Examining the Role of Social Information in the Relationship Between Repetition and Belief

By

Raunak M. Pillai

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

May 10, 2024

Nashville, Tennessee

Lisa K. Fazio, Ph.D.

Bethany Rittle-Johnson, Ph.D.

Melissa Duff, Ph.D.

Sarah Brown-Schmidt, Ph.D.

ACKNOWLEDGEMENTS

Throughout graduate school, I have had the fortune of knowing and working with a number of extraordinary people without whom this dissertation would not be possible. First and foremost, I would like to thank my advisor, Dr. Lisa Fazio. Seven years ago, as a sophomore with at most a perfunctory knowledge of research methods and a vague interest in studying how people learn, I reached out to join her lab as a research assistant. Since then, I have honed my research interests, developed (I hope) an ability to identify and experimentally address research questions and, perhaps most surprisingly, learned to not hate writing so much. I name these developments as just a small sample of the areas in which I have grown thanks to Lisa's mentorship over the many years.

I would also like to thank the rest of my committee members, Drs. Bethany Rittle-Johnson, Sarah Brown-Schmidt, and Melissa Duff for providing invaluable feedback that has guided the shape and direction of this dissertation. Beyond my committee, credit is also due to several other professors who have played a critical role in my training through coursework. My thanks go out especially to Dr. Sean Polyn for making me rethink what a good cognitive psychological explanation looks like and for making coding less scary, to Dr. Duane Watson for exposing me to the intricacies of language comprehension/production, and to Dr. Nicole Joseph for giving me the tools and space to approach my own area of research with a critical, historically-oriented lens. Outside of my coursework, I have also been fortunate to work under the guidance of Dr. Russell Fazio on the experiments that make up Chapter 2 of this dissertation; he has been a fount of knowledge and ideas for this project, and his enthusiasm for this work has been downright infectious.

Credit is also due to the many senior graduate students with whom I have overlapped. To Bethany Gardner for helping me finally understand how to interpret lower-order coefficient terms, for sharing with me meticulously annotated analysis code, and for somehow always having the right meme or Tumblr screenshot for any situation. To Yev Diachek for figuring out how to calculate Bayes Factors for mixed-effects models so that I didn't have to myself, for fielding innumerable questions about program requirements, and for being a source of joy and novel turns of phrase. Lastly, to Clair Hong and Darren Yeo for being model senior lab members.

To Emily Burgess, you have been a constant source of support. At every step of this journey in which I did not believe in myself, you believed in me, and for that I am eternally grateful. To Eeshan Hasan, I knew you would challenge me to grow personally and intellectually from the day we met at orientation, and I am glad to report that you still do, even after you abandoned me for Indiana. To Katrina Rbeiz and Ebony Pearson, oh man, oh man, where do I even begin. You both have been my rocks through tumultuous times and are continually sources of laughter and new memories.

My thanks go out also to my Mom, Dad, and brother for providing support in getting to this stage of my academic career and life.

This work was made possible with the help of research assistance from Shaun Karakkattu (Chapter 2), Leigh Farah (Chapter 3). Finally, this work as also made possible by funding from the National Science Foundation Graduate Research Fellowship Program under Grant No. 1937963; the views presented in this dissertation are my own and do not necessarily reflect the views of the National Science Foundation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction: The Effects of Repetition on Belief

"You really think someone would do that? Just go on the internet and tell lies?"

-Buster Baxter in *Arthur* (TV series) (Bailey, 2005)

From search engines to digital encyclopedias to social media platforms, the technological advances of the digital age have rapidly enhanced our ability to access knowledge and share it with one another. At the same time, these very technologies have also been implicated in the creation and spread of *false and misleading* information (Aslett et al., 2023; Garrett, 2011; Kumar et al., 2016; Vosoughi et al., 2018). Of course, misinformation is not a 21st-century problem. In 1782, Benjamin Franklin fabricated an issue of a Boston newspaper to claim that the Seneca people had scalped hundreds of colonists by order of King George (Parkinson, 2016). In the 1800s, White slave-owners in the American South spread false reports of uprisings and crimes by enslaved peoples (Soll, 2016). And in the mid-late 1900s, tobacco companies publicly downplayed scientific evidence of the health risks of smoking cigarettes (Cummings et al., 2007). While misinformation has long been present in human society, the 21st century has seen rapid transformations to human information flow that have reduced barriers to its creation and increased the speed and distance of its spread (Bak-Coleman et al., 2021). These changes make discerning truth from falsehood a routine challenge in everyday life. The overarching motivation behind this dissertation is thus to contribute to our understanding of how humans make these judgements of truth.

Undoubtedly, a number of factors affect what humans perceive to be true (see Brashier & Marsh, 2020 for a review). Of these, one perhaps surprising factor is repetition: people are more likely to believe statements they have heard before, a phenomenon known as the "illusory truth effect" (Hasher et al., 1977). In a typical study, participants first read a set of statements, and then, after some variable amount of delay, evaluate the truth (e.g., on Likert scale) of these same statements, intermixed with some new ones. In these studies, the typical finding is that the repeated statements are given higher truth ratings than the statements being seen for the first time.

Of course, it is a truism that repeating a statement does not *actually* make it any more true. And yet, dozens of psychological studies on the illusory truth effect show that people still reliably *think* that repeatedly encountered information is more likely to be true. Meta-analytic evidence indicates the effect is of medium size ($d = 0.49$; Dechêne et al., 2010). However, this figure may be an underestimate. Most studies warn people about the presence of false information prior to exposure, but, when people are not forewarned (as they often are not in real life), the effect doubles in size (Jalbert et al., 2020). The illusory truth effect is not only substantial in size, it is also long-lasting: repeating a claim even weeks after first exposure can increase belief (Henderson, Simons, et al., 2021). The effect also generalizes across a number of different types of stimuli, from trivia statements (Hasher et al., 1977) to political news headlines (Pennycook et al., 2018) to claims about consumer products (Roggeveen & Johar, 2002). Finally, the illusory truth effect occurs across the lifespan—from childhood (Fazio & Sherry, 2020) to late adulthood (Brashier et al., 2017; Lyons, 2023). In sum, the tendency to consider repeatedly-seen information truer is a core feature of how humans evaluate truth. Why, then, does repetition affect belief?

As I will describe in further detail below, prominent current accounts suggest that the illusory truth effect depend on low-level cognitive cues (Unkelbach et al., 2019). For instance, repeated statements feel easier to process (i.e., are more fluent), and people infer that more easily processed information is more likely true. Importantly, these theories are based on laboratory studies that largely investigate the effects of a single prior exposure to a statement without any identifiable source (see Henderson, Westwood, et al., 2021 for a systematic map of the literature). Of course, in real life, we repeatedly encounter statements in a social context. Statements often come from distinct sources whom we may trust or distrust. Further, we often see statements multiple times, either from a single repetitive voice or from an array of unique sources. Critically, there are good reasons, which I will also review below, to believe that these social factors of trustworthiness and consensus should affect what people deem true.

In this dissertation, I aim to examine whether and how higher-level social factors—namely, source trustworthiness and social consensus—affect the link between repetition and belief. In doing so, I will integrate theoretical perspectives on human truth judgements that emphasize low-level cognitive cues and higher-order social inferences, respectively.

Theories of the Illusory Truth Effect

A number of theoretical accounts have been put forth to explain why repeated statements seem truer, and all invoke low-level cognitive mechanisms (see Unkelbach et al., 2019 for a review). Here, I explain and discuss the evidence supporting the three most prominent accounts found in the literature: familiarity, processing fluency, and referential coherence.

Familiarity

One early explanation of the illusory truth effect relies on the idea that repeatedly-seen information is more familiar (Arkes et al., 1989). *Familiarity* generally refers to an automatic, context-free signal in memory that indicates a stimulus has been encountered before (see Yonelinas, 2002 for a review). The familiarity construct comes from dual-process theories of episodic memory which contrast familiarity with *recollection*, a form of memory involving reinstatement of vivid details about the original encounter with the stimulus. In a classic demonstration of the distinction, Mandler (1980) described the experience of seeing someone on the bus and feeling confident that you have seen them before without being sure where. Initially, you might have nothing more than this feeling, but after searching your memory, you may remember seeing this person as the butcher at your local supermarket. Where the initial feeling associated with seeing the face is best described as familiarity-driven, the latter experience of identifying the specific source of the face corresponds to the process of recollection.

Because familiarity with a statement is a memorial experience that lacks context, people may misattribute the source of this familiarity. Specifically, participants in a typical illusory truth experiment may incorrectly infer that a repeated statement seems familiar because they have seen it prior to the experiment. This misattribution of familiarity may make the statement seem truer. If a participant thinks they saw a statement prior to the experiment, they might think it is the kind of statement other people agree with and find true enough to repeat themselves (see, e.g., Arkes et al., 1991; Hawkins & Hoch, 1992).

Evidence for the familiarity account comes from studies that dissociate familiarity and recollection. In one set of studies, participants were exposed to statements from one of two voices and were instructed that statements from one voice were true and statements from the other voice were false (Begg et al., 1992). When these statements were seen again, participants generally regarded them as truer than new statements due to their familiarity—except when they could explicitly recall that the original voice was false. Follow-up studies dissociated these effects of familiarity and recollection by showing that they could be independently manipulated. For instance, dividing people’s attention during the original exposure to the statements reduced the effect of recollection of the voices on truth judgements, but left intact the effect of familiarity due to prior exposure.

Fluency

A second, more recent perspective on the illusory truth effect is that it occurs because repeated statements are more *fluently processed* (Reber & Schwarz, 1999). Processing fluency is a metacognitive experience of the ease with which a given mental process occurs (Oppenheimer, 2008). It is thought to arise spontaneously from a range of mental operations such as perception, memory retrieval, or language processing (Alter & Oppenheimer, 2009). In the context of truth judgements, two forms of fluency have been identified as particularly influential. First, repetition may increase *perceptual fluency*—the subjective ease associated with the visual or auditory processing involved in perceiving a statement. Second, repetition may increase *conceptual fluency*—the ease with which the meaning of a statement is processed. While both kinds of fluency are influential, recent work has identified conceptual fluency as more influential than perceptual fluency for judgements of truth (Vogel et al., 2020).

The core line of evidence for the fluency account is that other known fluency manipulations (besides repetition) also increase perceived truth. For instance, presenting statements in a high-contrast font color (Reber & Schwarz, 1999), making statements rhyme (McGlone & Tofighbakhsh, 2000), and having participants read passages about a semantically related topic beforehand (Arkes et al., 1991) are all manipulations known to increase processing fluency, and all of these manipulations also increase belief. Thus, ease of processing is one determinant of truth judgements, and may also explain why repetition makes statements seem truer.

This discussion begs the question: why does fluency affect judgements of truth? Current models emphasize two non-mutually-exclusive reasons (Unkelbach & Greifeneder, 2013). First, people may have implicitly learned from experience over the course of their life that fluency correlates with truth. For instance, in school, through caregivers, and in books, speakers may often repeat true information much more than false information, providing people a basis to acquire an association between fluency and truth (see Reber & Unkelbach, 2010 for further discussion on why true information may more likely be fluent). Consistent with this learning account, people can also *unlearn* the relationship between fluency and truth through a feedback-learning task in which easily-processed information is often labelled as false (Unkelbach, 2007).

In addition to the learning account, a second possibility is that people explicitly hold beliefs about how to interpret fluency—namely, that fluently processed information is likely to be true. Evidence for this possibility comes from related work showing that directly manipulating how people interpret fluency can shape its effects on other judgements. For instance, when told that unpleasant times tend to be forgotten, people who experience difficulty while retrieving childhood memories (i.e., retrieval disfluency) think of their childhoods as more negative. However, this pattern reverses when people are given an opposite interpretation of fluency (that *pleasant* experiences tend to be forgotten). After hearing this framing, disfluent retrieval experiences make people think of their childhoods as more *positive* (Winkielman & Schwarz, 2001). This work shows that the link between fluency and judgement can depend on people's explicit interpretations of fluency. In the context of repetition and truth judgements, people may interpret perceptual/conceptual fluency as a signal that the statement has been seen before and thus repeated because it is endorsed by others (Schwarz, 2004).

Referential Coherence

While the fluency account is the dominant account in the literature, a more recent competing account has been proposed. This account argues that the underlying mechanism behind the illusory truth effect is changes in the coherence of people's semantic representations of the information in a given statement (Unkelbach & Rom, 2017). According to this theory, the information in a given statement corresponds to underlying representations in a semantic memory network, which may be modelled as a propositional, symbolic network.[1] For instance, for a statement like "the cheetah is the fastest land animal," the network contains nodes referring to the concepts "cheetah," "fastest," "land," and "animal."

The nodes in this semantic network are associated by links with varying levels of strength representing the coherence of the proposition that they collectively represent. In turn, judgements of the truth of a statement are a direct function of the number and strength of the links between the semantic memory nodes corresponding to a statement. Critically, the account argues that each time a statement is encountered, the connections between these statements' corresponding nodes in the semantic memory network are immediately strengthened. Thus, the illusory truth effect arises because previously-encountered statements have more coherently linked memorial representations, and these coherent representations give rise to greater perceptions of truth.

In line with this referential coherence account, tasks involving the encoding of information in relation to oneself increase the effects of repetition on belief (relative to a more superficial task like counting the number of vowels in a statement; Unkelbach & Rom, 2017, Experiment 3). By encoding information in relation to oneself, the theory argues, the network representing the statement grows larger (i.e., by including nodes pertaining to one's self-concept) and this larger network, in turn, results in greater perceptions of truth. Also in line with the referential coherence account, hearing a statement that was implied from an earlier statement increases belief in it (Unkelbach & Rom, 2017, Experiment 4), as the references for the implied statement should be more coherent in memory due to prior exposure of the related statement.

