also known as the continued influence effect.

work during message-processing that increases accessibility of misinformation-consistent information (i.e., making misinformation-consistent information more readily activated from memory). As a result, receivers judge misinformation as plausible because misinformation-consistent information is more accessible and more strongly influences judgments about the misinformation. This rationale draws from models of bounded rationality, models of human memory, and theories of attribution.

Other factors have also been implicated in the implicitness effect. For example, Reynolds (2018) explored the phenomenon of misintegration, a process by which receivers interpret misinformation as being consistent with its correction. Receivers may explain how the misinformation and correction can both be true. Misintegration may be an alternative mechanism of MP as it does not imply monitoring failure and it may increase MP. Reynolds (2018) found that implicitness only increased MP among participants did not misintegrate. In other words, the effect of message implicitness disappeared when people interpreted the correction to be consistent with the misinformation. The conditional effect of implicitness suggests that constructing mental representations (i.e., understandings of the situation; Fuster, 1997; Roskos-Ewoldsen et al., 2009) is an information-processing mechanism linked to MP. A secondary goal of this study is to replicate Reynolds' (2018) finding on the moderating role of misintegration. To that end, I first review basic features of human memory and then articulate the FTM hypothesis and the role of information accessibility. Next, I discuss the impact of misintegration in response to corrective messages and introduce the present experiment.

RATIONALE

Misinformation Persistence and Limited Cognitive Capacity

Bounded rationality perspectives (e.g., Simon, 1972; Gigerenzer & Goldstein, 1996) recognize that people have limited cognitive capacity and dedicate limited cognitive resources to a given task. These limitations are particularly relevant in the context of message processing (Lang, 2009). The failure to monitor (FTM) hypothesis states that MP is a product of this limited capacity, resulting from failures to retrieve information and/or evaluate the consistency of beliefs when making judgments (Rich & Zaragoza, 2016). Information accessibility plays a critical role, therefore, representing the ease with which information is retrieved from memory and made available for processing (Rholes & Pryor, 1982; Roskos-Ewoldsen, 1997).

Contemporary models of human memory suggest that information accessibility influences MP, and that implicit messages may increase accessibility of misinformation. Human memory is a system of interconnected information processing mechanisms (Baddeley, 2012). Working memory enables immediate attention to small amounts of information but its contents degrade rapidly after exposure if not encoded in long term memory or attended to repeatedly (Baddeley, 2003). Once stored in long term memory, information can be retrieved and activated in working memory. Here, the strength of activation corresponds with the information's prominence in conscious awareness (Anderson, 1983). The structure of long-term memory is typically modelled as a network, where nodes represent information, and links represent cognitive associations (e.g., Anderson, 1983; Nelson et al., 2013). Patterns of neural activation are responsible for mental representations, or the understanding a person has of an object, event, or situation (Fuster, 1997; Roskos-Ewoldsen et al., 2009). For example, the evaluation of an

object is strongly influenced by the cognitions that are associated with and come to mind at the time of evaluation (Roskos-Ewoldsen, 1997; Ewoldsen & Fazio, 1997).

The process of information retrieval and activation is governed by several features of human memory systems (Anderson, 1983; Ayers & Reder, 1998; Fuster, 1997). First, prior activation tends to increase information accessibility, especially when activation has been recent or frequent. Second, the structure of long-term memory influences accessibility such that activation more likely spreads to strongly associated information than weakly associated information. Third, the strength of association increases when information is activated simultaneously (Anderson, 1983), as when multiple features of a stimulus become associated with one another. In addition, information perceived as causally related tends to be strongly associated in memory (Trabasso & Van Den Broek, 1985).

In the context of MP, because activation spreads within memory networks as a function of network structure, and because of limited cognitive capacity, a correction may fail if it is not activated while a person is evaluating the misinformation. When corrections succeed, they are perceived as both true and inconsistent with the misinformation, therefore appealing to consistency motivations (Festinger & Carlsmith, 1959; Gawronski, 2012; McGuire, 1960). Receiving a correction requires updating prior beliefs, a process that involves activating connected memory networks, retrieving related beliefs, bringing them into working memory, and reinterpreting the situation and its associated cognitions (Fuster, 1997; Kendeou et al., 2013; Rich & Zaragoza, 2016). This process can be difficult and time-consuming due to attentional bottlenecks (despite potential parallel processing; Fuster, 1997; Reder, 1982). Coupled with other motivational, time, and cognitive constraints (e.g., Lang, 2009; Simon, 1972), these limitations may cause MP through a failure to monitor inconsistencies.

Activation, Sense-Making, and Implicit vs. Explicit Messages

If monitoring failure can cause MP, it remains to be shown how implicit misinformation may lead to greater monitoring failure. Social psychologists have long demonstrated that people try to infer the cause of observed behavior, a process known as attribution (Heider, 1958; Kelley & Michela, 1980; Weiner, 2008). People will generate a causal explanation that makes sense of events, drawing upon information from observations, messages, and prior beliefs such as expectations (Kelley & Michela, 1980). Explicit messages contain all needed causal information and do not require the additional processing that attribution entails. Because implicit messages withhold a causal explanation of a key narrative event,[2] they lead receivers to engage in greater mental elaboration to infer a cause and satisfy a basic sense-making threshold.

This attributional process has multiple consequences for accessibility of information, monitoring failure, and MP. First, generating attributions promotes attention to information that can help identify a cause (Berscheid et al., 1976), and attribution, as part of the process of message comprehension, occurs continuously as relevant information is presented (Kendeou et al., 2013). When misinformation is presented, the message is consistent with the misinformation and resulting attributions promote attention to misinformation-consistent details. Based on the models of memory previously discussed, the increased attention will tend to increase accessibility of this information (Roskos-Ewoldsen, 1997; Roskos-Ewoldsen & Fazio, 1997), therefore, promoting accessibility of misinformation-consistent information.

Implicitness should also increase accessibility of misinformation-consistent cognitions through co-activation of attributions and the misinformation. In implicit messages, attributions function to explain the misinformation and should be activated simultaneously with the

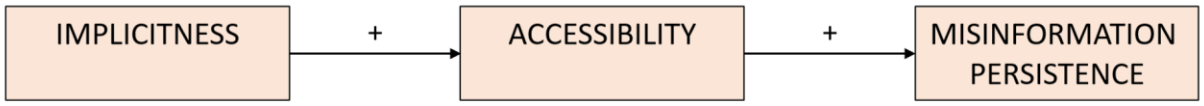
[2] Other research has looked at messages that omit the conclusion of an argument, rather than omitted explanatory information (e.g., Kardes, 1988).

misinformation. As Fuster (1997) describes, cognitions that are activated simultaneously tend to become associated. This means the attribution is likely linked in memory with the misinformation, making subsequent co-activation more likely. Empirical evidence has likewise shown that causally linked information tends to be strongly associated in memory networks and is more readily retrieved (Black & Bern, 1981; Keenan et al., 1984; Trabasso & Van Den Broek, 1985). For these reasons, implicit messages are expected to prompt an attributional process, creating interconnected attributions that are consistent with the misinformation and readily activated. Misinformation-consistent information should therefore be more accessible in response to implicit rather explicit messages.

The outcome of this attributional process may promote MP by making misinformation-consistent inferences more accessible in memory. As a result, highly accessible attributions should reduce the attention available for the correction when evaluating the misinformation. In summary, message implicitness should increase accessibility of misinformation-consistent information, meaning this information will more strongly influence evaluations of the misinformation, thus making it appear more plausible. See Figure 1 for a visualization of the proposed mediation model.

Figure 1.

Accessibility Model of Misinformation Persistence



Note. Accessibility refers to the ease of activating misinformation-consistent information.

In accord with these propositions, **[H1]** implicit misinformation should lead to greater accessibility of misinformation-consistent inferences than explicit misinformation, and that **[H2]** accessibility of misinformation-consistent beliefs will be positively associated with MP. In addition, **[H3]** a mediation model should fit the data such that differences in accessibility are a plausible mechanism of the effect of implicitness on MP.

Misintegration and Misinformation Persistence

As argued above, inconsistencies between information are not always recognized when making a judgment; however, recognizing (potential) inconsistency does not guarantee a receiver will reject inconsistent information. For example, a person may reassess the apparent inconsistency and explain how the beliefs are reconcilable. In other words, a receiver may accept both propositions without a failure to monitor. In so doing, a person will have integrated the two propositions into a unified account (for more on integration during comprehension, see Ecker et al., 2011; Kendeou et al., 2013). Integrating misinformation with a correction is a special case of integration which I label *misintegration* because it permits MP without ignoring or rejecting the correction (Reynolds, 2018).

Similar phenomena have been referenced in other literature. For example, Ecker et al. (2014) found that bigoted subjects were less prone to MP than non-bigoted subjects when corrections reinforced stereotypical beliefs. In the messages presented, an ethnic minority was portrayed to have heroically stopped a robbery attempt. Later, a correction stated the hero had, in fact, not belonged to the ethnic minority. One subject persisted in the belief about the hero's ethnicity, but supposed the hero had conspired with the robber. In that case, the respondent created a new interpretation of events that was consistent with their negative stereotype and supported the misinformation.

When misintegration occurs, it does not reflect a failure to monitor inconsistencies between correction and misinformation because the account represents them as being compatible. People who misintegrate should be able to resist correction because they have articulated how the correction does not conflict with the misinformation. For these reasons, I expect to replicate the finding of Reynolds (2018) that [**H4**] misintegration will be associated with greater MP. I also expect to replicate the finding of Reynolds (2018) that [**H5**] misinformation implicitness will increase MP only when participants do not misintegrate because receivers who integrate should be less sensitive to effects of the correction.

Corrective-Message Strength and Baseline Tendencies

MP does not merely imply that the misinformation is persuasive, but also that the correction lacks efficacy. For this reason, the response elicited by the correction is central to research on MP. The effect of misinformation implicitness on MP would be less puzzling if it were caused by initial differences in persuasiveness between the implicit and explicit messages. Rich and Zaragoza (2016) tested this possibility and found no significant effect of misinformation implicitness when there was no correction. Moreover, Rich and Zaragoza (2016) found no significant effect of misinformation implicitness on the believability of the correction. Similarly, Reynolds (2018) found no significant association between misinformation implicitness and participant denials of the correction. These findings increase confidence that misinformation is more resistant to corrective information when communicated implicitly vs. explicitly, and that the effect is not due to differences in the initial persuasiveness of the message.

Because the rationale for **H1**-**H3** involves the role of information accessibility in reinforcing the misinformation, I consider how accessibility might differ if the correction is withheld. Although misinformation implicitness apparently does not influence MP in the absence

of a correction, it is unknown whether accessibility will differ. Therefore, I wonder: [**RQ1**] how will accessibility of misinformation-consistent information differ on account of misinformation implicitness when a correction is absent? Answering this question will clarify whether misinformation implicitness results in a different associative structure upon initial message processing, or (assuming **H1-H3** are supported) whether differences arise only in response to a correction.