---

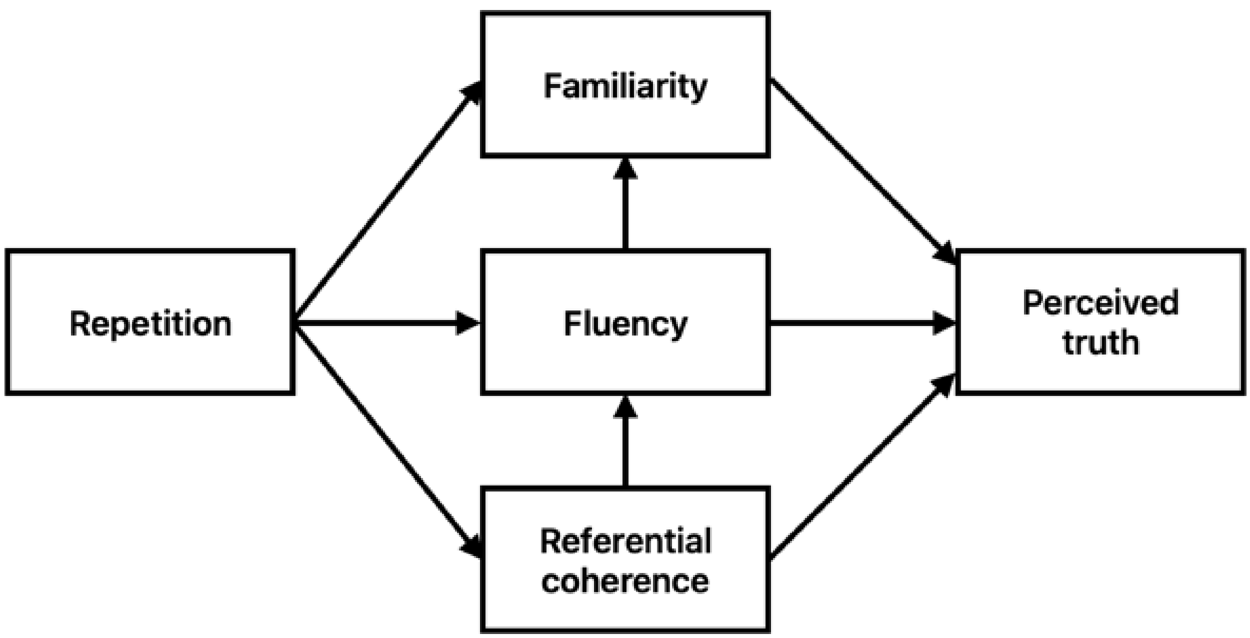
[1] Note that the authors do not make an explicit commitment to this formalism, arguing that the theory may just as well be instantiated in a sub-symbolic network (Unkelbach & Rom, 2017, p. 111).

Comparing and Contrasting the Core Explanations

All of the explanations reviewed thus far—familiarity, fluency, and referential coherence—share a similar structure. Repeatedly encountered statements evoke a low-level, readily accessible cognitive cue, and these cues in turn enhance perceptions that the statement is true. Figure 1.1 below provides a schematic outline of these explanations. How do these explanations differ, then? First, these cues have different relationships to one another, and second, they affect belief through different processes.

Figure 1.1

Relationship Between Explanations of the Illusory Truth Effect



Note. Figure adapted from Unkelbach et al. (2019).

Relationship between constructs. Familiarity, fluency, and referential coherence are all enhanced by repetition and, in turn, can increase perceptions that a statement is true. However, this broad summary misses the ways in which these cues are connected with one another, as summarized in the middle section of Figure 1.1. There are two connections of note. First, familiarity can be thought of as an interpretation of fluency: the easier a statement is to process, the more it may seem that it has been encountered before somewhere (Jacoby, Kelley, & Dywan, 1989). In this way, fluency may be a more fundamental, parsimonious explanation for the illusory truth effect than familiarity, as it may give rise both to judgements of truth and to familiarity (which may then also shape judgements of truth). Second, the coherence of people's semantic representations for a statement can give rise to fluency, making the statement conceptually easier to process. This makes the referential coherence account an even simpler explanation still than the fluency and familiarity accounts. However, the value of this simplification is unclear, as the referential coherence account does not entirely eschew fluency as a construct involved in truth judgements. Indeed, the referential account argues that processing fluency is needed to explain why manipulations other than repetition (e.g., rhyming or font color) increase belief (Unkelbach & Rom, 2017, p. 111).

Relationship between cues and belief. The other key difference between the explanations lies in the process by which the respective cues affect belief (rightmost arrows in Figure 1.1). Consider, first, the familiarity and fluency accounts. There is no inherent reason either of these cues should increase belief. Why, then, do they have this effect? As reviewed earlier, both of these explanations involve a process of interpretation in order for the cue to affect belief. For instance, people may infer that familiarity means they have seen the statement prior to the experiment and then then interpret this prior exposure as evidence of the statement's validity (Arkes et al., 1991). Further, people may interpret fluency as evidence of truth because they have learned from a lifetime of experiences that easily processed information tends to be true (Reber & Unkelbach, 2010).

In contrast to the familiarity and fluency accounts, the referential coherence account posits a cue that requires no interpretation at all to affect belief. According to this theory, referential coherence is inherently cognitively marked as a measure of the truth value of information (Unkelbach & Rom, 2017). As such, this account purportedly offers even more parsimonious explanation than the fluency account, because it does not leave open the further question of why people interpret familiarity or fluency as cues to truth (Unkelbach & Rom, 2017). However, as described above, the assumptions needed to link fluency with belief are not so onerous. This link may arise as long as people are generally exposed to more true than false information, an assumption that likely holds assuming communicators speak with good intent (Reber & Unkelbach, 2010).

Summary. To conclude, current theories of the illusory truth effect are largely compatible, emphasizing the tendency for truth judgements to be influenced by low-level cognitive cues. However, they differ in two respects. First, the referential coherence has been said to underly fluency, and fluency to underly familiarity. Second, where the familiarity and fluency accounts involve extra assumptions about why people interpret these cues as relevant to truth judgements, the referential coherence account does not involve this extra step. As a result, the accounts increase in complexity from the referential coherence to fluency to familiarity accounts.

For the purposes of this dissertation, I will articulate key hypotheses in terms of the fluency account, as this account a) offers a greater level of parsimony than the familiarity account and b) is able to account for a broader range of phenomena (e.g., effects of perceptual fluency on truth judgements) than the referential coherence account. However, for the questions being tested, the other accounts largely make identical predictions.

Social Factors Relevant to Belief

The empirical basis for the theories reviewed thus far are studies which largely examine the effects of a single, verbatim exposure to a statement without any identifiable source (Henderson, Westwood, et al., 2021). However, in real life, people repeatedly encounter information in a defined social context—one in which there is typically some identifiable source whom a) the speaker may hold a certain attitude towards and b) who may or may not be the same source who later repeats the statement. The accounts offered above have little to offer in the way of predictions about how variations in this social context will affect belief. They provide only the coarse expectation that as long as information is repeated and grows more fluent, belief should increase. However, there are good reasons to expect source factors to matter for belief. Here, I review two such factors for which there is relevant evidence—one pertaining to *who* says a statement (source trustworthiness) and the other pertaining to *how many different people* repeat a statement (social consensus).

Source Trustworthiness

In daily life, the information that repeatedly reaches us tends to come from distinct sources—a friend, a news organization, a suspicious website. In each of these cases, we likely hold some evaluation of the source as a reliable or unreliable place to get information. These evaluations of the trustworthiness of a source are known to affect how people perceive their messages (see Pornpitakpan, 2004 for a review). But does source trustworthiness affect the relationship between repetition and belief? In particular, does repetition make statements seem more true even when that statement comes from someone known to state falsehoods?

Current cognitive theories of the illusory truth effect suggest that the answer is yes: repetition should increase belief, irrespective of social evaluations like the trustworthiness of the speaker. These theories all focus on low-level cognitive cues that are thought to arise quickly and spontaneously— impressions of familiarity occur automatically (Yonelinas, 2002), processing fluency is experienced as a "by-product [of] mental operations" (Greifeneder & Unkelbach, 2013), and spreading activation through semantic networks leads to rapid evaluations of the coherence of incoming information (Unkelbach & Rom, 2017). Accordingly, these theories suggest that our cognitive systems readily signal to us when we are in the presence of repeated information through familiarity, fluency, or referential coherence, providing easily accessible cues for truth that may be difficult to discount.

Indeed, empirically, repetition increases belief even when people have other, contradictory cues for truth at their disposal. Repetition increases belief when the information contradicts one's prior knowledge (Fazio, 2020; Fazio et al., 2015), runs counter to one's political views (Pennycook et al., 2018), or seems implausible (Fazio et al., 2019; Lacassagne et al., 2022). Repetition even increases belief when a reliable adviser explicitly tells people that a statement is false (Unkelbach & Greifeneder, 2018). To be clear, each of these manipulations does affect truth judgments. For example, people give lower truth ratings for implausible statements or ones that contradict their prior knowledge. However, repetition increases belief even when confronted with these contradictory cues. In statistical terms, there is a main effect of prior knowledge (higher truth ratings for plausible claims) and a main effect of repetition (higher truth ratings for repeated claims) but no interaction.

These findings can be easily accommodated by the theories described earlier: the effects of repetition on belief stem from low-level cognitive processes and are thus unaffected by higher-level processes such as retrieving prior knowledge or engaging in politically motivated reasoning. By extension, these theories predict that repetition should increase belief even in the face of contradictory social information, like evaluations of the source as untrustworthy. That is, since repeating a statement automatically increases its fluency, familiarity or coherence, repetition should increase belief, regardless of who does the repeating.

However, there are also reasons to expect source trustworthiness to alter the effects of repetition on belief. First, people may be able to disregard fluency as a cue for truth when they remember that the fluency experience stems from repetition by an unreliable source. In related work, repeated exposure to non-famous names (e.g., "Sebastian Weisdorf") makes these names seem more famous, unless people remember having seen the name before (Jacoby, Kelley, Brown, et al., 1989; Jacoby, Woloshyn, et al., 1989). Presumably, seeing a name multiple times makes it more familiar in memory, and people infer that familiar names are more likely to be famous. However, when people explicitly recall seeing the name, they can attribute the experience of familiarity to the experimental context, rather than to general fame. Similarly, people may generally attribute the familiarity or fluency associated with a statement to it being true. However, if they are able to remember that the statement came from a source perceived as untrustworthy, they may be able to disregard fluency as a cue for truth.

Second, engaging with untrustworthy sources can trigger deeper, more skeptical information processing strategies (e.g., elaborating on alternatives to presented information; see Mayo, 2015 for a review). This sort of distrustful, accuracy-focused processing may, in turn, make repetition less influential on later belief. For instance, simply warning people that they may see false information prior to exposure reduces the effect of these exposures on later belief (Jalbert et al., 2020). Similarly, forcing people to directly consider the accuracy of claims (rather than their interestingness) can also make these exposures less impactful for subsequent belief (Brashier et al., 2020; Nadarevic & Erdfelder, 2014). In sum, people may spontaneously engage in skeptical information processing strategies (like thinking about how information could be false) when seeing information from an untrustworthy source. When people then see these statements again, their original doubts may be activated, reducing the effects of repetition on belief.

Overall, there are distinct, competing predictions about whether and how source trustworthiness moderates the effects of repetition on belief. On one hand, current cognitive theories of the effect predict source trustworthiness should not matter. On the other hand, source trustworthiness is known to affect cognition in ways that may disrupt the typical effects of repetition on belief. Chapter 2 reports the results of five experiments designed to examine these questions in further detail.

Social Consensus

Thus far, we have discussed the effects of a factor characterizing a single source of information. However, by definition, repetition involves multiple exposures to a statement, each potentially involving a different source or context of exposure. For instance, we may hear about an event separately from multiple witnesses or repeatedly from a single, incessant source. Assuming all sources are comparably trustworthy, how might this variation in the number of sources affect belief?

Current theories largely suggest this variation should not matter: by evoking low-level cognitive signals, repetition should increase belief to a similar degree—irrespective of how many unique people repeat the claim. On the other hand, classic research on conformity (Asch, 1956) and opinion formation (Festinger, 1954) suggests that people alter their thoughts and actions in line with what is socially agreed upon. More recent evidence echoes this basic social psychological insight. For instance, people are more confident in claims supported by multiple, independent sources of evidence than by claims supported multiple times a single source (Connor Desai et al., 2022). Relatedly, people are more likely to accept that humans are causing climate change after reading about the overwhelming scientific consensus (Lewandowsky et al., 2013). Importantly, people are not only sensitive to expert consensus—consensus among peers is also influential. For example, people are more likely to agree with simulated blogs and social media posts when there is strong agreement among the repliers or commentors beneath it (Lewandowsky et al., 2019; Ransom et al., 2021).

However, while it is clear that people are sensitive to social consensus, it is unclear whether this sensitivity will influence belief in repeatedly encountered statements. In past research, participants received information about social consensus in a single, uninterrupted experience, like reading a thread of comments under a post (e.g., Ransom et al., 2021), a summary statistic (e.g., that 97 of 100 experts agree; Lewandowsky et al., 2013) or a series of articles on the same topic (Connor Desai et al., 2022). In these cases, consensus is either directly communicated or can be easily comprehended by comparing adjacent pieces of information (e.g., noticing a comment agrees with the ones below it). Here, we are instead concerned with people's sensitivity to social consensus cues around *repeated* information, where consensus must be inferred across discrete instances of exposure to information. For example, imagine you are scrolling through a social media newsfeed, seeing the same news repeatedly, but separated by several other, unrelated posts. This situation requires that people a) track whether information was repeated by multiple sources or one in memory b) interpret this memory for multiple sources as a signal for broad agreement and c) use this perceived consensus as a cue for truth judgements. While past work justifies the latter assumption, it is less clear whether the first two processes will occur.

In sum, there is unclarity about how repetition affects belief when it implies varying degrees of social consensus. Current cognitive theories suggest this variation should not matter. By contrast, a number of studies have shown that social consensus is a powerful cue for belief.

However, this pattern has not been investigated in the context of repeatedly-seen claims, where information about consensus may be more difficult to track. Chapter 3 addresses these questions through a series of three experiments.

CHAPTER 2

Source Trustworthiness, Repetition, and Belief

Introduction

As reviewed above, repetition increases belief because repeatedly-encountered statements are easier, or more fluent, to process, and people infer that fluently processed information is more likely to be true (Reber & Schwarz, 1999). Using this fluency heuristic can help us arrive at quick, relatively accurate judgements of truth, as we likely repeatedly encounter more true than false information in daily life (Reber & Unkelbach, 2010). However, in some cases, fluency from prior exposure is not a reliable cue for truth—like when hearing statements from a suspected or known liar. The goal of this chapter is to investigate whether repetition still increases belief when statements come from untrustworthy sources, or whether repetition increases belief to a similar degree regardless of the source.

Past Research

Chapter 1 reviewed the theoretical arguments on either side of the question asked by this chapter. This section will focus more narrowly on two lines of prior empirical work that bear on this question, and it will describe how the present research will build on this past work.

First, some research has examined how repetition affects belief when people initially encounter statements in a context indicating that the statement is true or false (e.g., Begg et al., 1992; Unkelbach & Stahl, 2009). In these studies, participants are exposed to statements from one of two sources (e.g., male or female voices), and are instructed that statements from one source are always false, while those from the other source are always true. They are then asked to rate the truth of the repeated statements (alongside new ones) without any source information. These studies broadly show that repetition generally increases belief through automatic, low-level processes like fluency or familiarity, but that people can discount these effects through effortful recollection of the original context in which the statement was encountered. That is, when people remember the original source and that it was labelled “false”, repetition does not increase belief. However, when the source is forgotten (e.g. after a month-long delay) repetition again increases belief (Brown & Nix, 1996).

Second, other work has examined how fluency from prior exposure affects belief when people also have advice from a reliable source at the very moment they are judging truth. In these studies, people initially read a set of statements without any source cues and are then asked to evaluate the truth of repeated and new statements presented alongside advice by sources labelling the statement as true or false. Critically, participants are explicitly informed where these sources lie along a continuum of reliability from completely guessing (i.e., their advice is right 50% of the time) to completely accurate (i.e., 100% right). This work finds that, while people are less likely to believe statements when a reliable (versus unreliable) adviser indicates the statement is false, repetition increases belief similarly for statements repeated alongside advice from either adviser (Unkelbach & Greifeneder, 2018).

Overall, these studies shed insight into the cognitive processes by which source evaluations may matter—indicating, for instance, that effortfully recollecting source information can counteract the effects of repetition on belief. Critically, however, these studies provide participants information bearing directly on whether individual claims are true or false (e.g., instructions that statements from one voice are false, advice that a given statement is false). Instead, here we ask what happens when a statement comes from a source deemed untrustworthy. These evaluations of trustworthiness differ from past work in two key ways. First, these evaluations are social and interpersonal in nature. They concern evaluations of a speaker or source of a statement, rather than information about whether a given statement is true or false. Second, and relatedly, these social evaluations are less directly related to the truth of a statement. For instance, in the experiments by Begg et al. (1992), participants are explicitly told whether statements from one source are true or false, and in the experiments by Unkelbach & Greifeneder (2018), participants are directly told how valid a piece of advice is. By contrast, perceiving a speaker as trustworthy or untrustworthy is not completely diagnostic of whether any given assertion they offer is true or false. Reliable sources may err, and liars may stumble upon the truth, leaving it up to their audiences to decide how to relate perceptions of the speaker to perceptions of the information.

In the current studies, we attempted to capture the effects of receiving information from sources perceived as trustworthy or untrustworthy. In order to experimentally control trust in sources, we created two novel sources (Mr. Green and Mr. Red), and we introduce a source training task that allows participants to form social evaluations about their level of trust in each source. In this task, participants encounter statements from each source and are asked to decide whether the statement is true or false before receiving feedback indicating the truth of the statement. Critically, Mr. Green tends to state true information while Mr. Red tends to state false information, allowing participants to learn to regard the former as trustworthy and the latter as untrustworthy. We then examined how participants' beliefs are shaped by exposure to statements attributed to these two sources.

Present Research

Across five experiments, we examine whether and how source trustworthiness moderates the effects of repetition on belief. After completing the source training task described above, participants took part in a typical lab-based illusory truth paradigm, using a different set of statements than the ones used in the training phase. In Experiment 1, participants were first exposed to a series of statements without any source attributions, and then underwent a test phase in which they saw new and repeated statements attributed either to Mr. Red or Mr. Green. The key question was whether or not repetition would increase belief more when statements were repeated by Mr. Green than Mr. Red.

Experiments 2-5 then flipped the phase at which source information was presented. In an initial exposure phase, participants read a series of statements, with half attributed to Mr. Green and the other half to Mr. Red. Then, in a test phase, participants rated the truth of a series of statements presented without a source. Here, one third of statements were new, one third were originally stated by Mr. Green, and one third were originally stated by Mr. Red. Across experiments, we tested conditions under which sources cues were increasingly more salient (e.g., adding instructions to attend to source information in Experiments 3 & 4; making Mr. Green more and Mr. Red less trustworthy in Experiment 4) We expected that statements repeated by the trustworthy source (Mr. Green) would be perceived as more true than new statements, in line with the illusory truth effect. The main question across experiments was how truth ratings for statements repeated by the untrustworthy source (Mr. Red) would compare to ratings for new statements, and to those repeated by the trustworthy source (Mr. Green). In exploratory analyses, we also examine whether the effects of source may differ across individuals and whether this variability can be explained by predictors that capture theoretically-relevant processes such as trust in the two sources, memory for the source of the statements, and willingness to engage in thoughtful processing.

Open Practices

The hypotheses, design and analysis plan for all experiments were pre-registered. For Experiment 1, the pre-registration documents are available at the project's Open Science Framework (OSF) site, along with the materials, participant instructions, data, and analysis code for each experiment (https://osf.io/myc56). For Experiments 2-5, this information can be found at a separate OSF link (https://osf.io/z7bqy/?view_only=f418560e50c349a6ac1e68f7bad9ba0f).

Experiment 1

Experiment 1 is designed to examine whether repetition increases belief even when the statement is later repeated by an untrustworthy source. After learning to associate two sources with conveying true/false statements, participants were exposed to a series of statements from no source, rated the truth of these statements and new ones presented by either of the two sources, and then answered some questions about their attitudes towards the sources themselves. We hypothesized that there would be an interaction between source trustworthiness and repetition, such that the difference in perceived truth of new and repeated statements would be greater for statements coming from the trustworthy than the untrustworthy source.

Method

Experiments 1-2 received ethics approval under protocol #220388 from the Vanderbilt University Institutional Review Board.

Participants.

Statistical Power. We pre-registered a sample size of 192 based on an *a priori* power analysis conducted using the *Superpower* package in R (Lakens & Caldwell, 2021). A sample size of 192 participants provides 80% power to detect an interaction effect between source trustworthiness and repetition of size $\eta_p^2 = 0.044$, or $f = 0.216$.

This interaction effect size was calculated by specifying a pattern of means in *Superpower* such that, when a statement was presented by the trustworthy source, repetition increased truth ratings by 0.3 points (based on observed mean differences from past studies we have conducted; Pillai & Fazio, 2022), but when the statement was presented by an untrustworthy source, repetition only increased truth ratings by 0.1 points. In addition, we specified a standard deviation of 0.6 points and a correlation among repeated measures of 0.33, as found in similar past studies with similar materials and numbers of observations per participant per cell of the design (Pillai et al., under review).

Recruitment. One hundred and ninety-two adult participants on Amazon's Mechanical Turk platform (MTurk) were recruited to complete the experiment using CloudResearch (Litman et al., 2017). We restricted participation to U.S. residents and people with full color vision (as our source manipulation required participants to discern between red and green stimuli). To ensure data quality, we recruited participants from the approved participants list on CloudResearch (Peer et al., 2021) and blocked duplicate IP addresses.. While we had pre-registered that we would exclude any participants that a) failed two attention checks at the beginning of our survey (typing the name of a 2d image of a black and white cartoon animal and selecting a requested responses on a 5-point Likert question) or b) failed more than two of five Ishihara colorblindness test trials (Ishihara, 1917), no participants met either of these criteria for exclusion.

Demographics. The mean age of participants was 31.50 (*SD* = 13.24; Range = 20-79). Our final sample was predominantly White (80.73%, 6.77% Black, 5.73% Multiracial, 3.65% Asian, 0.52% Middle Eastern and North African, 1.04% Other, 1.56% not reporting) and non-Hispanic (83.85%, 13.55% Hispanic, 2.60% not reporting), and most participants were women (52.60 %, 44.80% men, 1.56 % nonbinary, 1.04 % not reporting). In addition, most participants (73%) held at least a college degree.

Design. This experiment manipulated repetition (repeated, new), source trustworthiness (trustworthy, untrustworthy), and statement truth (true, false) within-subjects. Repetition was counterbalanced by splitting items into two sets, with each set containing even numbers of true and false items. Participants were assigned to see one of these two sets during the exposure phase. Source trustworthiness was then counterbalanced by assigning participants to see the even-numbered items from the trustworthy source and the odd-numbered items from the untrustworthy source or vice versa. This ensured that for each participant, trustworthy and untrustworthy sources were attributed to an equal number of true/false and repeated/new items.

Materials. The key stimuli in this experiment were 56 relatively unknown trivia statements that were true (e.g., "Bullet was the name of Roy Roger's dog.") or false (e.g., "Napoleon was born on the island of Sardinia."). Forty statements were adapted from Fazio (2020) and an additional 16 statements were identified from a general knowledge norming study (Tauber et al., 2013). When asked as trivia questions (e.g., "What was the name of Roy Roger's dog?" or "What is the name of the island on which Napoleon was born?"), our 56 items were correctly answered by 2.6% of participants on average (range: 0-11%) in the Tauber et al. (2013) norms. True statements stated the correct trivia information (e.g., "Bullet"), while false statements stated plausible, yet incorrect, alternatives to the correct information (e.g., "Sardinia" instead of "Corsica").

In addition to the 56 key trivia statements, we also used separate set of 40 true and false trivia statements about which participants likely *did* hold relevant prior knowledge. These statements were used to train participants about the trustworthiness of the sources, as we describe in the Procedure. These statements were taken from Fazio (2020), who identified "known" statements as those corresponding to questions that were answered correctly by 60% (range: 42-80%) of participants in the Tauber et al. (2013) norms.

Procedure. This experiment was administered online via the gorilla.sc platform (Anwyl-Irvine et al., 2020). Participants began the experiment by reading the study information sheet and completing two attention checks. Next, participants completed five Ishihara plate test trials to check for possible colorblindness. On each trial, participants were shown an image with a circle of dots with a number in a different color depicted inside it and were asked to type the number present in the image. There were four phases to the experiment: training, exposure, test, and manipulation check.

Training Phase. Next, participants began the source training phase, designed to allow participants to learn the trustworthiness of the two sources used in the study. Participants were told that we wanted their opinion on whether statements were true or false, and that they would receive feedback about whether their responses were correct or incorrect. In addition, participants were instructed that statements would come from one of two sources, Mr. Red and Mr. Green, and that these sources would be indicated by the background color behind the statement.

Then, participants saw 24 likely-known statements, one at a time, and were asked "Is the statement above true or false?" (options: True, False). Statements were centered inside a white rectangle which was centered inside a larger, colored rectangle, indicating the source of the statement (Figure 2.1). After selecting a response, participants received feedback about whether their response was accurate or inaccurate in a white box placed just below the statement. Feedback began with "Correct" or "Incorrect" followed by "This statement is true" or "This statement is false." This feedback remained on screen for 3 seconds before automatically proceeding to the next trial.

Figure 2.1

Sample Statements

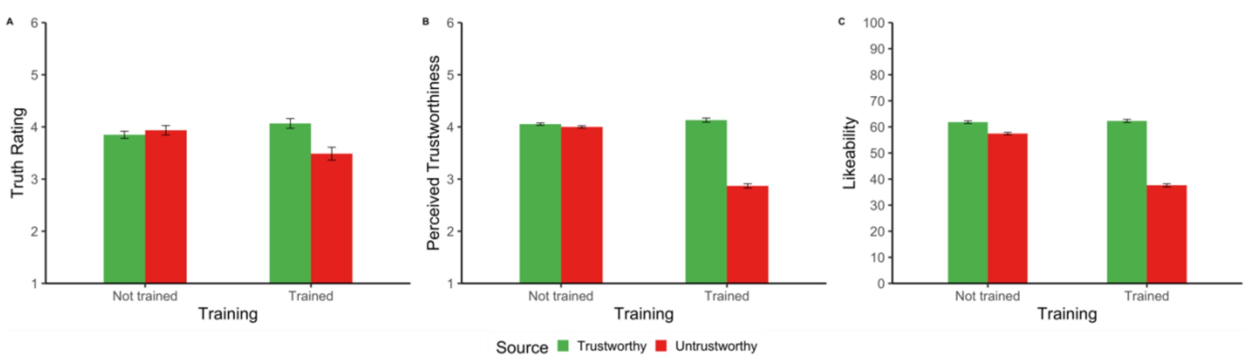
Note. Images correspond to sample statements attributed to Mr. Green (top left) or Mr. Red (top right) and statements not attributed to any source (bottom).

Critically, during this training phase, statements were probabilistically associated with each source. Specifically, Mr. Green presented nine true and three false statements, while Mr. Red presented nine false and three true statements. In order to counterbalance assignment of items to the two sources, we created four lists. In the first list, 15 true items were assigned to Mr. Green and the remaining five true items to Mr. Red. Similarly, 15 false items were assigned to Mr. Red and the remaining five to Mr. Green. To create the remaining three lists, we rotated which five true statements were assigned to Mr. Red and which five false statements were assigned to Mr. Green. Then, as described above, 24 of the full set of 40 statements were randomly selected for each participant, constrained to ensure the following breakdown: nine true from Mr. Green, three false from Mr. Green, nine false from Mr. Red, three true from Mr. Red.

We confirmed the effectiveness of this training phase in a pilot study ($N = 60$; see Appendix A for full details) in which participants were randomly assigned to complete or not complete the training phase. All participants were then asked to rate the truth of a series of new, unknown statements from either Mr. Green or Mr. Red, and were asked a series of manipulation check questions (see Manipulation Check below). Participants who completed the training phase rated Mr. Red as less trustworthy ($M_{Red} = 2.87$, $M_{Green} = 4.13$), $t(22) = -5.14$, $p < .001$, $d = 1.07$, and likeable ($M_{Red} = 37.6$, $M_{Green} = 62.3$), $t(22) = -4.33$, $p < .001$, $d = 0.90$ than Mr. Green. In addition, participants who received training were more likely to believe new statements coming from Mr. Green ($M = 4.07$) than Mr. Red ($M = 3.49$), $t(22) = 3.24$, $p = .004$, $d = 0.68$. Critically, there was a significant interaction between training and source for all three measures (Fs > 9.94, ps < .003) and there were no significant differences between ratings for Mr. Green and Mr. Red among participants who did not receive training (ts < 1.20, ps > .238), Figure 2.2.