In addition, the correction used in prior research also reveals a methodological limitation. Previous research on the effect of implicit misinformation has used the same corrective message across stimulus material (e.g., Reynolds, 2018; Rich & Zaragoaza, 2016). Results show that the correction has been relatively weak. In their third study, when Rich and Zaragoza (2016) directly asked participants how much they believed the correction, the mean was about 3 on a scale ranging from 1 to 6 (for participants who could recall the correction). In addition, Reynolds (2018) observed that about 26% of participants denied that the correction was true, and furthermore, the averages for MP were above the scale's midpoint for both the implicit and explicit condition (after correction). Although the effect of misinformation implicitness on MP emerges in response to correction, the effect may be suppressed due to the correction's weakness. If implicitness makes misinformation more resistant to correction, then I ask: [**RQ2**] will the effect of misinformation implicitness on MP be more pronounced when the correction is stronger (defined as when the correction more definitively rules out the misinformation)? If a stronger correction produces greater differences between implicit and explicit misinformation, this would enhance the opportunity to detect the mechanisms responsible for the effect. This finding would also inform researchers about the robustness of the implicit effect in response to a less weak correction.

Increasing correction strength could also clarify the role of misintegration in facilitating MP. Arguably, people who misintegrate should be especially able to resist a strong correction because they can explain how the correction is consistent with the misinformation. In contrast, people who do *not* misintegrate may be more strongly influenced by a stronger correction. If this were true, one would expect correction strength to moderate the association between MP and misintegration. Specifically, that **[H6]** the association between misintegration and MP will be stronger after a strong correction condition than a weaker correction.

The Present Study

The present study seeks to test the mechanisms responsible for the effect of implicit misinformation on MP. The FTM hypothesis has been proposed, but empirical results have not addressed key variables, such as accessibility. This study also seeks to explore the phenomenon of misintegration in MP and replicate the findings that misintegration moderates the implicitness effect (Reynolds, 2018). In pursuit of these goals, this study will assess a mediation model that places accessibility of misinformation-consistent attributions as a cause of the implicitness effect. Accessibility measures will be used to accomplish this goal. The strength of the corrective information will also be experimentally varied to improve upon the design and to answer the additional questions described above.

Importantly, the model predicts that accessibility of misinformation-consistent attributions will mediate the implicitness effect; however, the specific attributions that play a role should depend on the kind of information presented. For example, if the misinformation concerns a negative event, attributions may involve blame (rather than praise) for the person deemed responsible. Theories of attribution describe the types of inferences that people likely form when interpreting situations. Work by Malle et al., (2014) describes the inferential process

when assigning blame for a negative event. According to this research, when a person is perceived to violate a norm, a primary inference concerns the intentions of that person. In particular, that person's motives tend to be inferred because motive represents the reason for performing the behavior and therefore explains why the action was taken.

Research also indicates that people readily infer character traits after observing a person's behavior. This tendency is sometime called the fundamental attribution error because it supposes that people prioritize characterological judgments while downplaying situational factors (Lieberman et al., 2002). This perspective would predict that, for example, tardiness may prompt judgments about a person's diligence rather than the circumstances that may have prompted the tardiness. If a person commits an act judged to be morally wrong, that person is likely judged as having an immoral character under most circumstances.

The current experiments present information about a jewelry theft, an intentional, morally transgressive act. For the reasons described above, blame attributions and motive inferences should be relevant to receivers trying to understand the situation. In addition, judgments of the perpetrator's moral character should likewise emerge as misinformation-consistent inferences (see details about stimulus materials below). In the present study, accessibility measures should map onto motive and characterological attributions, as these are likely involved in MP. Below, the method is described in detail.

METHOD

Participants

For the study, 644 participants were recruited from subject pools of undergraduate students at a large midwestern university. Of these, 538 completed the relevant tasks and indicated honestly answering the questions (see below); only these participants will be analyzed further. The sample was 65% female. Native English speakers made up 88% of the sample, while native Chinese speakers made up 8%. Regarding race and ethnicity, 15% identified as Asian, 5% identified as Black or African American, 74% identified as White or Caucasian, and 2% identified as Hispanic.

Design

A 2 by 3 between-subjects experiment was conducted. The first factor, *message implicitness*, varied the way misinformation was presented to participants. It consisted of two levels, *explicit* and *implicit* misinformation, referring to whether Evan, the character from the stimulus, is directly implicated in a crime (explicit) or only suggestively implicated (implicit). The second factor, *correction strength*, varied the amount of evidence provided in the correction. This factor consisted of three levels, *strong, moderate*, and *uncorrected*. In the moderate condition, the same correction was used as in prior research (e.g., Reynolds, 2018; Rich & Zaragoza, 2016). In the strong condition, the correction added details about the way the police were able to confirm Evan's whereabout; it refers to multiple sources and video surveillance[3]. The uncorrected condition included no correction.

[3] A pilot study was conducted to assess whether the strong correction was perceived as stronger. Participants received the moderate and strong correction and indicated a) which provided more evidence of Evan's innocence, and b) which made them more confident in Evan's innocence (n = 19). One hundred percent of participants selected the strong correction. These results were consistent with the intended induction of correction strength.

Participants were randomly assigned to conditions with one constraint; because the uncorrected condition was only needed for RQ1 (see power analysis below), after 182 subjects were assigned to the uncorrected condition, subsequent participants were assigned to either the moderate correction (n = 241 total) or the strong correction (n = 221 total).

Materials and Procedures

Both recruitment and participation occurred online. Participants were informed that approximately 45 minutes would be required to complete the study. Participants were instructed to complete the study in one sitting—independently and free from distractions—and to follow all instructions carefully. A laptop computer was required as the reaction tasks would not load on mobile devices.

Part 1: Presentation of Stimuli

Participants read a story presented as a series of messages on a computer screen and based on materials used by Johnson and Seifert (1994) and Rich and Zaragoza (2016). The story was written to resemble a news report, and each set of stimuli consisted of messages that conveyed a unique aspect of the narrative. Each message was displayed to participants in the same sequence. Participants were not permitted to return to prior messages, and they clicked to the next page to indicate they had finished reading.

The narrative described a jewelry theft from the home of Mrs. Harter. Implicitness was manipulated in Message 5 by either stating that “Police suspect that the Harters’ son, Evan, may have taken the box from the house” (explicit), or withholding that information (implicit). In this case, Evan’s guilt is considered the misinformation. Depending on the experimental condition, message 11 would contain the correction, stating that Evan was out of town during the theft.

Correction strength was manipulated by varying the amount of evidence that corroborates Evan's whereabouts. A third control condition simply omits the correction.

Part 2: Filler Tasks

Consistent with Rich and Zaragoza (2016), participants did not proceed directly to dependent measures but completed filler tasks consisting of un-related questions (e.g., literature preferences, need for cognition, perspective taking, locus of control). Adopting these procedures is designed to allow more direct comparison with Rich and Zaragoza (2016).

Part 3: Dependent Measures, Writing Task, and Recall Check

After completing the filler tasks, participants proceeded to the accessibility measures, then to the MP measures, then to the writing task (see Appendix A), and then to the recall check. Lastly, participants answered demographic questions and an item about their level of honesty. This order is consistent with prior research (e.g., Smith et al., 2013) due to explicit measures' tendency to influence accessibility by activating cognitions. Accessibility measures, in contrast, are not thought to influence explicit measures. See Appendix B for item text.

The writing task was a free-response question; participants were asked to describe how Evan could have been involved in the theft (if he were involved). The recall check was a free response item assessing whether participants could recall the correction when prompted.

Measurement

Misinformation Persistence

Two Likert-type items measured MP. Participants indicated how much they agreed or disagreed with each of the following statements: "Evan Harter was guilty," and "Evan Harter was responsible for the theft." A 5-point response scale was used, ranging from 1 (strongly

disagree) to 5 (strongly agree). For the scale, M = 3.10, SD = 1.04, *min*. = 1.00, *max*. = 5.00, *skewness* = 0.01, α = .94.

Accessibility Measures

Consistent with prior literature (e.g., Ewoldsen & Fazio, 1997; Fazo, 1995), accessibility is conceptualized as a cognition's proclivity be activated from memory, or the ease with which it is activated. Accessibility was measured with response latency to a cue that presented particular information; the more rapid the response (low latency), the more accessible the cognition. For the accessibility task, participants were instructed to press either the 'Agree' or 'Disagree' key when statements were displayed on the computer screen. Statements were designed to reflect misinformation-consistent information, specifically, attributions made about Evan, whom the misinformation implicates in the theft.

Based on theories of blame attribution (Malle et al., 2014) and dispositional models of attribution (Lieberman et al., 2002), reaction-time items included statements that reflect motives for Evan to commit the theft and statements about Evan's trait-level unethicality. Accessibility of the misinformation itself was also assessed with items about Evan's involvement in the crime. Participants were instructed to perform the task as rapidly as possible but without making many errors. Participants were instructed to select the response that reflects their own thoughts/feelings. The first block contained practice trials, not used in this analysis. The second block included items about Evan's unethical traits, and the third block included items about Evan's motive for committing the theft (see Appendix B for item text). The order of items was randomized within each block.

Initially, reaction times were recorded as the number of milliseconds between stimulus onset and participant response. Because extreme outliers are expected in reaction-time data (e.g.,

Ratcliff, 1993; Whelan, 2008), the "absolute deviation" method (Leys et al. , 2013) was adapted to identify and either drop or recode outliers. To address extremely slow responses (representing inattention or failure to follow instructions), for each item, the median latency was calculated, and responses were dropped if they were more than 10 times the median absolute value above the median. Next, the process was repeated but values more than 8 times the median absolute value above the median were recoded to equal 8 times the median absolute value above the median. This represents a conservative threshold (Leys et al., 2013) to avoid overly transforming the data. Extremely fast responses (less than 150 milliseconds) were dropped due to the likelihood that they do not represent a genuine reaction to the stimulus (Whelan, 2008). Next, responses were recoded to equal 250 milliseconds if they were below that value (and above 150 milliseconds).

In total there were 3780 trials across all items and all participants. Using the criteria described above, 78 trials (2.1%) were dropped, and 59 (1.6%) were recoded. For the purpose of analysis, the multiplicative inverse of reaction times was calculated to make higher numbers represent higher levels of accessibility, consistent with Rhodes et al., (2008).

Unethicality Attributions. Consistent with the procedures described above, accessibility of Evan's trait-unethicality was assessed with four reaction-time items. Believing that Evan is unethical is consistent with the misinformation (that he is guilty of stealing from his mother). Each item was a statement about Evan's character. Example items include, "Evan is dishonest," and "Evan is corrupt." See Appendix B for these items. For each item, the multiplicative inverse was calculated so higher scores represent greater accessibility. For the scale, M = 1.00, SD = 0.50, *min*. = 0.29, *max.* = 4.00, *skewness* = 2.80, α = .81.

Motive Attributions. Consistent with the procedures described above, accessibility of Evan's motive for committing the theft[4] was assessed with three reaction-time items. Each item was a statement about a likely motivation for Evan to commit the theft. Example items include, "Evan needed the jewelry," and "Evan wanted money." See Appendix B for these items. For each item, the multiplicative inverse was calculated so higher scores represent greater accessibility. For the entire scale, $M = 0.97$, $SD = 0.51$, *min.* = 0.17, *max.* = 4.00, *skewness* = 3.13, $\alpha = .75$.

Misinformation. Accessibility of the misinformation was assessed with three reaction-time items. Each item was a statement about Evan's involvement in the crime, for example, "Evan took the jewelry," and "Evan was guilty." See Appendix B for these items. For all items, the multiplicative inverse was calculated so higher scores represent greater accessibility. For the scale, $M = 0.99$, $SD = 0.39$, *min.* = 0.21, *max.* = 4.00, *skewness* = 2.59, $\alpha = .70$.

Misintegration

Two coders worked independently to evaluate the presence or absence of misintegration in participant responses produced during the writing task. Misintegration was evident when the response described how both the misinformation and the correction could be true, for example, by stating that Evan told his friends to steal the jewelry while he was away. The variable was coded dichotomously (1 = misintegration present, 0 = misintegration absent). See Appendix C for more details about the coding protocol.

Training data from a prior study were used during initial coding sessions. Training occurred over three weeks with a total of six hours of meetings and four practice assignments. Intercoder reliability was assessed with Gwet's AC1 (γ); it adjusts for chance agreement,

[4] That Evan committed the theft is the misinformation in this case. Evan's motive is an attribution that provides an explanation for the misinformation.

provides more stable estimates of inter-rater reliability than other metrics, and is less susceptible to biasing effects of skewed category and response distributions (Wongpakaran et al., 2013). Perfect agreement would produce $\gamma = 1$, whereas agreement at chance levels would produce $\gamma = 0$. Raw percent agreement (*PA*) was also calculated.

On the final training assignment ($n = 54$), coders achieved the following intercoder reliability for misintegration: $PA = .89$, $\gamma = .82$. For the messages used in analysis, both coders evaluated all messages ($n = 644$). Here, coders demonstrated the following intercoder reliability: $PA = .95$, $\gamma = .93$. Consensus was reached for all disagreements to provide final codes. All coders were blind to experimental conditions.

Correction Recall

Participant's ability to recall the correction was assessed with a single free-response question. Consistent with Rich and Zaragoza (2016), participants were asked, "What did the story report about where Evan Harter was during Mrs. Harters' vacation?" Participant responses were content analyzed by two independent coders. Coders were trained to code 1 if participant messages referred to Evan being away, and 0 otherwise (missing excluded). Coders demonstrated the following reliability: $PA = .91$, $\gamma = .88$. Consensus was reached for all disagreements to provide final codes. All coders were blind to experimental conditions. In total, 56.1% of participants recalled the correction.

Honesty Check

Participants indicated their level of honesty during the study on a 5-point scale ranging from "very low" to "very high". Higher scores indicated greater honesty: $M = 4.19$, $SD = .75$. As stated above, participants were excluded from analysis if they did not indicate honestly answering the questions (above the midpoint for the scale).

Power Analysis

Consistent with the recommendations of Balkin & Sheperis (2011), statistical power of greater than .8 was targeted for all tests, using an alpha level of .05. To account for the more restrictive inclusion criteria for the corrected condition, and due to the predicted non-additivity involving the moderate and strong correction groups, more participants were recruited for the corrected than the uncorrected conditions. Obtained power was subject to exclusion criteria discussed above as well as the correction-recall check (see below). Only the corrected conditions involved the correction-recall check, which resulted in an exclusion of 107 additional participants.

After applying all exclusion criteria for the planned tests, 431 participants remained (see correction-recall check below). Using G*Power software (Faul et al., 2007) we calculated statistical power for multiple tests. Where the present analysis was replicating prior results (Rich and Zaragoza, 2016; Reynolds, 2018), effects had been in the "medium" range (e.g., Cohen's f = .2, Cohen's d = .5). In addition, I assumed a medium effect size for the novel tests. Results showed that for main effects involving the corrected conditions, achieved power ($1 - \beta$) was .96. For interactions involving the corrected conditions achieved power was .87. For main effects involving the uncorrected conditions, achieved power was .84. These results indicate that the study was adequately powered.

RESULTS

Measurement Model

Confirmatory factor analyses (CFA; see Reynolds, in press) assessed the fit of the measurement model. The model included the following nine factors: message implicitness (one indicator), correction strength-strong (one indicator), correction strength-moderate (one indicator), misintegration (one indicator), recall (one indicator), MP (two indicators), unethicality accessibility (four indicators), and motive accessibility (three indicators), and misinformation accessibility (three indicators). Using ML estimation, the model showed good fit: $\chi^2(88) = 150.64$, $p < .001$, *TLI* = .97, *CFI* = .98, *RMSEA* = .04, *SRMR* = .03 (n =474; listwise deletion was used). These results are consistent with valid measurement. Appendix D contains factor loadings.

Correction-Recall Check

As stated above, a majority of total participants recalled the correction when asked. Recall did not differ significantly between the moderate and strong correction, χ^2 (1) = 0.64, p = .46, but recall was much lower in the uncorrected condition than the other two, $\chi^2(1) = 169.9$, $p <$ .001, r = -.56. Recall by correction strength was as follows: strong = 71.1%, moderate = 74.6%, uncorrected = 9.8%. Recall did not differ significantly between the implicit (53.0%) and explicit (52.5%) conditions, $\chi^2(1) = 0.06$, p = .81. Participants in the moderate or strong conditions who failed to recall the correction will be excluded from further analysis, leaving 431 participants remaining across all conditions. See Figure 2 for n by condition.

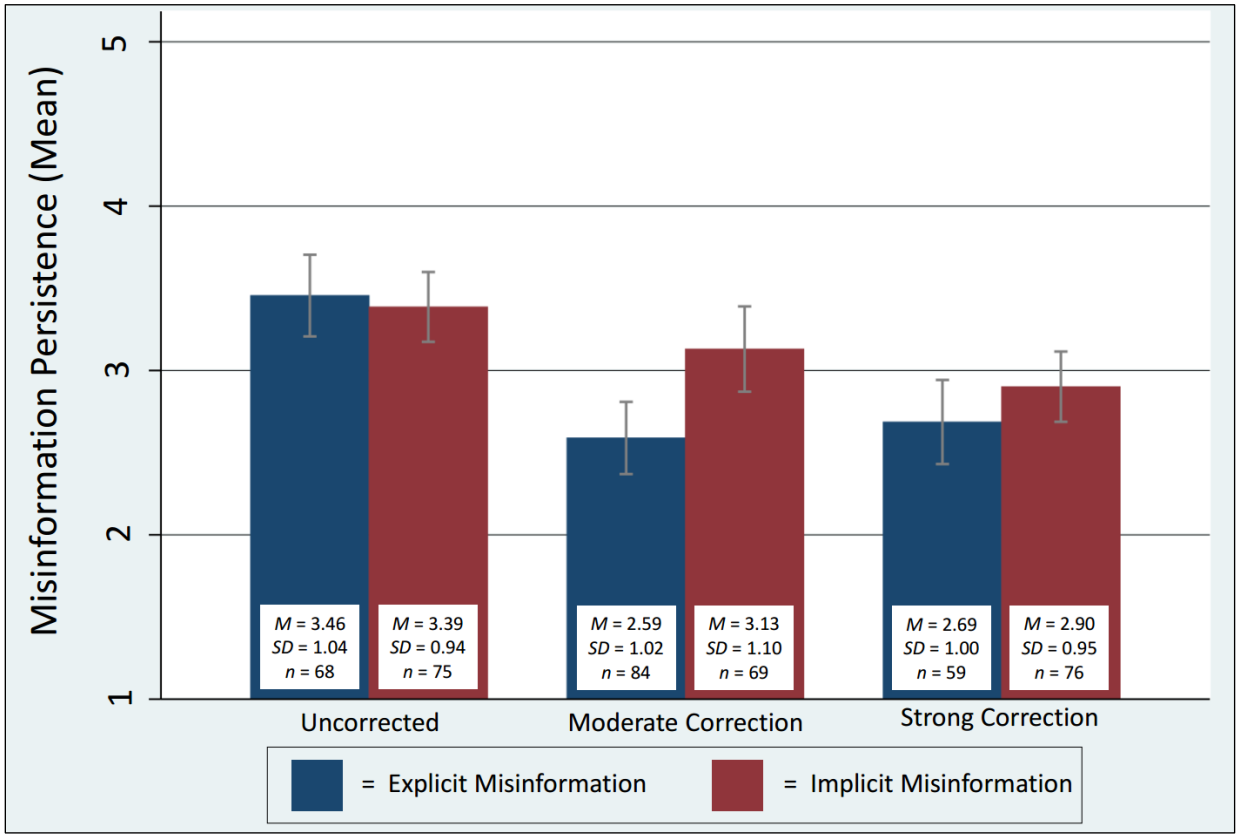
Replication of Implicitness Effect

A replication was expected wherein implicitness would positively correlate with MP after a correction. For this test, I exclude the uncorrected condition. Results show the implicit message was associated with greater MP, $r(286) = .18$, p = .002, replicating the implicitness effect. See

Figure 2 for a visualization of MP across all conditions. Consistent with Rich and Zaragoza (2016), in the uncorrected condition there was no significant association between implicitness and MP, $r(286) = -.04$, $p = .68$. I further address the effect of correction strength after testing the mediation model.

Figure 2.

Misinformation Persistence by Experimental Condition



Note. $N = 431$. Error bars represent 95% confidence intervals.

Information Accessibility and Mediation Model

H1 predicted that implicit misinformation would produce greater accessibility of misinformation-consistent information following a correction. Excluding the uncorrected condition, there was no significant association between implicitness and accessibility of

unethicality attributions, $r(285) = -.06$, $p = .34$, nor accessibility of motive attributions, $r(285) = .05$, $p = .41$, nor accessibility of misinformation $r(285) = .01$, $p = .83$.

Follow-up regressions were conducted to probe for an interaction between correction strength and implicitness. For each accessibility variable (trait attribution accessibility, motive attribution accessibility, and misinformation accessibility) there were no significant main or interaction effects. These results are inconsistent with H1. See Appendix E (Table 3-5) for detailed regression outputs.

RQ1 asked whether implicitness would affect accessibility in the uncorrected condition. Including only the uncorrected condition, results showed no significant association between implicitness and either trait accessibility, $r(139) = -.03$, $p = .75$, motive accessibility, $r(139) = .12$, $p = .17$, or misinformation accessibility, $r(137) = -.00$, $p = .98$.

H2 predicted that the accessibility of misinformation-consistent information would be positively associated with MP following a correction. Excluding the uncorrected condition, there was no significant association between trait attribution accessibility and MP, $r(285) = .01$, $p = .84$, or misinformation accessibility and MP, $r(285) = .09$, $p = .09$. However, there was a significant association between motive attribution accessibility and MP, $r(285) = .19$, $p = .001$. These results are consistent with H2, but only with respect to motive accessibility.

H3 predicted that the following mediation model would fit the data: implicitness increases accessibility of misinformation-consistent information which increases MP. Because no direct effect was found between implicitness and accessibility, the mediation model is not consistent with the data.

Misinformation Representations

H4 predicted that misintegration would be associated with MP following a correction. Excluding the uncorrected condition, there was no significant association between misintegration and MP, $r(286) = -.05$, $p = .42$. These results are inconsistent with H4.