Figure 2.2

Pilot Study Results



Note. Figure shows results from the following measures: A) truth ratings (1 = definitely false; 6 = definitely true) for new statements attributed to Mr. Green or Mr. Red B) perceived trustworthiness (1 = very untrustworthy, 6 = very trustworthy) for Mr. Green and Mr. Red and C) likeability ratings (0 = don't like, 100 = like very much) for Mr. Green and Mr. Red. In all conditions in all panels, bars on the left indicate responses for the trustworthy source, Mr. Green, and bars on the right represent responses for the untrustworthy source, Mr. Red.

Exposure Phase. After the training phase, participants were asked to "help us by providing ratings of statements that we plan to use in another research project." Participants were instructed that they would see 28 statements and rate each one on a 6-point scale from Very Interesting to Very Uninteresting. In addition, participants were told that all statements would come from a source other than Mr. Red or Mr. Green and thus all statements would be on a white background. Participants were not told whether statements would be true or false.

Then, participants saw the 28 statements that were assigned to be "repeated" for that participant. For each statement, participants responded to the prompt "How interesting is the statement above?" (*Very Uninteresting*, *Uninteresting*, *Slightly Uninteresting*, *Slightly Interesting*, *Interesting*, or *Very Interesting*). Statements were formatted like those shown in the training phase (statements centered in a white rectangle inside of a larger rectangle), except that the larger background rectangle was white, rather than red or green.

Test Phase. After the exposure phase, participants were told that they would now "continue with the main study." Participants were instructed to rate a series of 56 statements on a 6-point scale from Definitely True to Definitely False, without any feedback, and were told that some statements would be repeated, while others would be new. Participants were also told that statements would "again come from Mr. Red or Mr. Green," and were reminded that the sources would be indicated by background color. Unlike some prior studies of the illusory truth effect (e.g., Fazio, 2020), participants were not told whether statements would be true or false.

Then, participants saw the full set of 56 key statements, one at a time, and were asked "How true or false is the statement above?" (*Definitely True, Probably True, Possibly True,* *Possibly False, Probably False, Definitely False,* coded from 1 = Definitely False to 6 = Definitely True in the analyses below). The 56 statements were evenly split across levels of three key factors in this experiment's design: repetition (repeated, new), source trustworthiness (trustworthy, untrustworthy), and statement truth (true, false).

Manipulation Check. Then, participants were asked some follow-up questions. First, we asked participants "In the first set of statements you rated, which of the following was most accurate regarding the background colors?" (*Statements from Mr. Red were much more likely to be true, Statements from Mr. Red were somewhat more likely to be true, Statements from Mr. Red or Mr. Green were equally likely to be true, Statements from Mr. Green were somewhat more likely to be true, Statements from Mr. Green were much more likely to be true,* coded from 1 to 5). Next, we asked participants about the trustworthiness of each source across two questions ("How trustworthy is [Mr. Red/Mr. Green]?": (*Very Untrustworthy, Untrustworthy, Slightly Untrustworthy, Slightly Trustworthy, Trustworthy, Very Trustworthy*, coded from 1 to 6). Finally, we asked participants about the likeability of each source ("How much do think you would like [Mr. Red/Mr. Green]?": numeric slider entry from *0 = don't like,* to *100 = like very much*, with default of 50). After completing the study, participants answered optional demographic questions (gender, race, ethnicity, education), were thanked for their time, and were informed about the purpose of the study.

Results

For all experiments, all statistical tests are conducted at the .05 alpha level and are pre-registered unless labelled as exploratory.

Manipulation Checks. As a check of our source trustworthiness manipulation, we conducted exploratory analyses of the three sets of follow-up questions. Responses to all three sets of questions indicated that participants regarded the “trustworthy” source more favorably and as more reliable. First, on a 5-point scale from “Statements from Mr. Red were much more likely to be true” (1) to “Statements from Mr. Green were much more likely to be true,” (5) participant gave ratings higher than the midpoint, indicating that participants thought Mr. Green was more likely to provide true statements, $M = 3.41$, $SD = 0.92$, $t(191) = 6.11$, $p < .001$, $d = .44$. Next, participants gave higher trustworthiness ratings to Mr. Green ($M = 4.23$) than they did to Mr. Red ($M = 3.63$), $t(191) = 7.33$, $p < .001$, $d = 0.53$, 95% CI [0.45, 0.77]. Finally, participants also reported liking Mr. Green ($M = 62.70$) more than Mr. Red ($M = 53.69$), $t(191) = 6.26$, $p < .001$, $d = 0.45$, 95% CI [6.18, 11.89]

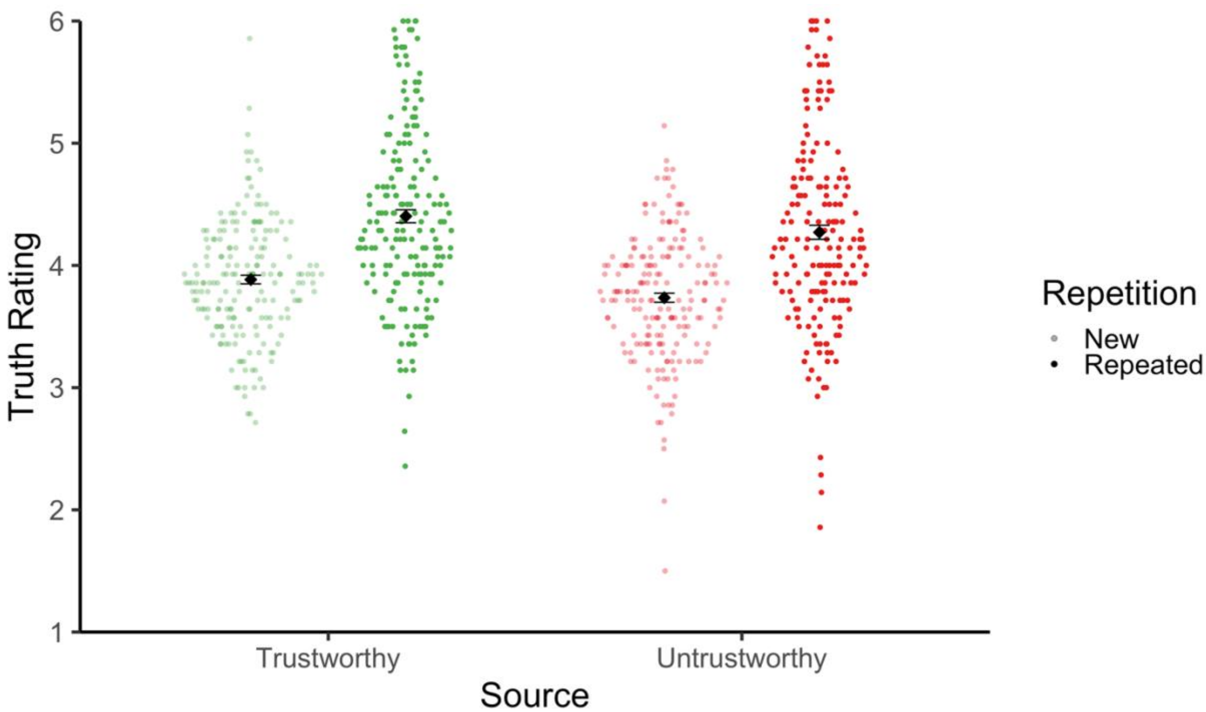
Truth Ratings. Our main hypothesis was that repetition would increase belief in claims to a greater degree when the claim came from a trustworthy source than when it came from an untrustworthy source. However, as shown in Figure 2.3 below, repetition increased belief similarly when claims came from trustworthy and untrustworthy sources.

We evaluated these data statistically using a 2 (repetition: repeated, new) × 2 (source trustworthiness: trustworthy, untrustworthy) ANOVA on participants’ mean truth ratings. Replicating the illusory truth effect, we observed a main effect of repetition such that participants gave higher truth ratings to repeated ($M = 4.34$) versus new statements ($M = 3.81$), $F(1, 191) = 95.00$, $p < .001$, $\eta_p^2 = 0.33$. In addition, consistent with our source trustworthiness manipulation, participants gave higher truth ratings to statements from the trustworthy source, Mr. Green ($M = 4.14$) than the untrustworthy source, Mr. Red ($M = 4.00$), $F(1, 191) = 13.48$, $p < .001$, $\eta_p^2 = 0.07$. However, contrary to our hypothesis, we did not observe an interaction effect between source trustworthiness and repetition, $F(1, 191) = 0.14$, $p = .71$, $\eta_p^2 < 0.01$. Repetition increased belief for statements coming from both trustworthy sources ($M_{new} = 3.88$, $M_{repeated} = 4.40$; $t(191) = 8.93$, p

< .001, $d = 0.63$, 95% CI [0.40, 0.63]), and untrustworthy sources ($M_{new} = 3.73$, $M_{repeated} = 4.27$; $t(191) = 9.13$, $p < .001$, $d = 0.66$, 95% CI [0.42, 0.65]).

Figure 2.3

Mean Truth Ratings by Repetition and Source Trustworthiness at Test (Experiment 1)



Note. Ratings were coded from 1 = *Definitely False* to 6 = *Definitely True*. Each dot represents one participant ($N = 192$) with values horizontally shifted to represent the density distribution. Black diamonds reflect group means and error bars reflect standard errors of the mean.

Discussion

Experiment 1 demonstrates that repetition increases belief to a similar degree regardless of whether statements are coming from a source thought to be trustworthy or untrustworthy. This finding is comparable to the results Unkelbach & Greifeneder (2018), who find that repetition and advice about the truth of a statement from an independent source additively—not interactively—

affect perceived truth. Further, this pattern is consistent with current cognitive theories of the illusory truth effect that emphasize that the effect is driven by low-level cognitive cues like fluency that may be difficult to discount, even when hearing a statement being repeated by an untrustworthy statement.

Critically, in Experiment 1, sources for statements were indicated at test—after fluency had already accrued for the statements through prior exposure. One possibility is that, at the time of rating truth, the trustworthiness of a statement's source does not necessarily indicate anything about the reliability of fluency from prior exposure as a cue for truth. Thus, in Experiment 2 onwards, we examine if *initially* exposing participants to statements from untrustworthy or trustworthy sources will have different effects on perceived truth when the statement is later repeated. We hypothesize that people will be able to discount fluency as a cue for truth when that fluency comes from earlier exposure to a statement from an untrustworthy source. In addition, we measure and examine participants' memory for the source of information, as recollection of the source is a key process likely to be involved in these effects (Begg et al., 1992)

Experiment 2

Method

Participants.

Statistical Power. We pre-registered a sample size of 281 based on an *a priori* power analysis conducted using G*Power (Faul et al., 2009) for a two-tailed matched-pairs *t*-test. A sample size of 281 participants provides 80% power to detect an effect of size $d_z = 0.168$.

This minimal effect size of interest was determined by selecting 0.1 points as the smallest difference in truth ratings between statements repeated from the trustworthy versus untrustworthy source that we were practically interested in detecting for this experiment (on a 6-point scale). We converted this raw difference into the standardized effect size d_z by assuming a standard deviation of 0.77 and a correlation among repeated measures of 0.70. Both values were based on what we observed for participants' mean ratings of repeated items from the trustworthy and untrustworthy sources in Experiment 1 (and were similar to the observed values in this experiment of 0.71 and 0.79, respectively). For context, a past meta-analysis estimates that the effect of repetition on perceived truth is d = 0.49 (Dechêne et al., 2010). Here, we were powered to detect a difference in truth ratings between two types of repetition that is about three times smaller, d_z = 0.168.

Recruitment and Exclusions. Two hundred and eighty-one adult participants were recruited from MTurk in a similar manner as described in Experiment 1 (participants who completed Experiment 1 were ineligible for this study). Again, while we had pre-registered that we would exclude any participants that a) failed two attention checks at the beginning of our survey (typing the name of a black and white cartoon animal and selecting a requested responses on a 5-point Likert question) or b) failed more than two of five Ishihara colorblindness test trials (Ishihara, 1917), no participants met either of these criteria for exclusion.

Demographics. The mean age of participants was 41.50 (*SD* = 12.63; Range = 19-79). Our final sample was predominantly White (78%, 8.2% Asian, 7.5% Black, 5.0% Multiracial, 0.36% Middle Eastern and North African, 0.36% Pacific Islander, 0.36% reporting some other race, 0.71% not reporting) and non-Hispanic (89%, 6.1% Hispanic, 4.6% not reporting), and 48% of participants were women (48% men, 0.71% nonbinary, 3.2% not reporting). Most participants (73%) had received at least a college degree.

Design. This experiment manipulated repetition (new, trustworthy repetition, untrustworthy repetition) and statement truth (true, false) within-subjects. Repetition was counterbalanced by assigning participants to one of three stimulus sets. In the first set, we assigned one third of the true items as new, one third as trustworthy repetition, and one third as untrustworthy repetition and split the false items in the same way. To create the remaining two sets, we rotated the levels of repetition through each third of the true and false items.

Materials. The key stimuli were 42 likely unknown trivia statements were a randomly chosen subset of the key stimuli used in Experiment 1. In addition, the 42 key statements, we used 24 likely unknown statements during the training phase—these were a randomly chosen subset of the statements used for this purpose in Experiment 1.

Procedure. As in Experiment 1, this experiment was administered online via gorilla.sc (Anwyl-Irvine et al., 2020a) and participants began the experiment by reading the study information sheet, completing two attention checks, and responding to five Ishihara plate test trials to check for possible colorblindness. There were five phases to this experiment: training, exposure, test, memory check, and manipulation check.

Training Phase. The training phase was identical to that of Experiment 1, except that all participants saw the same set of 24 statements (as opposed to a subset of 24 statements from a full set of 40).

Exposure Phase. The exposure phase was similar to that of Experiment 1, except that the source of the statement was indicated on each trial. In the instructions for this phase, participants were reminded that the source of statements would be indicated by the background color of the statement. Participants were not told whether statements would be true or false. Participants then saw 28 of the 42 key statements (half true, half false; half from Mr. Red, half from Mr. Green) and, for each statement, were asked “How interesting is the statement above?” (*Very Uninteresting*, *Uninteresting*, *Slightly Uninteresting*, *Slightly Interesting*, *Interesting*, or *Very Interesting*).

Test Phase. The test phase was similar to that of Experiment 1, except that sources for statements were not indicated. Participants were correctly instructed that some statements would be repeated, and some would be new, but were not told anything about the truth of the statements. Then, participants saw all 42 key statements (one third each new, repeated from Mr. Green during the exposure phase, and repeated from Mr. Red during the exposure phase). Each statement was presented individually on a white background not indicating any source (see Figure 2.1). For each statement, participants were asked "How true or false is the statement above?" (*Definitely True, Probably True, Possibly True, Possibly False, Probably False, Definitely False,* coded from 1 = Definitely False to 6 = Definitely True in the analyses below).

Memory Check. Next, participants were told that we would ask them some questions about the second phase of the experiment in which they rated their interest in statements (i.e., the exposure phase). Participants were told that they would see 12 statements and that they should indicate "whether the statement came from Mr. Red, Mr. Green, or it was not presented in the second part of the experiment." Then, participants saw 12 statements in plain text in the center of the screen without any surrounding rectangle above the question "Who did this statement come from?" (*Mr. Red, Mr. Green, Not Presented*). The 12 statements were randomly selected for each participant such that 4 were actually presented by Mr. Red during the exposure phase, four were presented by Mr. Green and four were not presented.

Manipulation check. Participants then completed the same manipulation questions as in Experiment 1. After completing the study, participants answered optional demographic questions (gender, race, ethnicity, education), were thanked for their time, and were informed about the purpose of the study.

Results

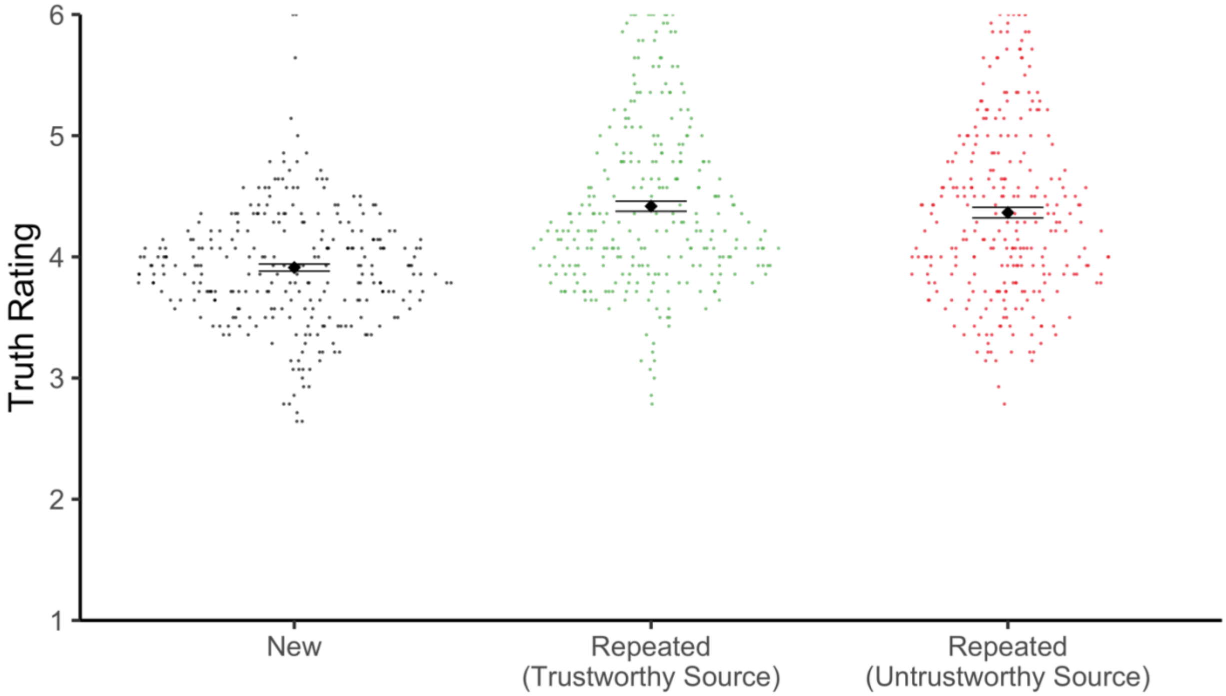
Manipulation Checks. As in Experiment 1, we first report exploratory analyses of the three sets of follow-up questions to examine our source trustworthiness manipulation. Participants thought Mr. Green was more likely to provide true statements than Mr. Red, $M = 3.48$, $SD = 0.98$, $t(280) = 8.29$, $p < .001$, $d = .49$; and also found Mr. Green to be more trustworthy ($M = 4.35$) and likeable ($M = 67.67$) than Mr. Red ($M_{trust} = 3.52$; $M_{liking} = 53.25$), $t(280) = 10.49$, $p < .001$, $d = 0.62$, 95% CI [0.67, 0.98], $t(280) = 9.07$, $p < .001$, $d = 0.54$, 95% CI [11.29, 17.55], respectively.

Truth Ratings. Turning to our pre-registered analyses, we hypothesized that, in line with the illusory truth effect, statements repeated from the trustworthy source would be perceived as more true than new statements. Critically, we also hypothesized that statements repeated from the trustworthy source would be perceived as more true than statements repeated from the untrustworthy source. We did not make a prediction about whether repetition would increase truth ratings for statements repeated by the untrustworthy source.

As shown in Figure 2.4, repetition increased truth ratings regardless of whether the information was originally presented by a trustworthy or untrustworthy source. A one-way repeated measures ANOVA on participants' mean truth ratings revealed a significant main effect of repetition status (new, repeated from a trustworthy source, repeated from an untrustworthy source), $F(2, 560) = 99.80$, $p < .001$, $\eta_p^2 = 0.26$. As hypothesized, participants rated statements repeated from a trustworthy source ($M = 4.42$) as more true than new statements ($M = 3.91$), $t(280) = 11.60$, $p < .001$, $d = 0.69$, 95% CI [0.42, 0.59]. In addition, statements repeated from an untrustworthy source ($M = 4.37$) were rated as more true than new statements, $t(280) = 10.17$, $p < .001$, $d = 0.61$, 95% CI [0.37, 0.54]. The difference in ratings between statements repeated by trustworthy versus untrustworthy sources was marginally significant in the direction predicted by our hypothesis (larger truth ratings for statements repeated by trustworthy sources), $t(280) = 1.92$, $p = .056$, $d = 0.11$, 95% CI [-0.001, 0.11].

Figure 2.4

Mean Truth Ratings by Repetition Status (Experiment 2)



Note. Ratings were coded from 1 = *Definitely False* to 6 = *Definitely True*. Each dot represents one participant (*N* = 281) with values horizontally shifted to represent the density distribution. Black diamonds reflect group means and error bars reflect standard errors of the mean.

Source Memory. Next, we report descriptive statistics regarding participants' responses to the memory questions. Recall that participants were presented with 12 statements (four each per repetition condition) and asked to indicate whether the statement was not presented, presented by Mr. Green, or presented by Mr. Red in the exposure phase. Current theories of memory suggest this sort of task involves two abilities, which we examine separately: *item memory*, or the ability to distinguish between items that had versus had not been presented, and *source memory*, or the ability to determine which of the two sources presented previously-seen items (see Ghetti & Angelini, 2008 for a similar breakdown). To examine item memory, we collapsed responses into two categories: new (not presented) and old (presented by Mr. Green or presented by Mr. Red). Then, we calculated two measures from signal detection theory: *d'*, a measure of participants ability to discriminate between new and old items, and *c,* a measure of participants general bias to say "old" versus "new." Then, to examine source memory, we calculated the number of "old" items for which participants correctly identified the correct source divided by the number of "old" items that participants correctly identified as old (whether or not they selected the correct source). For instance, a participant who gave a response of either "presented by Mr. Green" or "presented by Mr. Red" to 6 of the 8 items that were actually presented by either of these sources, but only correctly identified the source for 3 of these items, they would receive score of 3/6 = 50%. Participants who classified all old items as "not presented" (N = 7) were excluded from this calculation, as they would have a denominator of 0.

Overall, participants were somewhat sensitive to the difference between new versus old items, indicated by a positive *d'* (M = 0.76, SD = 0.87), though participants had an overall bias to characterize items as "old," indicated by a negative *c* (M = -0.85, SD = 0.69). By contrast, source memory was rather poor, with participants only correctly identifying the source for 52.25% (SD = 20.62%) of items correctly recognized as having been presented in the exposure phase, just above chance level.

Individual Differences in Truth Ratings. While we did not observe a significant main effect of source on belief in the repeated statements, we do note that there was variability in this effect across participants, with 51.6% of participants providing greater truth ratings to repeated statements from the trustworthy versus untrustworthy source. Thus, as an exploratory analysis we examined whether we could predict individual-level variability in the effect of source (trustworthy versus untrustworthy) on people's beliefs. We predicted participant-level difference scores in truth ratings for statements repeated by the trustworthy minus untrustworthy source using linear regression. This regression included two predictors: a) difference scores in *perceived trust* for the trustworthy minus untrustworthy source and b) *source memory ability*, as indexed by the proportion of old items whose source was correctly classified. Full results are available at Appendix A, but to preview, we find that only source memory predicted the effect of sources on truth ratings ($b = 0.30$, $t(271) = 0.13$, $p = .028$). Participants who were better able to remember the source of statements tended to give higher truth ratings to statements from the trustworthy relative to the untrustworthy source.

Discussion

In Experiment 2, we find that repetition increased belief, even when the statement originally came from a seemingly untrustworthy source. However, we also obtained marginally significant evidence that these statements were less likely to be believed than those repeated by the trustworthy source, in line with the possibility that social information moderates the effects of repetition on belief. To follow up, we conducted a second experiment which was identical to Experiment 2 except for three changes. First, we increased our sample size, providing us with 80% power to detect the difference between truth ratings for the trustworthy and untrustworthy sources observed in Experiment 2 in a pre-registered one-tailed test. Second, we increased the number of observations per cell of our design from 14 to 20, to further increase the precision of our estimates, and thus power. Third, we heightened the proportion of true/false statements that Mr. Green/Mr. Red said in the training phase from nine of 12 (75%) to 10 of 12 (83.33%), to make the two sources somewhat more clearly trustworthy or untrustworthy.

Experiment 3

Method

Experiments 3-5 received ethics approval under protocol 2023E0154 from the Institutional Review Board of The Ohio State University.

Participants.

Statistical Power. We pre-registered a sample size of 500 based on an *a priori* power analysis for a matched-pairs *t*-test in G*Power. Given our strong directional hypothesis that statements repeated by the trustworthy would be rated as more true than those from the untrustworthy source, we powered for and pre-reregistered a one-tailed test. This power analysis revealed that 478 participants would be needed to provide 80% power to detect an effect of size $d_z = 0.114$, the effect observed in Experiment 2. Rounding up, we pre-reregistered a sample size of 500.

Recruitment and Exclusions. Five hundred participants were recruited via the Connect platform (Hartman et al., 2023).[2] As in Experiment 2, we restricted participation to individuals based in the U.S. and advertised the study as only eligible to people with full color vision.

We again pre-registered that we would exclude participants that a) failed both attention checks or b) failed a colorblindness test, but no participants were excluded for either of these reasons. Upon inspection of the data, we noticed technical errors in the study for two participants (i.e., one participant failed to see a training phase trial, and another participant saw an exposure phase trial twice). Prior to data analysis, we decided to exclude these participants as a conservative (though not pre-registered) measure, leaving us with 498 participants in our analyses below.

---

[2] While we pre-registered that we would recruit participants from MTurk, we used Connect instead due to logistical issues with uploading funds to MTurk.

Demographics. The mean age of participants was 38.84 (*SD* = 11.87; Range = 18-79). Our final sample was predominantly White (76 %, 9.8% Black, 8.0% Asian, 4.2% Multiracial, 0.20% Middle Eastern and North African, 0.20% Pacific Islander, 1.4% reporting some other race, 0.40% not reporting) and non-Hispanic (88 %, 8.6% Hispanic, 3.0% not reporting), and 53% of participants were men (45% women, 0.71% nonbinary, 3.2% not reporting). Most participants (70%) had received at least a college degree.

Design. This experiment used the same design as Experiment 2

Materials. The key stimuli were 60 trivia statements (half true, half false) about which participants likely did not directly hold prior knowledge. 56 statements were used from Experiment 1, and an additional 4 were adapted from a general knowledge norming study (Tauber et al., 2013), as described in Experiment 1. In addition to these 60 key trivia statements, we again used the 24 well-known trivia statements from Experiment 2 in the training phase.

Procedure. The procedure for this experiment was identical to that of Experiment 2, except that participants read and rated more key statements. During the exposure phase, participants saw 40 statements, and during the test phase, participants saw 60 statements.