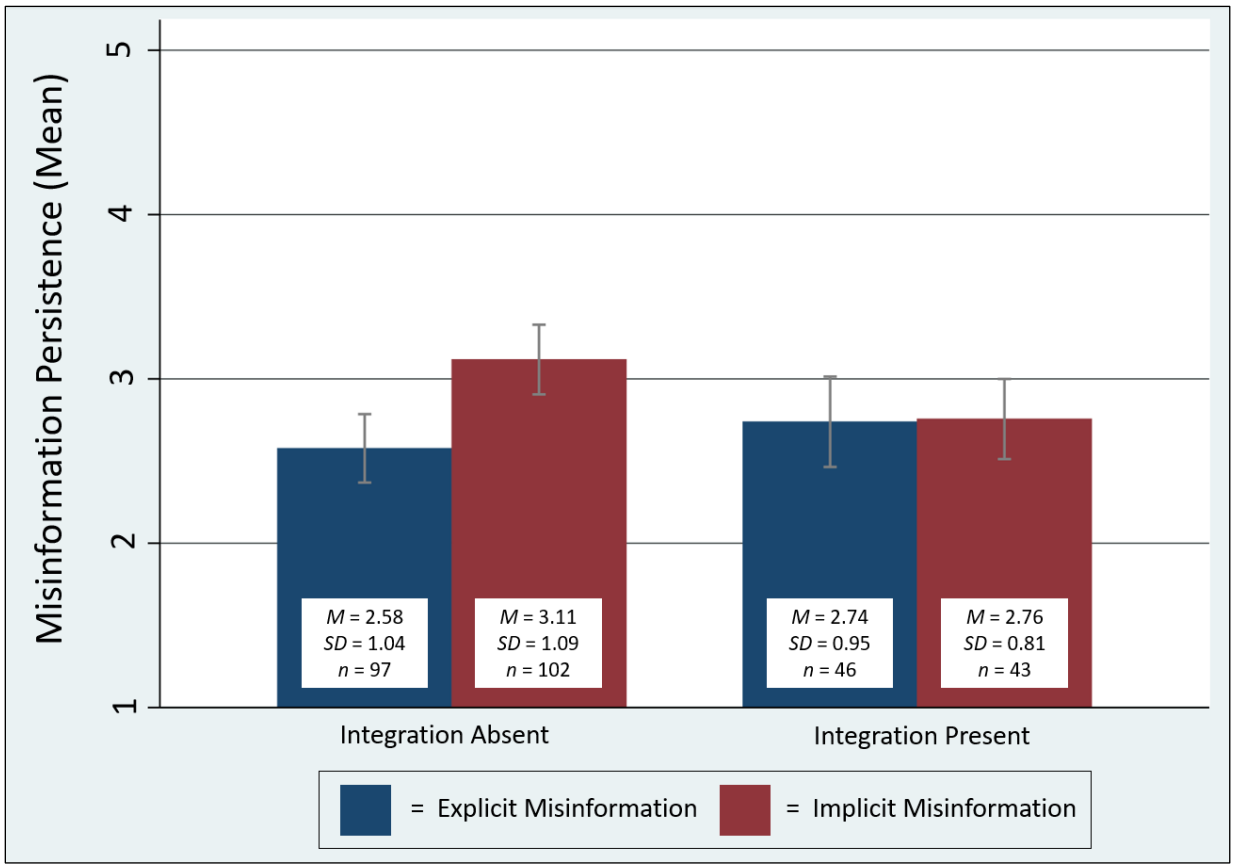
H5 predicted an interaction between implicitness and misintegration. Specifically, that misinformation implicitness will lead to greater MP following a correction only when participants do not integrate. Excluding the uncorrected condition, multiple regression revealed a significant interaction between implicitness and misintegration. The overall model was significant, $F(3, 284) = 4.94$, $p = .002$, $R^2_{adj} = .04$, as was the interaction term, $\beta = -.18$, $p = .04$ (see Appendix E for detailed regression output [Table 6]). Follow-up analyses were consistent with H5; for participants who did integrate there was no association between implicitness and MP, $r(87) = .01$, $p = .93$, but for participants who did not integrate, the association was significant, $r(197) = .25$, $p < .001$. See Figure 3 for a visualization and descriptives by condition.

Correction Strength

RQ2 asked whether the association between implicitness and MP would be stronger in the strong correction condition than in the moderate correction condition. Multiple regression was conducted to test the interaction, with MP as the dependent variable. Excluding the uncorrected condition, the overall model was significant, $F(3, 284) = 4.08$, $p = .007$, $R^2_{adj} = .03$, but the interaction term was not, $\beta = -.14$, $p = .18$. In the model, only the effect of implicitness was significant, $\beta = .26$, $p = .001$ (see Appendix E for detailed regression output [Table 7]). These results do not suggest the association between implicitness and MP differs between the strong and moderate correction.

Figure 3.

Implicitness by Misintegration Interaction



Note. N = 288. Error bars represent 95% confidence intervals.

H6 predicted that the association between misintegration and MP would be stronger in the strong correction condition than the moderate correction condition. Multiple regression was conducted to test the interaction. Excluding the uncorrected condition, the overall model was not significant, $F(3, 284) = 0.30$, $p = .83$, $R^2_{adj} = -.01$, and neither was the interaction, $\beta = .04$, $p = .65$ (see Appendix E for detailed regression output [Table 8]). These results are not consistent with H6; the association between implicitness and MP did not differ across the strong and moderate correction. In fact, contrary to the findings of Reynolds (2018), misintegration was not significantly associated with MP (following a correction), $r(286) = -.05$, $p = .42$.

Supplemental Analyses

Conditional Association of Accessibility and MP

In testing H2, I found that accessibility of motive attributions was positively correlated with MP, and no evidence of moderation was observed when differentiating between the moderate and strong corrections (excluding uncorrected conditions). It may be that the association between accessibility and MP varies between the corrected and uncorrected conditions. To test this possibility, a multiple regression model was estimated with MP as the DV, predicted by motive-attribution accessibility (as a continuous factor), a dichotomous correction-strength indicator (0 = no correction; 1 = moderate or strong corrections), and their interaction term. The overall model was significant, $F(3, 424) = 14.43$, $p < .001$, $R^2_{adj} = .09$, as was the interaction term, $\beta = .16$, $p = .030$ (see Appendix E for detailed regression output [Table 9]). This result indicates the association between motive accessibility and MP significantly differs between the uncorrected and corrected conditions. Follow-up analysis showed there was no association between motive-accessibility and MP in the uncorrected conditions, $r(139) = -.00$, $p = .99$, whereas the association was positive in the corrected conditions, $r(285) = .19$, $p = .001$ (per H2).

Motive Integration

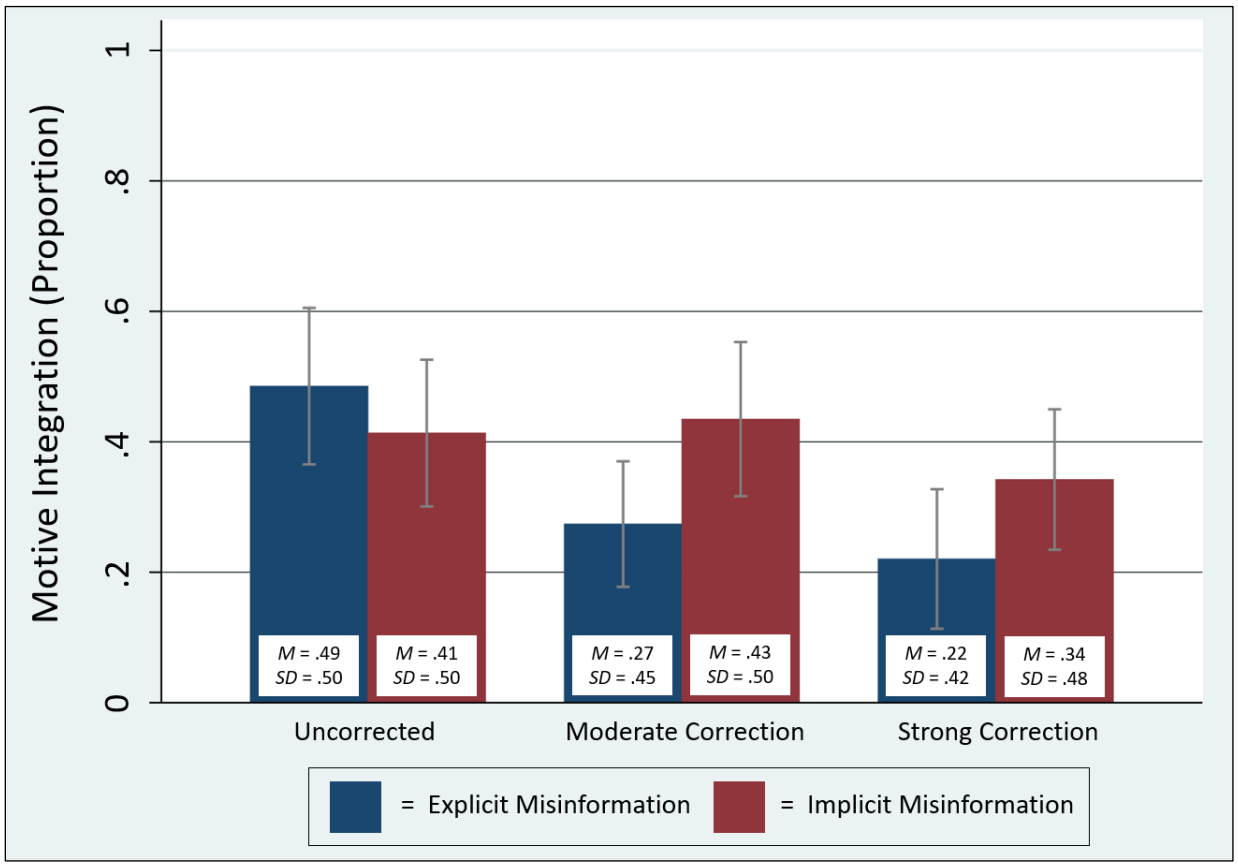
Because motive accessibility was positively associated with MP, the role of motive attributions was further explored. Specifically, the inclusion of motive attributions within misinformation accounts was coded as a dichotomous variable (1 = Evan's motive was referenced [e.g., he needed money, he had to pay his debts], 0 otherwise [missing excluded]). This variable indicates whether a motive attribution was integrated with the misinformation

representation. Two coders coded all messages; n = 644, PA = .88, γ = .80 (see Appendix C for coding details).

In the corrected conditions, implicitness was positively associated with motive integration, $r(286) = .14$, $p = .014$, whereas no significant association was found in the uncorrected conditions, $r(141) = -.07$, $p = .39$. Implicit misinformation lead to more motive integration than explicit misinformation, but only in response to a correction (moderate or strong). See Figure 4 for the proportion of motive-references by condition.

Figure 4.

Motive Integration by Condition



Note. N = 431. Error bars represent 95% confidence intervals.

The role of motive integration in misinformation accounts was also explored as a predictor of MP. Including only the corrected conditions, there was a positive association

between motive integration and MP, $r(286) = .27$, $p < .001$. Follow-up analyses showed no significant interaction between motive references and either correction strength or implicitness, which does not suggest the association differs across conditions. Compared with people who displayed no motive integration, people who did so tended to believe the misinformation more strongly, with or without a correction.

Next, motive integration was tested as a mediator of the implicitness effect. Addressing this issue requires examining the intermediate paths as well as the fit of a mediation model where implicitness (I) causes motive integration (MI) which cause MP. Only the uncorrected conditions were included to focus on the MP effect in response to a correction. Results showed that although each intermediate paths was positive and significant (as discussed above), the model fit was poor, $\chi^2(1) = 7.80$, $p = .009$, *TLI* = .45, *CFI* = .82, *RMSEA* = .14, *SRMR* = .06 ($n = 288$).