Results

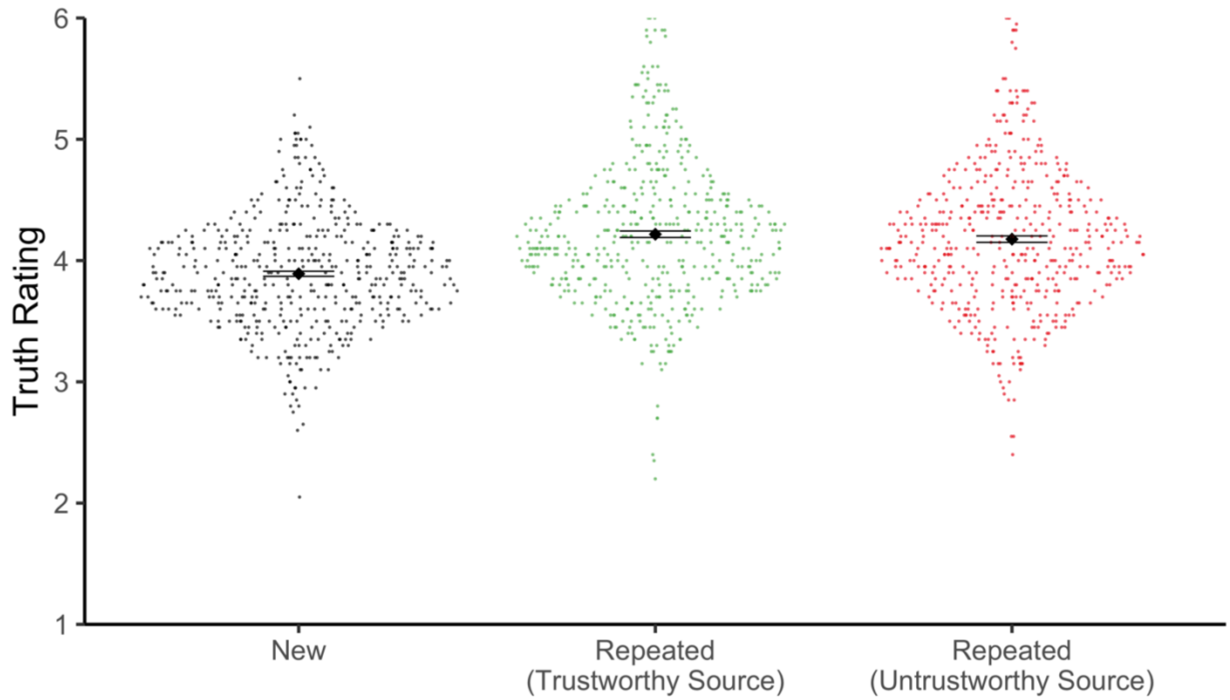
Manipulation Checks. As in Experiment 1, we first report exploratory analyses of the three sets of follow-up questions to examine our source trustworthiness manipulation. Participants thought Mr. Green was more likely to provide true statements than Mr. Red, $M = 3.78$, $SD = 1.12$, $t(497) = 15.01$, $p < .001$, $d = .70$; and also found Mr. Green to be more trustworthy ($M = 4.43$) and likeable ($M = 66.24$) than Mr. Red ($M_{trust} = 3.17$; $M_{liking} = 45.75$), $t(497) = 18.17$, $p < .001$, $d = 0.81$, 95% CI [1.12, 1.39], $t(497) = 16.45$, $p < .001$, $d = 0.73$, 95% CI [18.04, 22.94], respectively.

Truth Ratings. Our pre-registered hypotheses made three directional predictions. First, statements repeated from the trustworthy source should be perceived as more true than new statements. Second, statements repeated from the untrustworthy source should also be perceived as more true than new statements. Third, statements repeated from the trustworthy source should be perceived as more true than statements repeated by the untrustworthy source. As shown in Figure 2.5, our results were consistent with all of these predictions.

A one-way repeated measures ANOVA on participants' mean truth ratings revealed a significant main effect of repetition status (new, repeated from a trustworthy source, repeated from an untrustworthy source), $F(2, 994) = 118.13$, $p < .001$, $\eta_p^2 = 0.19$. As hypothesized, participants rated statements repeated from a trustworthy source ($M = 4.22$) as more true than new statements ($M = 3.89$), $t(497) = 13.02$, $p < .001$, $d = 0.58$, 95% CI [0.27, 0.38]. In addition, statements repeated from an untrustworthy source ($M = 4.18$) were rated as more true than new statements, $t(497) = 11.61$, $p < .001$, $d = 0.52$, 95% CI [0.24, 0.33]. Finally, statements repeated by the trustworthy source were perceived as more true than those repeated by the untrustworthy source, $t(497) = 2.09$, $p = .019$, $d = 0.09$, 95% CI [0.002, 0.08]. Note that, as pre-registered, we report one-tailed *t*-tests in line with our directional hypotheses, but for all analyses we report two-tailed confidence intervals for ease of interpretation.

Figure 2.5

Mean Truth Ratings by Repetition Status (Experiment 3)



Note. Ratings were coded from 1 = *Definitely False* to 6 = *Definitely True*. Each dot represents one participant (*N* = 498) with values horizontally shifted to represent the density distribution. Black diamonds reflect group means and error bars reflect standard errors of the mean.

Source Memory. Next, we report descriptive statistics regarding item and source memory, based on participants' responses to the memory questions. As in Experiment 2, participants were somewhat sensitive to the difference between new and old items (mean *d*' = 0.69, *SD* = 0.81), though they exhibited a general bias to categorize items as "old" (mean *c* = -0.91, *SD* = 0.61). However, again, participants recognized the source of old items at just above chance level (*M* = 54.23%, *SD* = 19.38%, 7 participants excluded for not correctly recognizing any old items). Overall, memory ability as indexed by these measures was generally comparable to that observed in Experiment 2.

Individual Differences in Truth Ratings. While there was a significant main effect of statement source (trustworthy versus untrustworthy) in this experiment, there was again variability in this effect across participants, with only 52.2% of participants giving higher ratings to the repeated statements from the trustworthy relative to the untrustworthy source. As an exploratory analysis, we examined whether these differences could be predicted by participants' relative trust in the two sources or their source memory ability. Full results are available in Appendix A, but we did not find evidence that trust ($b = 0.00$, $t(488) = 0.29$, $p = .770$) or source memory ($b = 0.16$, $t(488) = 1.59$, $p = .113$) predicted the effects of source on belief.

Discussion

Experiment 3 confirms the pattern observed in Experiment 2. Repetition increased belief in statements regardless of source. However, these effects were slightly smaller when statements originally came from the untrustworthy source, confirming that social evaluations can moderate the relationship between repetition and belief.

However, memory for the source of the statements was far from perfect—participants were barely above chance at correctly identifying the source of previously encountered statements. Thus, in Experiment 4, we examined whether repetition of statements from Mr. Red would still increase belief when participants were better able to remember the source of each statement. To increase attention and memory, we made two changes to the procedure of Experiment 3. First, we specifically instructed participants that their task is to "remember WHO said WHAT" during the exposure phase. Second, rather than having participants rate their interest in each statement, we asked participants to rate how likely they would be to remember who said that statement. These judgements of learning often improve associative learning between a cue (here, the statement) and a target (here, the source of the statement) (e.g., Myers et al., 2020; Rivers et al., 2021; Soderstrom et al., 2015).

Still, Experiments 2 & 3 provided inconsistent evidence as to whether source memory would dictate the effects of repetition of statements from these two sources. In exploratory analyses in Experiment 2, we found evidence that participants with better source memory show greater relative belief in statements from the trustworthy than the untrustworthy source, but we failed to observe this effect in Experiment 3. As described in the introduction, one alternate mechanism through which source trustworthiness may have an impact is by eliciting a skeptical, elaborative encoding process when participants are presented with statements from an untrustworthy source. If this is the case, we might expect that the effect of the initial source of a statement to matter more for participants higher in need for cognition, or the tendency to engage in thinking (Cacioppo & Petty, 1982), as these people may be more prone to deep, elaborative processing. Thus, we also introduced a 6-item measure of participants' need for cognition at the end of the experiment (NCS-6; Coelho et al., 2020) and explore whether this moderates our key effects.

Experiment 4

Method

Participants.

Statistical Power. We pre-registered a sample size of 500 based on the same rationale outlined in Experiment 3.

Recruitment and Exclusions. Four hundred and ninety-nine participants were recruited via the Connect platform in the same manner as described in Experiment 3 (participants who completed Experiment 3 were ineligible for this study). Note that this is one short of pre-registered sample size due to an incomplete submission.

No participants were excluded per our pre-registered exclusion criteria (failing attention checks or failing a colorblindness test). However, like in Experiment 3, we noticed that 3 participants faced technical errors during the study (i.e., two participants failed to see all training phase trials and one participant had several repeated trials during the test phase). Again, we conservatively decided to exclude these participants, leaving us with 496 participants in our analyses below.

Demographics. The mean age of participants was 38.63 (*SD* = 11.97; Range = 18-77, 1 not reporting). Our final sample was predominantly White (69%, 14% Black, 9.5% Asian, 4.8% Multiracial, 1.4% reporting some other race, 0.81% not reporting) and non-Hispanic (87%, 10% Hispanic, 2.8% not reporting), and 51% of participants were men (45% women, 1.0% nonbinary, 2.8% not reporting). Most participants (71%) had received at least a college degree.

Design & Materials. This experiment used the same design and stimuli as Experiment 3.

Procedure. The procedure for this experiment was identical to that of Experiment 3 with three modifications. First, participants were explicitly instructed to pay attention to the source of each statement before the exposure phase. Specifically, participants were told:

> "Now, you will be shown a second block of 40 statements. In this part, your goal is to learn WHO said WHAT. The statements will again come from Mr. Red or Mr. Green. Remember that these sources will be indicated by the background color behind the statement. Mr. Red's statements will have a red background and Mr. Green's a green background.
>
> Please pay close attention to the source of each statement. At the end of this experiment, we will show you some of these statements, and ask you for your memory of whether they were said by Mr. Red or Mr. Green. We hope that you will be able to identify WHO said each statement."

Second, rather than rating their interest in each statement, participants provided a judgement of how likely they were to remember the source of the statement during the exposure phase. On each trial, participants first saw the statement for 2 seconds, then were asked "How likely are you to remember who said this statement?" with a 100-point slider scale (0 = not at all likely, 100 = very likely, default of 50). Finally, we added an exploratory measure of need for cognition (NCS-6; Coelho et al., 2020) consisting of 6 items rated on 5-point scales (e.g., "I would prefer complex to simple items"), at the end of the study, just before participants were debriefed.

Results

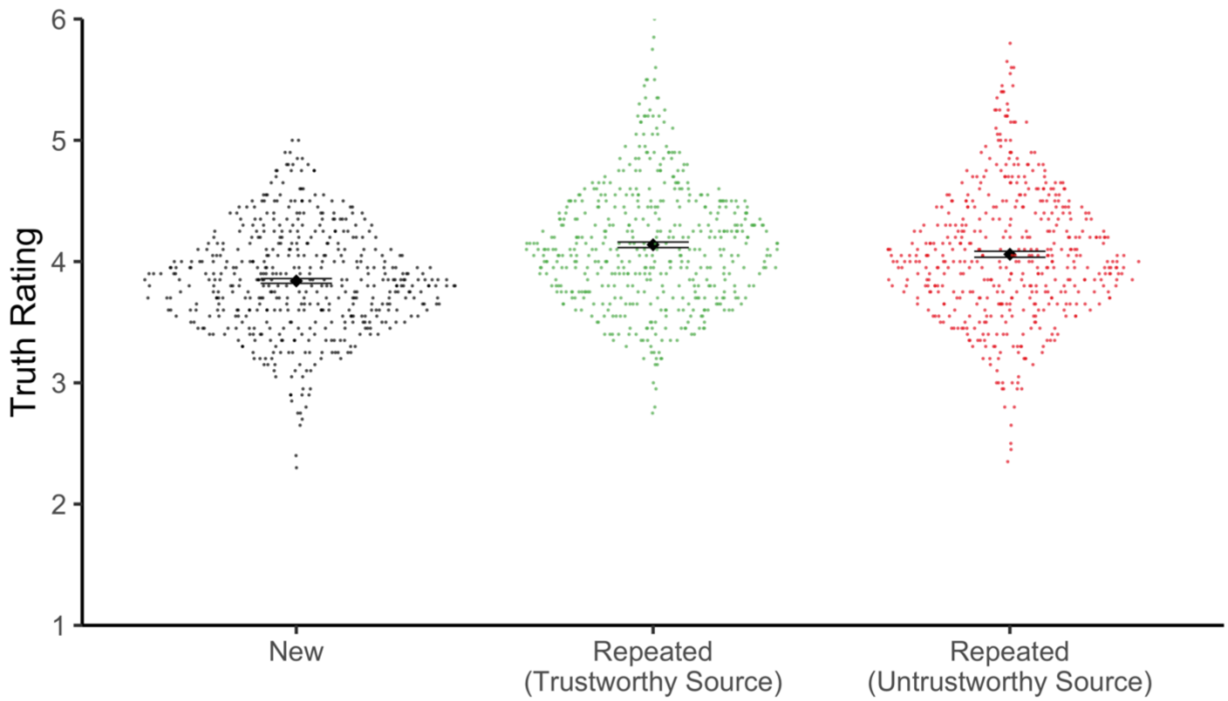
Manipulation Checks. As in Experiments 2 & 3, exploratory analyses of the three sets of follow-up questions revealed that the source manipulation was effective. Participants thought Mr. Green was more likely to provide true statements than Mr. Red, $M = 3.81$, $SD = 1.06$, $t(495) = 17.05$, $p < .001$, $d = .77$; and also found Mr. Green to be more trustworthy ($M = 4.41$) and likeable ($M = 67.09$) than Mr. Red ($M_{trust} = 3.15$; $M_{liking} = 43.56$), $t(495) = 18.44$, $p < .001$, $d = 0.83$, 95% CI [1.13, 1.40], and $t(495) = 15.25$, $p < .001$, $d = 0.69$, 95% CI [17.19, 22.28], respectively.

Truth Ratings. As shown in Figure 2.6, the results were similar to those from Experiment 2. As predicted, statements repeated from the trustworthy source were perceived as more true than new statements and statements repeated from the trustworthy source were perceived as more true than statements repeated by the untrustworthy source. Turning to our key question, statements repeated from the untrustworthy source were still perceived as more true than new statements. Repetition still increased belief, even for statements originally presented by an untrustworthy source.

These observations were confirmed with our preregistered analyses. A one-way repeated measures ANOVA on participants' mean truth ratings revealed a significant main effect of repetition status (new, repeated from a trustworthy source, repeated from an untrustworthy source), $F(2, 990) = 85.06$, $p < .001$, $\eta_p^2 = 0.15$. As hypothesized, participants rated statements repeated from a trustworthy source ($M = 4.14$) as more true than new statements ($M = 3.84$), $t(495) = 12.52$, $p < .001$, $d = 0.56$, 95% CI [0.25, 0.35], and as more true than statements repeated by the untrustworthy source ($M = 4.18$), $t(495) = 3.53$, $p < .001$, $d = 0.16$, 95% CI [0.03, 0.12]. Note that, as pre-registered, these *t*-tests are one-tailed in line with our directional hypotheses. Turning to our final research question, we also find that statements repeated by the untrustworthy source are still rated as more than new statements, $t(495) = 8.78$, $p < .001$, $d = 0.39$, 95% CI [0.17, 0.27].

Figure 2.6

Mean Truth Ratings by Repetition Status (Experiment 4)



Note. Ratings were coded from 1 = *Definitely False* to 6 = *Definitely True*. Each dot represents one participant (N = 496) with values horizontally shifted to represent the density distribution. Black diamonds reflect group means and error bars reflect standard errors of the mean.

Source Memory. Next, we report descriptive statistics regarding item and source memory, based on participants' responses to the memory questions. As in Experiments 2 & 3, participants were sensitive to the difference between new and old items (mean d' = 0.90, SD = 0.89), though they exhibited a general bias to categorize items as "old" (mean c = -0.82, SD = 0.59). Participants were also above chance at correctly recognizing the source of old items (M = 60.56%, SD =

21.80%; 0 exclusions). In exploratory analyses, we compared these statistics with those of Experiments 2 & 3. Full results are available in Appendix A, but to summarize, *d'* was higher in this experiment ($M_{\text{Exp 2}} = 0.76$, $M_{\text{Exp 3}} = 0.69$; *p*s < .035), as was source memory ($M_{\text{Exp 2}} = 52.25\%$, $M_{\text{Exp 3}} = 54.23\%$; *p*s < .001). Thus, participants were better able to discriminate old from new items and to remember the source of items, consistent with our manipulations designed to improve memory in this experiment.

Individual Differences in Truth Ratings. As in prior experiments, the effect of statement source (trustworthy versus untrustworthy) varied notably across participants, with 54.0% of participants giving higher truth ratings to statements from the trustworthy than the untrustworthy source. Thus, we again conducted an exploratory linear regression to predict variation in this effect across participants. In addition to the two predictors used in prior experiments (relative trust in the trustworthy versus untrustworthy source and source memory), we also added participants' average scores on the measure of need for cognition. Responses to each of the 6 items were scored from 1-5 and summed to produce the final scores ($M = 20.65$, $SD = 5.66$).

Full results of the regression are available in Appendix A, but to preview, we find that the effects of statement source are greater for participants who are more trusting of Mr. Green over Mr. Red ($b = 0.05$, $t(492) = 3.62$, $p < .001$) and who had better memory for the sources ($b = 0.20$, $t(492) = 1.97$, $p = .049$). We did not find an effect of participants' scores on the measure need for cognition ($b = 0.00$, $t(492) = 0.91$, $p = .361$).

Discussion

Once again, repetition increased belief in statements coming from an untrustworthy source, though not to the same degree as statements coming from an untrustworthy source. Interestingly, the difference in belief in statements coming from the trustworthy versus untrustworthy source was larger in this experiment (d = 0.16) than in Experiment 3 (d = 0.09). These results are consistent with the possibility that the effects of source trustworthiness depend on memory-based mechanisms. In this experiment, participants received instructions and completed a task (judgements of learning) designed to enhance memory for the source of the statement. In fact, source memory was higher in this experiment than in prior experiments. Further, exploratory analyses revealed that participants who had better memory in this experiment also showed greater differences in truth ratings for statements by the trustworthy vs untrustworthy sources.

By contrast, our results were less consistent with a thoughtful encoding mechanism. Variation in participants' need for cognition failed to explain differences in the effect of initial source on belief, suggesting that these latter effects may not depend on the degree to which participants initially thoughtfully elaborated on claims from the untrustworthy source. That being said, these analyses were exploratory, and the difference in belief across sources was small to begin with, preventing strong conclusions.

Experiment 5

Experiments 2-4 demonstrate that social evaluations about the initial source of a claim can moderate the extent to which people rely on cues like fluency from prior exposure when evaluating truth. However, in all experiments, repetition still increased belief, even when the statement was repeated by a source explicitly deemed by participants as untrustworthy. One possible explanation for this is that this source is not *fully* considered untrustworthy. For instance, during the initial training phase, Mr. Red said *some* true statements (i.e., two out of 12, or 17% of the time), leaving participants with the sense that this source may not be *completely* unreliable. To examine this possibility, we conducted a final experiment where during the source training phase, Mr. Green *always* said true statements and Mr. Red *always* said false statements. Our key question was whether, with this more extreme source information, participants would continue to be influenced by prior exposure to statements from the untrustworthy source.

Method

Participants.

Statistical Power. We pre-registered a sample size of 500 based on the same rationale outlined in Experiment 3.

Recruitment and Exclusions. Four hundred and ninety-six participants were recruited via the Connect platform in the same manner as described in Experiment 3 (participants who completed Experiments 3 or 4 were ineligible for this study). Note that this is four short of pre-registered sample size due to four incomplete submissions.

No participants were excluded per our pre-registered exclusion criteria (failing attention checks or failing a colorblindness test).

Demographics. The mean age of participants was 38.00 (*SD* = 13.02; Range = 18-85, 1 not reporting). Our final sample was predominantly White (71%, 11% Black, 7.7% Multiracial, 7.1% Asian, 0.40% Pacific Islander, 1.0% reporting some other race, 1.6 % not reporting) and non-Hispanic (86%, 10% Hispanic, 4.05% not reporting), and 50% of participants were men (47% women, 1.6% nonbinary, 1.68% not reporting). Most participants (67%) had received at least a college degree.

Design & Materials. This experiment used the same design as Experiments 2-4 and the same materials as Experiments 3-4.

Procedure. The procedure for this experiment was identical to that of Experiment 4 with one exception. During the training phase, information from the trustworthy source (Mr. Green) was always true, and information from the untrustworthy source (Mr. Red) was always false.

Results

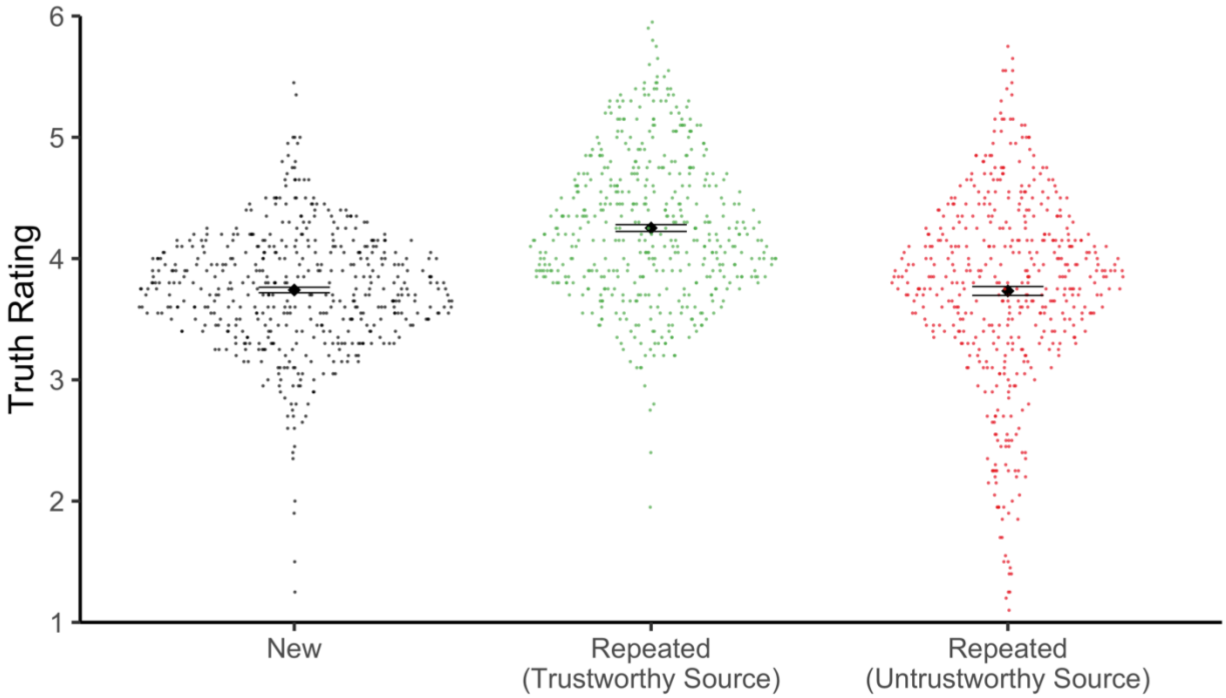
Manipulation Checks. As in experiments 2-4, exploratory analyses of the follow-up questions revealed the source manipulation to have been successful. Participants thought Mr. Green was more likely to provide true statements than Mr. Red, $M = 4.32$, $SD = 1.03$, $t(493) = 28.50$, $p < .001$, $d = 1.28$; and also found Mr. Green to be more trustworthy ($M = 4.85$) and likeable ($M = 71.56$) than Mr. Red ($M_{trust} = 2.42$; $M_{liking} = 36.38$), $t(493) = 26.62$, $p < .001$, $d = 1.20$, 95% CI [2.25, 2.60], and $t(493) = 22.06$, $p < .001$, $d = 0.99$, 95% CI [32.05, 38.32], respectively.

Truth Ratings. As shown in Figure 2.7, the pattern of results was similar to the previous findings, but strikingly different in one respect. As predicted, statements repeated from the trustworthy source were perceived as more true than new statements and statements repeated from the trustworthy source were perceived as more true that statements repeated by the untrustworthy source. However, statements repeated from an untrustworthy source were perceived similarly to new statements.

A one-way repeated measures ANOVA on participants' mean truth ratings revealed a significant main effect of repetition status (new, repeated from a trustworthy source, repeated from an untrustworthy source), $F(2, 986) = 120.94$, $p < .001$, $\eta_p^2 = 0.20$. As hypothesized, participants rated statements repeated from a trustworthy source ($M = 4.25$) as more true than new statements ($M = 3.74$), $t(493) = 15.59$, $p < .001$, $d = 0.72$, 95% CI [0.45, 0.57], and as more true than statements repeated by the untrustworthy source ($M = 3.73$), $t(493) = 11.30$, $p < .001$, $d = 0.51$, 95% CI [0.43, 0.61]. Note that, as pre-registered, these t-tests are one-tailed in line with our directional hypotheses. Turning to our final research question, we fail to find evidence that statements repeated by the untrustworthy source are rated differently than new statements, $t(493) = -0.23$, $p = .821$, $d = 0.01$, 95% CI [-0.08, 0.06].

Figure 2.7

Mean Truth Ratings by Repetition Status (Experiment 5)



Note. Ratings were coded from 1 = *Definitely False* to 6 = *Definitely True*. Each dot represents one participant (*N* = 494) with values horizontally shifted to represent the density distribution. Black diamonds reflect group means and error bars reflect standard errors of the mean.

Source Memory. Next, we report descriptive statistics regarding item and source memory, based on participants' responses to the memory questions. As in experiments 2-4, participants were sensitive to the difference between new and old items (mean *d'* = 1.15, *SD* = 0.95) but exhibited a general bias to categorize items as "old" (mean *c* = -0.77, *SD* = 0.56). Again, participants were above chance at recognizing the source of old items (*M* = 63.66%, *SD* = 22.15%; 0 exclusions).

In exploratory analyses, we compared these memory data to those from prior experiments. Full results are available in Appendix A, but to summarize, *d'* was higher in this experiment (M_{Exp2} = 0.76, $M_{Exp\ 3}$ = 0.69, $M_{Exp\ 4}$ = 0.90; *p*s < .001), as was source memory ($M_{Exp\ 2}$ = 52.25%, $M_{Exp\ 3}$

= 54.23%, $M_{\text{Exp 4}}$ = 60.56%; *p*s < .026). Thus, participants were better able to discriminate old from new items and were better able to remember the source of previously-seen items. While one possibility for the increase in source memory is that sources were more distinctive in terms of trustworthiness in this experiment, it is less clear why we observed differences in item memory.

Individual Differences in Truth Ratings. In experiments 2-4, 52-54% of participants gave higher truth ratings to statements from the trustworthy than the untrustworthy source. Here, a greater number of participants showed this effect, 64.0%. We again conducted an exploratory linear regression to predict variation in this effect of statement source (trustworthy versus untrustworthy) from participants relative trust in the two sources, their memory for the sources, and their need for cognition scores. As in Experiment 3, we found that the effects of statement source were greater for participants who showed a greater difference in trust for the trustworthy versus untrustworthy source (b = 0.22, $t(490)$ = 10.75, p < .001) and who were better able to remember the sources of the statements (b = 0.95, $t(490)$ = 5.16, p < .001). However, we again did not find an effect of the need for cognition measure (b = -0.01, $t(490)$ = -1.59, p = .112). Full results are available in Appendix A.

Discussion

In Experiment 5, we finally demonstrate a condition in which source evaluations can fully override the effects of repetition on belief. Specifically, we show that when participants learn via feedback that a source consistently states false information (and no true information), repetition of claims from this source do not increase belief.

Interim Discussion

Across five experiments, we examine how repetition of statements by trustworthy and untrustworthy sources affects belief. In Experiment 1, we find that prior exposure to a statement increases belief to a similar degree when it is later repeated by a trustworthy or untrustworthy source. In subsequent experiments, we examine the effects of the initial source of a statement on belief when that statement is later repeated. In Experiments 2-4, we show that repetition of a statement is less likely to increase belief when that statement was originally made by a source that participants learned to find untrustworthy (versus trustworthy). Further, in Experiment 5 we show that for extremely untrustworthy sources (i.e., sources that always state falsehoods), repetition no longer increases belief, posing a novel boundary condition for the illusory truth effect. Together, this pattern of findings suggests that the context in which statements are initial encountered seem to be most impactful for determining the effects of repetition on belief. Hearing a statement from someone you do not trust makes that exposure less impactful (Experiments 3-5), but once you have heard the statement, you will be more likely to believe it regardless of who repeats it (Experiment 1).

Importantly, the findings from Experiments 3-5 highlight limitations of current theories of the relationship between repetition and belief, which focus exclusively on low-level cognitive processes like processing fluency or familiarity and ignore the role of social information. These theories broadly suggest that repetition should increase belief regardless of who originally repeated the information, as repetition provides easily accessible cues for truth that are hard to disregard. In contrast, our data reveal that social evaluations about the initial source of a claim can deeply affect how people process and engage with information they are exposed to, and, ultimately, how these exposures affect their beliefs.

What mechanisms might allow source evaluations to have these effects? Our data are consistent with the possibility that source memory is a key process involved. Across Experiments 3-5, participants were less likely to trust repeated statements initial made by the untrustworthy versus trustworthy source. However, these effects were larger in Experiment 4 ($d = 0.16$) and 5 ($d = 0.51$), where participants were explicitly tasked with keeping track of who said each statement and were given a task (judgements of learning) that should enhance this associative memory as compared to Experiment 3 ($d = 0.09$). In addition, exploratory analyses revealed that in all experiments except for Experiment 3, participants with greater source memory showed greater relative belief in repeated statements originally attributed to the trustworthy versus untrustworthy source. Of course, the evidence from these cross-experimental comparisons and exploratory analyses warrants further confirmation.