Motive and Unethicality Attribution Beliefs

The same items used to measure response times for the accessibility measures can be used to indicate belief in the motive and unethicality information by evaluating whether participants agreed or disagreed with the statements. First, a motive-attribution belief scale was constructed representing the proportion of motive items participants agreed with, out of the three items. Across all conditions, the mean was .74, *SD* = .28, min. = 0, max. = 1, $\alpha = .55$. Similarly, a four-item unethicality-attribution belief scale was created. Across all conditions, it had a mean of .46, *SD* = .37, min. = 0, max. = 1, $\alpha = .75$.

A modified test of H1, H2, and H3 is possible using the motive- and unethicality-attribution belief scales. First, correlations revealed no significant association between implicitness and either motive attribution beliefs, $r(285) = .01$ $p = .87$, or unethicality attribution beliefs, $r(285) = .01$ $p = .92$ (excluding uncorrected conditions). On the other hand, MP was

positively associated with both motive attribution beliefs, $r(285) = .21$ $p < .001$, and unethicality attribution beliefs, $r(285) = .47$ $p < .001$ (excluding uncorrected conditions). Neither of these variables fit the mediation model, however, which requires them to be correlated with implicitness.

There was no evidence that the association between the attribution beliefs (motive and unethicality) differed between the corrected and uncorrected conditions. Although the overall models were significant, the regressions revealed no significant interaction terms. See Appendix E for detailed output (Tables 11-12).

Distinguishing Predictors of MP with Multiple Regression

Thus far, several variables in the present study have been associated with MP following a correction: misinformation implicitness, motive attribution accessibility, motive integration, motive attribution beliefs, and unethicality attribution beliefs. To assess whether these predictors are associated with unique variance in MP, multiple regression was used. Because significant interactions were also found, two interaction terms were also entered as predictors: the implicitness by misintegration interaction, and the motive integration by motive accessibility interaction. See Table 1 for the full model output. Results showed the overall model was significant. In addition, several variables remained significant predictors of MP, suggesting they share unique variance with MP. Interestingly, motive attribution accessibility appears to be uncorrelated with motive integration, $r(285) = .02$, $p = .78$, and motive attribution beliefs, $r(285) = -.05$, $p = .36$. In contrast, motive integration was positively correlated with motive attribution beliefs, $r(285) = .21$, $p < .001$.

Table 1.

Predictors of Misinformation Persistence

	Regression Coefficient	
	B (SE)	β
Motive Accessibility (mean centered)	0.27 (0.17)	.11
Motive Integration	0.44 (0.12)***	.20***
Interaction: Motive Integration by Accessibility	0.77 (0.32)*	.14*
Misinformation Implicitness	0.42 (0.13)***	.20***
Misintegration	0.13 (0.16)	.08
Interaction: Implicitness by Misintegration	-0.42 (0.21)*	-.15*
Motive Attribution Belief	0.22 (0.19)	.06
Unethicality Attribution Belief	1.13 (0.14)***	.41***
Constant	1.89 (0.15)***	-
F (8, 278) = 21.92, $p < .001$, $R^2_{adj} = .31$, $VIF = 1.50$		

Note. $*p < .05$,$**p < .01$, $***p < .001$ (two-tailed). Standard errors are reported in parentheses. B = unstandardized, β = standardized OLS estimates. VIF = mean variance inflation factor. Sample includes corrected conditions only (n = 287).

Alternative Indicator of MP

As discussed above, response latencies were used to measure misinformation accessibility. The response-time items can also indicate MP by evaluating the actual response (agree or disagree). A scale was constructed indicating the proportion of "agree" responses among the three total items. Across all conditions, the mean was .47 (SD = .45), min. = 0, max. = 1, α = .88.

This new indicator was also used as a supplemental test of the predictions, including the implicitness effect, and the effect of accessibility (H2). Excluding the uncorrected conditions, results showed the alternative MP indicator was not significantly associated with implicitness, $r(285) = .08$, $p = .17$, failing to replicate the implicitness effect. Regarding H2, there was a significant association between the new MP indicator and both motive accessibility, $r(285) = .30$, $p < .001$, and misinformation accessibility, $r(285) = .16$, $p = .07$, but not unethicality

accessibility, $r(285) = .09$, $p = .12$. This is consistent with H2 with respect to motive and misinformation accessibility.

A regression also probed whether the association between motive accessibility and the alternative MP indicator varied between the corrected and uncorrected conditions. The overall model was significant, $F(3, 424) = 21.57$, $p < .001$, $R^2_{adj} = .13$, as was the interaction, $\beta = .28$, $p = .01$. See Appendix E for detailed regression output (Table 13). A follow-up analysis revealed that the association between motive accessibility and the alternative MP indicator was not significant in the uncorrected condition, $r(139) = .10$, $p = .23$, whereas it was significantly positive in the corrected conditions (reported above).

The effect of motive integration on the alternative MP indicator was also tested. Excluding the uncorrected conditions, there was a significant positive correlation, $r(285) = .30$, $p < .001$. In addition, the alternative MP indicator was positively associated with motive attribution beliefs, $r(285) = .28$, $p < .001$, and unethicality attribution beliefs, $r(285) = .50$, $p < .001$ (excluding uncorrected conditions). Overall, the alternative MP indicator appears to function similarly to the Likert MP scale, however, it did not demonstrate a significant implicitness effect.

DISCUSSION

Misinformation persistence (MP) remains an important topic for researchers and practitioners alike; all people require new information to survive but are susceptible to undue influence from prior beliefs. Furthermore, within social structures, beliefs lead to decisions with wide-ranging consequences (Lewandowsky et al., 2012). This study addressed an important aspect of the MP phenomenon, that implicit misinformation—mere suggestion—can reduce the effectiveness of corrective messages, relative to explicit misinformation. Prior research had only provided limited evidence about the mechanisms that may account for this effect (e.g., Reynolds, 2018; Rich & Zaragoza, 2016). To expand understanding of this topic, this study tested the failure-to-monitor hypothesis by directly measuring the accessibility of misinformation-consistent inferences. In addition, this study sought to replicate the exploratory finding of Reynolds (2018) on the role of misintegration in MP. The present results provide insight into the factors responsible for MP and suggest fruitful avenues of future research.

Summary of Results

Misinformation-consistent inferences are beliefs generated by a receiver that extrapolate from message content. Attributions are a type of inference that explains an event or assign responsibility (Weiner, 2008). This study applied the concept of accessibility to attributions in order to assess the ease with which the attribution is retrieved from memory. The present study found that MP was greater when misinformation-consistent attributions were more accessible following a correction (consistent with **H2**). Specifically, the accessibility of motive attributions was positively associated with ratings of guilt, as guilt was the misinformation in the present case. The accessibility of dispositional attributions (trait unethicality), however, were not associated with MP. In addition, accessibility of the misinformation itself was not associated

with MP. Although attribution accessibility appears to be linked with MP, the present results were not consistent with its role as a mediator of the implicitness effect; accessibility of misinformation-consistent information was not associated with the implicitness induction.

In terms of misinformation representations, this study further showed the relevance of the misintegration phenomenon where a correction is represented as consistent with misinformation. Replicating Reynolds (2018), the present results showed that integration moderates the implicitness effect (consistent with **H5**). Specifically, the implicitness effect seems not to apply when participants misintegrate. The present study did not find a direct association between integration and MP, however, failing to replicate an aspect of Reynolds (2018).

Although pretesting showed the experimental induction successfully varied correction strength, the predicted effect of correction strength was not found. The strong correction did not produce a greater implicitness effect (on MP) than the moderate correction. If anything, the implicitness effect was weaker in response to a strong correction, although not significantly different. In addition, correction strength did not moderate the association between misintegration and MP.

Supplemental analyses revealed several additional findings. First, misinformation implicitness increased the likelihood that misinformation representations would contain motive attributions (motive integration), but this effect was only observed following a correction. Second, motive integration was positively associated with MP; people who referenced a motive in their misinformation accounts believed the misinformation more strongly. Although motive integration proved an important feature of misinformation representations, it did not appear to function as a sole mediator; the poor fitting mediation model suggests that other factors likely contribute to the effect of implicitness on MP. As stated above, the accessibility of unethicality

attributions was not correlated with MP, however, believing in the unethicality attributions was positively associated with MP. In addition, the multiple regression suggests that several distinct mechanisms play a role in MP simultaneously, and the model explained substantial variance in MP (R^2_{adj} = .31).

Implications

Research on MP has produced knowledge about how corrections fail, and several explanations of MP and the implicitness effect. This study provides important insight into this process, with implications for models of human memory, theories of attributions, and the role of mental representations in decision-making.

The primary aim of this study was to test the failure-to-monitor (FTM) hypothesis and to extend the work of Rich and Zaragoza (2016) and Reynolds (2018). The FTM hypothesis states that MP results from retrieval failure (failure to access corrective information) and/or reasoning failure. FTM predicts that highly accessible misinformation-consistent attributions reduce the ability of corrective information to impact judgments and reduce MP. The results of the present study did not support the FTM hypothesis as a whole, although parts of the model do align with the data; consistent with FTM, accessibility of motive attributions was positively associated with MP after a correction. The FTM hypothesis is also consistent with the supplemental analysis that found the accessibility-MP correlation depended in the presence of a correction. Because the accessibility mechanism of monitoring failure involves scarce cognitive resources, only when information is inconsistent should accessibility be associated with MP. In other words, if all beliefs reinforce the misinformation, the relative accessibility among those beliefs is less important. Indeed, the present results show that in the uncorrected conditions, the accessibility-MP correlation was 0, vs .19 in the corrected conditions. These results further suggest that causal

inferences create strongly interconnected networks in memory (see Kendeou et al., 2013; Trabasso & Van Den Broek, 1985) that interfere with belief modification when they are accessible (prone to activation) and inconsistent with new beliefs.

Results showed that the accessibility measures behaved differently, and literature on attribution theory may explain why. According to Malle et al.'s theory of blame (2014), detecting negative events triggers an attribution process that prioritizes inferences of motive over inferences of stable trait characteristics. In particular, observers wish to determine whether the event resulted from human agency, and then to determine whether it was intentional. In the theory, judgments of intentionality are central to the attribution process (Malle & Knobe, 1997; Malle et al., 2014), and mental states of the actor (such as motives) are central to judgments of intentionality, as they reflect goals of the actor in performing the behavior. On the other hand, literature has questioned the primacy of dispositional (trait) attributions, going as far as to call "dispositionism" a dogma in need of retirement (Malle, 2011). People may infer traits from observed or implied behavior, but traits are incomplete explanations because they provide no immediate cause of the behavior. Furthermore, as Malle (2011) reports in a comprehensive review, people rarely generate dispositional attributions spontaneously when seeking to explain a behavior. In the present study, the unexplained event was described as a theft—an intentional action for which blame is highly relevant. As countless crime dramas corroborate, establishing motive is a primary concern in this context. In other settings, the accessibility of other inferences may be more important. For example, if the behavior were accidental, inferences may focus on whether an individual had a duty to avoid a particular action, or other mitigating circumstances.

Although the accessibility of the dispositional attribution (unethicality) seemed not to play a role in MP, believing in the dispositional attribution was more strongly associated with

MP than any other predictor. This suggests that dispositional attributions are relevant to the MP process. Furthermore, the discrepancy between the accessibility and belief measures may indicate involvement of both automated and controlled cognitive processes (e.g., van den Broek et al., 2015). Differences may emerge when forming attributions if, for example, motive attributions are more automatically generated than dispositional attributions. Likewise, these attributions may produce different effects if they are retrieved more or less spontaneously. Although the current study does not allow direct comparisons of the accessibility of different attributions,[5] it may be that when motive attributions are accessible they particularly interfere with corrective information processing.

The lack of correlation between MP and misinformation accessibility may seem puzzling at first. Perhaps nothing could be more misinformation-consistent than the misinformation itself. Yet, this result may be explained through the attribution process and the activation-based model of memory (Anderson, 1983). A motive attribution serves to explain a behavior (Malle et al., 2014), increase its plausibility, and create associative links between these newly inferred attributions and the behavior. In the present study the corrective messages were targeted at the misinformation directly rather than motive attributions. In this way, the corrections were relevant to the misinformation and irrelevant to motive attributions. The corrections, therefore, should become more strongly linked in memory with the misinformation than with misinformation-consistent attributions. This means that a person may have both highly accessible misinformation and accessible corrections, potentially allowing the correction to work as intended if the correction is retrieved during a judgment. This may explain why misinformation accessibility was not associated with MP whereas motive accessibility was. If a correction had specifically

[5] Systematic differences in the task prevent meaningful comparisons. For example, dispositional attribution accessibility was assessed with single words, taking less processing time.

targeted the issue of motive, for example, by stipulating that Evan in fact had no need of money, the result may have differed. In addition, further research could investigate the accessibility of corrective information to address the associative network between misinformation, attributions, and various types of corrections (but see limitations below).

Despite the observed association between attribution accessibility and MP, no association was observed between the implicitness induction and any accessibility measure. This suggests that the accessibility mechanism of the FTM hypothesis may contribute to MP in general, but may not explain the implicitness effect. Combined with the finding of Reynolds (2018), that articulated monitoring failure within misinformation representations was uncorrelated with implicitness, these results call for new theorizing about how the implicitness effect operates. Alternately, the accessibility mechanism may account for the implicitness effect but other accessibility measures may be required. As previously discussed, accessibility of the correction could be directly assessed. Assuming correction accessibility were the mediator, however, additional questions would follow. According to the FTM hypothesis, implicit misinformation is supposed to lead to more elaborate inferences that compete with the correction for activation (Rich & Zaragoza; Ecker at al., 2011). In the absence of this mechanism, it would be unclear why implicitness leads to reduced accessibility of a correction.

Although the present study found no evidence that implicitness is associated with information accessibility, it did find that misinformation representations play a role in the implicitness effect. In particular, this study replicated the finding of Reynolds (2018), showing that misintegration moderates the implicitness effect. When people construe the correction as consistent with the misinformation, implicitness does not influence MP. This reiterates how corrections depend on additional assumptions shared by receivers, echoing interpretive

perspectives of information processing (Delia, 1977). In terms of reducing MP, if misintegration could be anticipated, it could be preemptively addressed in corrective messages.

Although misintegration itself was not associated with MP in the present data, it is relevant to the FTM hypothesis in multiple ways. Clearly, misintegration does not represent monitoring failure; it does not involve retrieval failure because it incorporates the correction into the misinformation representation, nor does it involve a failure to reason because it explicitly explains the lack of inconsistency. The FTM hypothesis would expect that, in the absence of monitoring failure, implicitness should have no effect. Therefore, the moderating role of misintegration is consistent with the FTM hypothesis, although it does not demonstrate monitoring failure.

Supplemental analyses also examined misinformation representations in terms of motive integration. Results showed that implicit misinformation was more likely to produce motive integration than explicit misinformation, and motive integration was positively associated with MP. Despite these significant correlations, the poor-fitting mediation model suggests that other variables are involved and a simple causal sequence is unlikely. Nonetheless, considering motive integration as a potential cause of the implicitness effect may be helpful. Arguably, motive integration would be expected to function in a similar way as motive accessibility. Specifically, the accessibility mechanism of FTM supposes that more explanatory inferences are generated in the implicit than explicit conditions. The FTM hypothesis would therefore predict that motive attributions would be integrated into misinformation representations more frequently following implicit rather than explicit misinformation.

It remains puzzling that motive accessibility and motive integration were uncorrelated and showed different patterns of association with implicitness. This finding has several

implications and demands further conceptual clarification. Whereas motive accessibility refers to the ease with which motive information is activated from memory, motive integration reflects whether motive is included in the misinformation representation. In other words, integrating information is not identical with forming a simple associative network between them; integration transforms the larger understanding of the object that consists of the information. For example, consider a set of directions from A to B to C. The fact that each step is associated does not provide the order of implementation or, more generally, the structure in which they hang together. Information representations consist of qualitative and structural components that enable identification of complex objects, events, situations, and people, rather than mere associations between features. This view is consistent with prior work on mental representations (Fuster, 1997; Roskos-Ewoldsen et al., 2009; Song & Ewoldsen, 2015) and the role of integration in transforming representations (Ecker et al., 2011; Kendeou et al., 2013; Todorov & Uleman, 2003; Yonelinas, 2013). In addition, this view aligns with Anderson's (2003) insight about human memory, that mental processes impose structure on associative networks and do not simply reflect patterns of activation among information stored in memory.

Research on the meta-cognitive model of attitudes (Petty et al., 2007) further distinguishes between associative and evaluative processes. The model follows conventional associative memory frameworks (e.g., Anderson 1983; Ayers & Reder, 1998), supposing that representations of objects become associated with features that define their characteristics. The meta-cognitive model further states that "validity tags" determine the confidence that one places in a given association (i.e., that the feature truly characterizes an object). Although direct evidence of validity tags has yet to be observed, they would explain the disjunction between belief accessibility (an indicator of associative strength) and belief integration (a process that

involves some level of validity judgments). The meta-cognitive model has implications for MP because it suggests the role of accessibility depends on these validity judgments; information tagged as invalid will be less influential, regardless of accessibility (assuming the tag itself is accessible). In the present study, this would mean that motive accessibility should correlate with MP to a greater extent when motive integration occurs, because motive integration involves judging the motive attribution as valid to some extent. A follow-up regression revealed a significant interaction between motive accessibility and motive integration when predicting MP. The overall model was significant, $F(3, 283) = 13.15$, $p < .001$, $R^2_{adj} = .11$, as was the coefficient for the interaction term, $\beta = .35$, $p = .03$ (see Appendix E for detailed regression output [Table 10]). Results show the correlation between motive accessibility and MP was higher when motive integration occurred, $r(90) = .32$, $p = .001$, than when it did not occur, $r(193) = .16$, $p = .03$.

These findings demonstrate key differences between information accessibility and information representations, but also their interdependence within information processing. When evaluating information, associations that are not considered valid may be discounted. If the FTM hypothesis is correct, it appears not to rely exclusively on retrieval failure or differences in accessibility. Breakdowns in reasoning about the information retrieved are another avenue of monitoring failure (Rich & Zaragoza, 2016) that should be further studied. This may pose a challenge because the reasoning process is not easy to observe directly, although thought-listing techniques may be useful (Cacioppo et al., 1997). The writing task in the present study is one example that allowed the observation of distinct forms of message-processing and reasoning (or failure to reason). Interestingly, the general category of information integration can be either consistent or inconsistent with monitoring failure, depending on how information is integrated. For example, misintegration uses reason to explain the correction without contradicting

misinformation. In contrast, integrating motive attributions may interfere with monitoring. The present data suggest several mechanisms of MP, but research should continue to explore how misinformation sticks and why suggestion leads to lasting effects.

The significant role of misintegration in MP also has implications for the very meaning of the terms *misinformation* and *correction*. The term *misinformation* assumes the information is false, and *correction* assumes that the information is actually inconsistent with the misinformation. Rich and Zaragoza (2016) echo this view when they remark about the correction: "This is a correction because the son could not have committed the theft if he was out of town" (p. 4). Particularly when causal attributions are the issue, these assumptions may not be empirically demonstrable. In fact, many of our participants provided reasons to reject those assumptions. Although the question of truth is important, arguably, the relevance one belief has to another depends on constructing the understanding that they are consistent or inconsistent (Delia, 1977). This position does not deny truth or reality, but encourages a search for multiple interpretations to map the boundaries of plausibility for any set of facts. Perhaps more interesting than the degree of consistency is the way that beliefs are made to seem consistent. Not every interpretation is possible or permissible, but new assumptions often transform the meaning of information previously obtained, and evidently, numerous assumptions are available, enabling countless conclusions. One should ask whether and under what conditions the label of misinformation is properly applied.

Limitations

The present results should be considered along with multiple limitations. First, the accessibility mechanism of the FTM hypothesis could be more fully examined by measuring accessibility of the correction directly. This could reveal whether correction accessibility is

associated with either MP or other accessibility measures. However, directly measuring accessibility of the correction poses methodological concerns. If conventional reaction-time tasks were used, they would require re-presenting corrective information to receivers. This would risk weakening the MP effect by priming the corrective information (Ecker et al., 2011), influencing memory for the correction, and/or increasing accessibility of the correction. Research has already shown that explicit measures can influence implicit measures (Smith et al., 2013), and future research could further study whether accessibility measures would influence explicit measures in this case. Such a design could manipulate the inclusion or order of accessibility measures.

One way to increase confidence in the measurements is to use a variety of instruments. Other forms of accessibility instruments include word-stem completion tasks (Benjamin Clarke & Butler, 2008) and lexical decision tasks (Ratcliff et al., 2004). Further research may reveal idiosyncrasies in these measures, or may help establish convergent validity for the constructs described in this manuscript. Beyond accessibility, this principle also applies to measures of integration and belief. Future experiments can continue to assess the extent to which these constructs are distinct and, if so, what their unique role is in MP and message processing generally.

This study also relies on a single set of stimuli designed to vary the relevant aspects of the misinformation-containing message. This method is suitable for testing mechanisms of MP but raises questions about generalizability. Increasingly, researchers are prioritizing message-sampling strategies (Reeves et al., 2016), as there may be idiosyncratic aspects of the stimulus that play a role in the observed effects. Although the stimulus was chosen specifically to replicate Rich and Zaragoza (2016), and observe the mechanisms responsible for their findings, there is a need to explore the boundary conditions of the effects reported in this manuscript.

In addition, this study helps explain MP in response to a correction but it does not directly predict what kinds of corrective messages should better help reduce MP. Researchers have long called for more attention to message tailoring (Noar et al., 2009), and the present study provides some insight in how this research could be extended. Because several cognitive tendencies seem to produce MP, and especially because several moderators of MP were identified, there may be no universal messaging strategy that consistently eliminates MP. For example, people who deny the correction will require different persuasive strategies compared with people who misintegrate; in particular, bolstering the credibility of the correction would likely be ineffective in response to misintegration. In addition, knowing the specific attributions that a person had made could help tailor messages for each individual. Future research can leverage the current results to better understand how corrections succeed—not just how they fail.

APPENDICES

APPENDIX A

Writing Task

Instructions

Think back to the news story you read. You may or may not believe that Evan was involved in the theft. In the box below, describe how the theft occurred, if it were true that Evan was involved.

APPENDIX B

Questionnaire Items

Direct Measure of Evan's Involvement (MP)

5-point scale items (Likert-type)

Indicate the extent you agree or disagree with the following statements

1. Evan Harter was guilty.
2. Evan was responsible for the theft.

Accessibility Items

Instructions: For the next items, statements about the news story and its characters will be displayed. Indicate if you disagree or agree with each statement. Press the "z" key if you DISAGREE. But press "/" key if you AGREE. Respond AS RAPIDLY AS POSSIBLE, but don't respond so fast that you make many errors. (Occasional errors are okay.)
Continue when you are ready.

Motive Attributions

1. Evan needed the jewelry.
2. Evan wanted money.
3. Evan desired money.

Dispositional Attributions (Unethicality)

1. Evan is bad
2. Evan is sneaky
3. Evan is dishonest
4. Evan is corrupt

Misinformation Accessibility

1. Evan took the jewelry
2. Evan was guilty
3. Evan committed a crime

Correction Recall Check (free response)

1) What did the story report about where Evan Harter was during the Harters' vacation?

APPENDIX C

Coding Manual

This manual contains instructions for coding messages. Messages are contained in an Excel file, each row containing a message, the first row indicating column names. Study this manual and the spreadsheet before coding and refer to it when necessary. Apply codes in the appropriate cell and check your work as you go.

To code:
Orient yourself to the spreadsheet and verify that messages (rows) are presented in the correct order. Read the text of each row and apply the code for each column. Instructions for each column are presented below. Each column represents the presence or absence of a message feature for each row. Use your best judgment to make a decision. If there is a problem or you are not sure make a note. Always code independently. You will want to "freeze" the top row in Excel to keep it visible. In Excel click View > Freeze Panes > Freeze Top Row. Do not make any edit to the Excel document except to enter the codes.
Remember, unless it is necessary to infer the intended meaning of the utterance, code the text based on what is explicitly said.
Contact me if you have questions: reedmr@msu.edu

Background Information:
The messages you are coding were written by participants who read a story about a jewelry theft. Mrs. Harter had her jewelry stolen. Police initially suspect her son, Evan Harter; however, later in the story a correction provides updated information. **The correction states that Evan Harter was out of town on a business trip at the time of the theft.** Participants were instructed to write a message explaining how Evan could have been involved in the crime. We are interested in the content of the messages that reflect the understanding of events in the story.

Important Notes:
Sometimes participants might only describe possibilities about what could have happened. For example:

- If Evan did it then he must have planned carefully

Notice how this statement doesn't say what actually occurred, but highlights one possibility. Even if messages contain only statements about what could or would have happened, we will consider that as their interpretation of the events. This means we will code based on what is stated as potentially true.

Column Name	Message Feature	Code	Examples
Correction Contradict	Message states information that contradicts correction. For example, statements indicating that Evan personally stole the jewelry, or was in town when the robbery occurred. **Note**: Correction is about location; contradiction must be about Evan's location at time of theft. If message states that Evan had access at an unspecified time in the past (not mentioning access during the theft) code 0.	1 if present Otherwise code 0	• "Evan broke into his parents' house" • "Evan was in town during the robbery" • "He had access to the house at the time of the theft" • "Evan stole the box"
Misintegration	Message provides explicit explanation about *how* Evan was involved in the crime without being present when the theft occurred. Misintegration does not necessarily state intentionality. **Note**: Message may state how Evan *could* have been involved without being at the scene. Message may be vague about his actual whereabouts.	1 if present Otherwise code 0	• "Evan told his friend to break in while he was away" • "Evan had an accomplice" • "Evan left the door unlocked for the thief"
Motive Integration	Message refers to Evan's motive for committing the theft. Any motive counts, as long as it proposes why he committed the crime. If unsure whether a motive is referenced, code 0.	1 if present Otherwise code 0	• "Evan had to pay off his debts" • "Evan needed to support his gambling habit • "Evan wanted the money"
Correction Denial	Message denies or expresses doubt that the correction is true. To deny the correction, *message must show awareness of correction*. **If any part of the correction is denied code as 1**.	1 if present otherwise code 0	• "Evan was supposed to be out of town but he wasn't" • They said Evan was at work but he drove back without anyone noticing.

APPENDIX D

Measurement Model Factor Loadings

Table 2.

Measurement Model Factor Loadings

Factor and Indicators	Factor Loadings								
	1	2	3	4	5	6	7	8	9
1. MP									
Item—Evan Harter was guilty	1.040								
Item—Evan Harter was responsible for the theft	0.857								
2. Misinformation Accessibility									
Item—Evan took the jewelry		0.536							
Item—Evan was guilty		0.650							
Item—Evan committed a crime		0.713							
3. Unethicality Accessibility									
Item—Evan is bad			0.706						
Item—Evan is sneaky			0.820						
Item—Evan is dishonest			0.593						
Item—Evan is corrupt			0.702						
4. Motive Accessibility									
Item—Evan wanted money				0.822					
Item—Evan desired money				0.639					
Item—Evan needed the jewelry				0.660					
5. Implicitness (Experimental Induction)					1.000				
6. Strong Correction (Experimental Induction)						1.000			
7. Moderate Correction (Experimental Induction)							1.000		
8. Misintegration (Coded Content)								1.000	
9. Recall (Coded Content)									1.000

Note: $n = 474$. Maximum likelihood estimation was used. The correction induction had 3 conditions: the uncorrected condition serves as the reference category. All cross-loadings were constrained to zero, consistent with classical test theory (Reynolds, 2020).

APPENDIX E

Regression Output

Table 3.

H1 Regression—Misinformation Accessibility

Dependent Variable – Misinformation Accessibility		
Predictor	Coefficient	
	B (SE)	β
Correction Strength Moderate = 0 Strong = 1	0.08 (0.07)	.11
Misinformation Type Explicit = 0 Implicit = 1	0.02 (0.06)	.02
Interaction Correction Strength by Misinformation Type	-0.03 (0.09)	-.04
Constant	0.95 (0.04)	-
F (3, 283) = 0.74, p = .53, R^2_{adj} = -.00		

Note. B = unstandardized, β = standardized OLS estimates. Sample includes corrected conditions only (n = 287).

Table 4.

H1 Regression—Unethicality Accessibility

Dependent Variable – Unethicality Accessibility		
Predictor	Coefficient	
	B (SE)	β
Correction Strength Moderate = 0 Strong = 1	-0.10 (0.08)	-.11
Misinformation Type Explicit = 0 Implicit = 1	-0.09 (0.07)	-.10
Interaction Correction Strength by Misinformation Type	0.10 (0.11)	.10
Constant	1.04 (0.05)	-
F (3, 283) = 0.85, p = .47, R^2_{adj} = -.00		

Note. B = unstandardized, β = standardized OLS estimates. Sample includes corrected conditions only (n = 287).

Table 5.

H1 Regression—Motive Accessibility

Dependent Variable – Motive Accessibility		
Predictor	Coefficient	
	B (SE)	β
Correction Strength Moderate = 0 Strong = 1	-0.01 (0.07)	-.01
Misinformation Type Explicit = 0 Implicit = 1	0.04 (0.07)	.05
Interaction Correction Strength by Misinformation Type	0.00 (0.10)	.00
Constant	0.91 (0.05)	-
F (3, 283) = 0.87, p = .87, R^2_{adj} = -.00		

Note. B = unstandardized, β = standardized OLS estimates. Sample includes corrected conditions only (n = 287).

Table 6.

H5 Regression

Dependent Variable – Misinformation Persistence		
Predictor	Coefficient	
	B (SE)	β
Misinformation Type Explicit = 0 Implicit = 1	0.54 (0.14)	.26
Misintegration Absent = 0 Present = 1	0.16 (0.18)	.07
Interaction Misinformation Type by Misintegration	-0.52 (0.26)	-.18
Constant	2.58 (0.10)	-
F (3, 284) = 4.94, p = .002, R^2_{adj} = .04		

Note. B = unstandardized, β = standardized OLS estimates. Sample includes corrected conditions only (n = 288).

Table 7.

RQ2 Regression

Dependent Variable – Misinformation Persistence		
Predictor	Coefficient	
	B (SE)	β
Misinformation Type Explicit = 0 Implicit = 1	0.54 (0.16)	.26
Correction Strength Moderate = 0 Strong = 1	0.10 (0.17)	.05
Interaction Misinformation Type by Correction Strength	-0.32 (0.24)	-.14
Constant	2.59 (0.11)	-
F (3, 284) = 4.08, p = .007, R^2_{adj} = .03		

Note. B = unstandardized, β = standardized OLS estimates. Sample includes corrected conditions only (n = 288).

Table 8.

H6 Regression

Dependent Variable – Misinformation Persistence		
Predictor	Coefficient	
	B (SE)	β
Correction Strength Moderate = 0 Strong = 1	-0.06 (0.15)	-.03
Misintegration Absent = 0 Present = 1	-0.16 (0.19)	-.03
Interaction Correction Strength by Misintegration	0.12 (0.27)	.04
Constant	2.88 (0.10)	-
F (3, 284) = 0.30, p = .83, R^2_{adj} = -.01		

Note. B = unstandardized, β = standardized OLS estimates. Sample includes corrected conditions only (n = 288).

Table 9.

Motive Accessibility by Correction Strength Interaction

Dependent Variable – Misinformation Persistence		
Predictor	Coefficient	
	B (SE)	β
Correction Strength Uncorrected = 0 Moderate/Strong = 1	0.56 (0.10)	.25
Motive Accessibility (mean centered)	-0.00 (0.17)	-.00
Interaction Correction Strength by Motive Accessibility	0.48 (0.22)	.16
Constant	3.41 (0.09)	-
F (3, 424) = 14.43, $p < .001$, $R^2_{adj} = .09$		

Note. B = unstandardized, β = standardized OLS estimates. Sample includes corrected and uncorrected conditions (n = 429).

Table 10.

Motive Accessibility by Motive Integration Interaction

Dependent Variable – Misinformation Persistence		
Predictor	Coefficient	
	B (SE)	β
Motive Integration Absent = 0 Present = 1	0.63 (0.13)	.29
Motive Accessibility (mean centered)	0.33 (0.15)	.13
Interaction Motive Integration by Motive Accessibility	0.77 (0.36)	.13
Constant	2.66 (0.07)	-
F (3, 283) = 13.15, $p < .001$, $R^2_{adj} = .11$		

Note. B = unstandardized, β = standardized OLS estimates. Sample includes corrected conditions only (n = 287).

Table 11.

Motive Attribution Beliefs by Correction Strength Interaction

Dependent Variable – Misinformation Persistence		
Predictor	Coefficient	
	B (SE)	β
Correction Strength Uncorrected = 0 Moderate/Strong = 1	-0.39 (0.29)	-.17
Motive Attribution Belief	1.00 (0.30)	.27
Interaction Correction Strength by Motive Attribution Belief	-0.24 (0.36)	-.10
Constant	2.65 (0.24)	-
F (3, 424) = 19.14, $p < .001$, $R^2_{adj} = .12$		

Note. B = unstandardized, β = standardized OLS estimates. Sample includes corrected and uncorrected conditions (n = 428).

Table 12.

Unethicality Attribution Belief by Correction Strength Interaction

Dependent Variable – Misinformation Persistence		
Predictor	Coefficient	
	B (SE)	β
Correction Strength Uncorrected = 0 Moderate/Strong = 1	-0.71 (0.29)	-.32
Unethicality Attribution Belief	0.91 (0.21)	.32
Interaction Correction Strength by Unethicality Attribution Belief	0.39 (0.25)	.13
Constant	2.96 (0.13)	-
F (3, 424) = 45.29, $p < .001$, $R^2_{adj} = .24$		

Note. B = unstandardized, β = standardized OLS estimates. Sample includes corrected and uncorrected conditions (n = 428).

Table 13.

Predictors of Misinformation Persistence (Alternative Indicator)

Dependent Variable – Misinformation Persistence (Alternative Indicator)		
Predictor	Coefficient	
	B (SE)	β
Correction Strength Uncorrected = 0 Moderate/Strong = 1	-0.47 (0.10)	.08
Motive Accessibility	0.08 (0.07)	.08
Interaction Correction Strength by Motive Accessibility	0.23 (0.09)	.28
Constant	0.56 (0.07)	-
F (3, 424) = 21.57, $p < .001$, $R^2_{adj} = .13$		

Note. B = unstandardized, β = standardized OLS estimates. Sample includes corrected and uncorrected conditions (n = 428).

REFERENCES

REFERENCES

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*, 261-295. https://doi.org/10.1016/S0022-5371(83)90201-3

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language, 49*, 415-445. https://doi.org/10.1016/j.jml.2003.08.006

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*, 1-21. https://doi.org/10.3758/BF03209454

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829-839. https://doi.org/10.1038/nrn1201

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1-29. https://doi.org/10.1146/annurev-psych-120710-100422

Balkin, R. S., & Sheperis, C. J. (2011). Evaluating and reporting statistical power in counseling research. *Journal of Counseling & Development, 89*, 268-272. https://doi.org/10.1002/j.1556-6678.2011.tb00088.x

Benjamin Clarke, A. J., & Butler, L. T. (2008). Dissociating word stem completion and cued recall as a function of divided attention at retrieval. *Memory, 16*, 763-772. https://doi.org/10.1080/09658210802261116

Berscheid, E., Graziano, W., Monson, T., & Dermer, M. (1976). Outcome dependency: Attention, attribution, and attraction. *Journal of Personality and Social Psychology, 34*, 978-989. https://doi.org/10.1037/0022-3514.34.5.978

Black, J. B., & Bern, H. (1981). Causal coherence and memory for events in narratives. *Journal of Verbal Learning and Verbal Behavior, 20*, 267-275. https://doi.org/10.1016/S0022-5371(81)90417-5

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306-307. https://doi.org/10.1207/s15327752jpa4803_13

Cacioppo, J. T., Von Hippel, W., & Ernst, J. M. (1997). Mapping cognitive structures and processes through verbal content: The thought-listing technique. *Journal of Consulting and Clinical Psychology, 65*, 928-940. https://doi.org/10.1037/0022-006X.65.6.928

Chan, M. P. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science, 28*, 1531-1546. https://doi.org/10.1177/0956797617714579

Delia, J. G. (1977). Constructivism and the study of human communication. *Quarterly Journal of Speech, 63*, 66-83. https://doi.org/10.1080/00335637709383368

Ecker, U. K., 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*, 292-304. https://doi.org/10.3758/s13421-013-0358-x

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*, 570-578. https://doi.org/10.3758/s13423-011-0065-1

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

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. https://doi.org/10.3758/BF03193146

Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strengths: Antecedents and consequences* (pp. 247–282). Mahwah, NJ: Erlbaum.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *The Journal of Abnormal and Social Psychology, 58*, 203-210. https://doi.org/10.1037/h0041593

Fuster, J. M. (1997). Network memory. *Trends in Neurosciences, 20*, 451-459. https://doi.org/10.1016/S0166-2236(97)01128-4

Gawronski, B. (2012). Back to the future of dissonance theory: Cognitive consistency as a core motive. *Social Cognition, 30*, 652-668. https://doi.org/10.1521/soco.2012.30.6.652

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650-669. https://doi.org/10.1037/0033-295X.103.4.650

Heider, F. (1958). *The psychology of interpersonal relations*. Hillsdale, NJ: Lawrence Erlbaum. https://doi.org/10.1037/10628-000

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*, 1420-1436. https://doi.org/10.1037/0278-7393.20.6.1420

Kardes, F. R. (1988). Spontaneous inference processes in advertising: The effects of conclusion omission and involvement on persuasion. *Journal of Consumer Research, 15*, 225-233. https://doi.org/10.1086/209159

Keenan, J. M., Baillet, S. D., & Brown, P. (1984). The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior, 23*, 115-126. https://doi.org/10.1016/S0022-5371(84)90082-3

Kelley, H. H., & Michela, J. L. (1980). Attribution theory and research. *Annual review of psychology, 31*, 457-501. https://doi.org/10.1146/annurev.ps.31.020180.002325

Kendeou, P., Smith, E. R., & O'Brien, E. J. (2013). Updating during reading comprehension: Why causality matters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 854-865. https://doi.org/10.1037/a0029468

Lang, A. (2009). The limited capacity model of motivated mediated message processing. *The SAGE handbook of media processes and effects*, 193-204.

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

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*, 764-766. https://doi.org/10.1016/j.jesp.2013.03.013

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflexion and reflection: A social cognitive neuroscience approach to attributional inference. In M. P. Zanna (Ed.), *Advances in experimental social psychology*, (pp. 199–249). Academic Press. https://doi.org/10.1016/S0065-2601(02)80006-5

Maheshwari, S. (2016). How fake news goes viral: A case study. The New York Times. Retrieved from http://www.nytimes.com/2016/11/20/business/media/how-fake-news-spreads.html

Malle, B. F. (2011). Time to give up the dogmas of attribution: An alternative theory of behavior explanation. In *Advances in experimental social psychology* (Vol. 44, pp. 297-352). Academic Press. https://doi.org/10.1016/B978-0-12-385522-0.00006-8

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry, 25*, 147-186. https://doi.org/10.1080/1047840X.2014.877340

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology, 33*, 101-121. https://doi.org/10.1006/jesp.1996.1314

McGuire, W. J. (1960). Cognitive consistency and attitude change. *The Journal of Abnormal and Social Psychology, 60*, 34-353. https://doi.org/10.1037/h0048563

Nelson, D. L., Kitto, K., Galea, D., McEvoy, C. L., & Bruza, P. D. (2013). How activation, entanglement, and searching a semantic network contribute to event memory. *Memory & Cognition, 41*, 797-819. https://doi.org/10.3758/s13421-013-0312-y

Noar, S. M., Harrington, N. G., & Aldrich, R. S. (2009). The role of message tailoring in the development of persuasive health communication messages. *Annals of the International Communication Association, 33*, 73-133. https://doi.org/10.1080/23808985.2009.11679085

Nyhan B., & Reifler J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior, 32*, 303-330. https://doi.org/10.1007/s11109-010-9112-2

Petty, R. E., Briñol, P., & DeMarree, K. G. (2007). The Meta–Cognitive Model (MCM) of attitudes: Implications for attitude measurement, change, and strength. *Social Cognition, 25*, 657-686. https://doi.org/10.1521/soco.2007.25.5.657

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510-532. https://doi.org/10.1037/0033-2909.114.3.510

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111*, 159-182. https://doi.org/10.1037/0033-295x.111.1.159

Reder, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification. *Psychological Review, 89*, 250-280. https://doi.org/10.1037/0033-295X.89.3.250

Reeves, B., Yeykelis, L., & Cummings, J. J. (2016). The use of media in media psychology. *Media Psychology, 19*, 49-71. https://doi.org/10.1080/15213269.2015.1030083

Reynolds, R. M. (2018). *Mechanisms of misinformation persistence: Sources of communicative and cognitive influence* (Unpublished preliminary paper). Michigan State University, East Lansing, MI.

Reynolds, R. M. (in press). Factor analysis, exploratory and confirmatory. *International Encyclopedia of Media Psychology*. Wiley & Sons.

Rhodes, N., Roskos-Ewoldsen, D. R., Edison, A., & Bradford, M. B. (2008). Attitude and norm accessibility affect processing of anti-smoking messages. *Health Psychology, 27*, S224-S232. https://doi.org/10.1037/0278-6133.27.3(Suppl.).S224

Rholes, W. S., & Pryor, J. B. (1982). Cognitive accessibility and causal attributions. *Personality and Social Psychology Bulletin, 8*, 719-727. https://doi.org/10.1177/0146167282084019

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, and Cognition*, *42*(1), 1-13. https://doi.org/10.1037/xlm0000155

Roskos-Ewoldsen, D. R. (1997). Attitude accessibility and persuasion: Review and a transactive model. *Annals of the International Communication Association, 20*, 185-225. https://doi.org/10.1080/23808985.1997.11678942

Roskos-Ewoldsen, D. R., & Fazio, R. H. (1997). The role of belief accessibility in attitude formation. *Southern Journal of Communication, 62*, 107-116. https://doi.org/10.1080/10417949709373045

Roskos-Ewoldsen, D. R., Roskos-Ewoldsen, B., & Carpentier, F. D. (2009). Media priming: An updated synthesis. In J. Bryant, & M. B. Oliver (Eds.), *Media Effects: Advances in Theory and Research* (pp. 74-93). New York: Routledge.

Simon, H. A. (1972). Theories of bounded rationality. *Decision and Organization, 1*, 161-176.

Song, H., & Ewoldsen, D. R. (2015). Metacognitive model of ambivalence: The role of multiple beliefs and metacognitions in creating attitude ambivalence. *Communication Theory, 25*, 23-45. https://doi.org/10.1111/comt.12050

Smith, C. T., De Houwer, J., & Nosek, B. A. (2013). Consider the source: Persuasion of implicit evaluations is moderated by source credibility. *Personality and Social Psychology Bulletin, 39,* 193-205. https://doi.org/10.1177/0146167212472374

Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology, 39,* 549-562. https://doi.org/10.1016/S0022-1031(03)00059-3

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

van den Broek, P., Beker, K., & Oudega, M. (2015). Inference generation in text comprehension: Automatic and strategic processes in the construction of a mental representation. In E. J. O'Brien, A. E. Cook, & R. F. Lorch, Jr. (Eds.), *Inferences during reading* (p. 94–121). Cambridge University Press.

Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs, 85,* 423-441. https://doi.org/10.1080/03637751.2018.1467564

Walter, N., & Tukachinsky, R. (2020). A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it? *Communication Research, 47,* 155-177. https://doi.org/10.1177/0093650219854600

Weiner, B. (2008). Reflections on the history of attribution theory and research: People, personalities, publications, problems. *Social Psychology, 39,* 151-156. https://doi.org/10.1027/1864-9335.39.3.151

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record, 58,* 475-482. https://doi.org/10.1007/BF03395630

Wongpakaran, N., Wongpakaran, T., Wedding, D., & Gwet, K. L. (2013). A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Medical Research Methodology, 13,* 61. https://doi.org/10.1186/1471-2288-13-61

Yonelinas, A. P. (2013). The hippocampus supports high-resolution binding in the service of perception, working memory and long-term memory. *Behavioural Brain Research, 254,* 34-44. https://doi.org/10.1016/j.bbr.2013.05.030