Another possibility is that seeing a statement from an untrustworthy source triggers more thoughtful encoding processes, such that participants associate the statements with a sense of doubt. When participants then see the statement again to judge its truth, those initial doubts may be activated, reducing the effects of repetition on belief. Thus, in Experiments 4 & 5, we added an exploratory measure of participants' need for cognition in order to examine whether the effect of source would be larger for participants who are more likely to engage in this sort of skeptical, elaborative encoding processes. While we failed to find evidence of such a relationship, this does not rule out the possibility that untrustworthy sources may trigger more thoughtful encoding of presented information.

Overall, our results are consistent with the possibility that when people see statements that they had previously heard from a trustworthy source, they may use their memory for the source of that statement to discount the effects of repetition. We did not find evidence that people who are more prone to elaborative thinking were more likely to avoid the effects of repetition for untrustworthy sources. Thus, the current results suggest that the effects of source trustworthiness may be due more to memory processes than to elaborative thinking. Of course, more work is needed to precisely pinpoint the cognitive mechanisms by which social information may moderate the effects of repetition on belief. Still, these results are consistent with prior work showing that remembering information about the source of a statement can moderate the effects of repetition on belief (e.g., Begg et al., 1992; Unkelbach & Stahl, 2009).

Critically, this past work has examined the effects of source information that is directly and completely diagnostic of whether individual statements are true or false (e.g., instructions that statements in a man's voice are always false). Moving beyond this work, we consider people's evaluations of source trustworthiness. This source information is more social in nature and less diagnostic of the truth of individual statements than the source information used in prior work. When people learn that a source communicates mostly (or even only) false information in one situation, they can recognize that the same source may nonetheless be conveying true information at a later time. Still, our findings suggest that the reputation a source develops by communicating false information does affect how people process the information communicated by that source. Later repetition of these assertions appears to activate the distrust evoked by the memory for the sources' repetition, attenuating (and sometimes eliminating), the effects of repetition on belief.

In sum, our results highlight a previously underappreciated role for social evaluations in shaping belief in repeated information. While repetition of a statement alone provides powerful cognitive cues (through fluency, familiarity or referential coherence) that heighten belief, our data show that negative evaluations of the initial source of a statement can diminish—and potentially even eliminate—these effects of repetition.

CHAPTER 3

Social Consensus, Repetition, and Belief

This chapter is adapted from *Repeated by Many Versus Repeated by One: Examining the Role of Social Consensus in the Relationship Between Repetition and Belief*, forthcoming at the *Journal of Applied Research in Memory and Cognition.*

Introduction

Chapter 2 examined how social evaluations of the source of a single statement may moderate the effects of repetition on belief. This chapter turns to the broader context of repeatedly seeing information multiple times, potentially from multiple distinct sources. Thus, instead of examining the impacts of evaluations of a single source, this chapter will examine the effects of the *relationship between* sources. Specifically, we examine belief in statements that are repeated multiple times by the same source or multiple different sources. For instance, when hearing about a breaking news story, we may see a single news headline from the source several times, or we may see several different news outlets separately report on the event, each in their own wording. These different kinds of repetition connote different levels of consensus and agreement, but will this variation affect belief?

Current cognitive theories of the illusory truth effect largely suggest that such variation should not matter much: as long as an idea is encountered, belief should increase. In contrast to this picture, research on social influence, reviewed in Chapter 1, suggests that higher-level cues to social consensus should affect belief (e.g., Ransom et al., 2021), but it is unclear to what extent people can track these cues across repeated exposure to statements. In sum, there are open questions about how repetition will affect belief when that repetition reflects a consensus versus a single, repetitive source. The present chapter examines this question by manipulating consensus in two ways.

Cues to Consensus

Wording Variation. Experiment 1 begins by manipulating whether statements are repeated in verbatim or paraphrased form. At first, this wording manipulation may not seem to indicate consensus. Indeed, past work on the effects of paraphrased repetition on belief has typically manipulated wording variability in a way this unlikely to provide social cues. For instance, some studies (Silva et al., 2017; Vogel et al., 2020) used strict linguistic rules to transform statements from verbatim to paraphrased form (e.g., “The pigeon has a lifetime superior to that of a rabbit” to “A rabbit has a lifetime inferior to that of a pigeon”), finding that verbatim and paraphrased repetitions generally have similar impacts on belief. Other work has shown that repeating a contradictory version of a statement (e.g., “Nerthus is a German goddess of [earth/water].”) can increase belief when participants fail to recall they are seeing a contradiction (e.g., after a long delay; Garcia-Marques et al., 2015; Unkelbach & Rom, 2017). Together, these findings highlight importance of conceptual over perceptual process in the illusory truth effect: as long as an idea is easier to process, it seems more true, regardless of how exactly it is worded. However, they do not shed light on the potential impacts of wording variation that may convey social information.

Thus, unlike these prior experiments, the present experiment used more naturalistic paraphrased stimuli, drawn from different news outlets reporting on the same event. These stimuli more closely resemble the type of non-verbatim repetition of information seen in daily life, in which information varies more drastically in its tone, details and wording (e.g., “A study found that an iPhone 12 can disable a cardiac rhythm management device” versus “Cardiologists Find

Apple iPhone 12 Magnet Deactivates Implantable Cardiac Devices"), creating greater perceptual variability. Critically, the paraphrased headlines may also convey social information, like the fact that these different headlines come from different sources who report on information with their own perspective, style, and tone.

Source Variation. In addition to variation in the wording of repeated information, we also examine a more direct manipulation of the level of consensus: variation in the distinct sources sharing a statement. Past work has manipulated this factor by, for instance, varying the number of different social media users agreeing with an idea (e.g., Ransom et al., 2021; Simmonds et al., 2023) or the number of unique news sources reporting on a topic (Connor Desai et al., 2022; Yousif et al., 2019). Experiments 2 & 3 use the former manipulation: varying the number of distinct social media users who post about a topic. Each time participants see a statement, it is associated with an image of a person sharing the information, and we manipulate whether the same person shares each headline or different people share the headline each time. Critically, and unlike in past work, we are interested in examining this form of consensus across distinct encounters with a statement, rather than during a single event (e.g., seeing a single social media post with one or many people agreeing with it). It is unclear whether cues to consensus can still be influential in such a context, where direct comparison of views across sources is more difficult.

Present Research

Across three experiments, we examine whether the effects of repetition on belief will be moderated by variation in the phrasing and/or number of unique sources—manipulations designed to convey varying levels of social consensus. In Experiment 1, participants read news headlines that were repeated verbatim or in paraphrased forms derived from different news outlets. In Experiment 2, participants encountered verbatim repetitions of headlines shared by one person multiple times or by multiple different people. Finally, in Experiment 3, participants encountered verbatim repetitions of headlines from a single source or paraphrased repetitions from multiple sources. The key question across experiments is how participants then rate the truth of these headlines relative to each other and to new headlines that were not previously encountered.

If higher-level inferences about social consensus matter for truth judgements, varying the phrasing or source of repeated headlines should increase belief. By contrast, if repetition affects belief through lower-level cognitive processes (i.e., fluency) alone, neither manipulation should increase belief. The conceptual fluency account suggests that repetition should increase belief so long as the same ideas are repeated, regardless of their wording or source. The perceptual fluency account suggests that paraphrased repetition should be *less* impactful, since perceptual features of the statement change on each exposure, and broadly suggests that source variability should not matter. Table 3.1 summarizes these predictions. To preview, our results are most consistent with a conceptual fluency account: repetition increases belief to a similar degree, regardless of the level of variation in wording or source.

Table 3.1

Predictions Across Experiments

Account	Experiment 1	Experiment 2	Experiment 3
Fluency Alone			
Conceptual	Verbatim = Paraphrased	1 source = 3 sources	Verbatim, 1 source = Paraphrased, 3 sources
Perceptual	Verbatim > Paraphrased	1 source = 3 sources	Verbatim, 1 source > Paraphrased, 3 sources
Social Consensus	Verbatim < Paraphrased	1 source < 3 sources	Verbatim, 1 source < Paraphrased, 3 sources

Note. Key predictions about the effects of different kinds of repetition on belief. Cells indicate which conditions are predicted to result in the highest perceived truth.

Open Practices

The hypotheses, design and analysis plan for all experiments were pre-registered. The pre-registration documents are available at the project's Open Science Framework (OSF) site (https://osf.io/z7bqy/?view_only=bb5273e426e3478585fd1b2438b13779), along with the materials, participant instructions, data, and analysis code for each experiment. For all experiments, we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Experiment 1

Experiment 1 compares the effects of verbatim versus paraphrased repetition on belief. Participants read a set of headlines repeated three times in verbatim or paraphrased form. They then rated the truth of the verbatim headlines, a fourth unique version of the paraphrased headlines, and some completely new headlines.

We predicted that headlines repeated in verbatim or paraphrased form would be rated as truer than new statements, replicating the illusory truth effect. Critically, we also predicted that participants would provide higher truth ratings for headlines repeated verbatim as compared to those repeated in paraphrased form, as verbatim headlines should receive the greatest boost in perceptual fluency.

Method

All experiments received ethics approval under IRB #170586 at Vanderbilt University.

Participants.

Statistical Power. Our pre-registered sample size was based on an *a priori* power analysis in G*Power (Faul et al., 2009), which revealed that 262 participants was needed to achieve 80% power to detect a an effect size of $f = .055$ in a two-group repeated measures ANOVA.

This two-group repeated measures ANOVA is equivalent to the paired *t*-test we planned to run comparing participants' mean responses to items in the repeated paraphrase and repeated verbatim conditions. We were minimally interested in a difference between these conditions of 0.1 points on our 6-point scale, and, assuming a standard deviation of these ratings of 0.88 from past studies with similar materials and numbers of items (Pillai et al., under review), this difference corresponds to a minimal effect size of interest of $f = .055$. For context, this minimal effect size of interest is about 4.5 times smaller than the overall expected effect of repetition (verbatim repetition vs. new) based on past meta-analytic evidence (Dechêne et al., 2010). Finally, our power analysis also assumed a correlation among repeated measures of 0.8 based on prior research (Pillai, et al., under review) and no correction for non-sphericity.

Recruitment. 262 adult participants were recruited from Amazon's Mechanical Turk (MTurk) platform to complete the experiment through the CloudResearch platform (Litman et al., 2017a) for a payment of $1.81. To ensure data quality, we recruited participants from CloudResearch's approved participants list (Peer et al., 2021) and excluded participants (not counting towards the above sample size) for failing two attention checks at the beginning of our survey (typing a response to "Puppy is to dog as kitten is to ___?" and selecting two requested responses on a 5-point multiple choice question).

Demographics. The mean age of participants was 39.69 (*SD* = 11.37; Range = 18-76). Our final sample was predominantly White (76%, 9.5% Black, 8.8% Asian, 2.3% Multiracial, 1.5% Other, 1.5% not reporting) and non-Hispanic (89%, 9.9% Hispanic, 0.8% not reporting), and most participants were men (51%, 46% women, 0.8% nonbinary, 2.3% not reporting). Most participants (72%) had at least a college degree.

Design. We manipulated repetition type (new, repeated verbatim, repeated paraphrase) within-subjects. We counterbalanced repetition across participants by splitting our 36 items into three sets of 12 and rotating these sets through each level of repetition type. This created three possible counterbalancing groups to which we assigned participants.

Materials. Stimuli consisted of news headlines describing 36 different events or facts that were either confirmed by the third-party fact-checking site Snopes as "true" or were reported by various reputable mainstream news outlets (e.g., The New York Times, The Washington Post). Example headlines include "Exquisitely-preserved wolf pup mummy discovered in Yukon permafrost" and "'Cocaine bananas' accidentally shipped to grocers in bungled drug deal". Note that participants were never shown the original source of the headline—only the text of the headline itself.

Each of the 36 events had four different headlines, each from a different online source. (e.g., "Cocaine found in banana shipment part of drug deal gone bad," "Cocaine-stuffed shipments of bananas ended up at Canadian grocery stores due to a drug-trafficking mix-up.," "'Cocaine bananas' accidentally shipped to grocers in bungled drug deal," "'Cocaine bananas' shipped to grocery stores in botched operation"). For each of the 36 items, we randomly selected one of the four versions as the key headline that was shown during the rating phase. For new items, this is the only version participants saw in the experiment. For repeated verbatim items, participants saw the key headline three times during the exposure phase and then again in the rating phase. For repeated paraphrased items, participants saw the other three versions of the headline during the exposure phase followed by the key headline in the rating phase. Table 3.2 shows an example item in each of these three conditions. Note that each participant saw only one of these three conditions (new, repeated verbatim or repeated paraphrased) for a given item.

Table 3.2

Sample Headlines

Condition	Exposure Phase			Rating Phase
New				A Hacker Tried to Poison a Florida City's Water Supply, Officials Say
Repeated Verbatim	A Hacker Tried to Poison a Florida City's Water Supply, Officials Say	A Hacker tried to Poison a Florida City's Water Supply, Officials Say	A Hacker tried to Poison a Florida City's Water Supply, Officials Say	A Hacker tried to Poison a Florida City's Water Supply, Officials Say
Repeated Paraphrased	In Florida City, a Hacker Tried to Poison the Drinking Water	Someone tried to poison Oldsmar, Florida's water supply during hack, sheriff says	Feds tracking down hacker who tried to poison Florida town's water supply	A Hacker Tried to Poison a Florida City's Water Supply, Officials Say

Note. Participants saw a given headline in one of the three conditions.

Procedure. This experiment was administered online via Qualtrics.

Exposure Phase. After reading the information sheet and completing two attention checks, participants were instructed that we wanted to get their opinion on "various claims that have been posted online." As all headlines were true, we did not inform participants of the truth of the headlines they were about to see. Starting on the following screen, participants saw 72 headlines, one at a time, in the center of the screen above the question "How interesting is the headline above?" Participant selected from the options *Very Uninterested*, *Uninterested*, *Slightly Uninterested*, *Slightly Interested*, *Interested*, or *Very Interested* to proceed. The 72 headlines consisted of 12 items repeated three times verbatim, and 12 items repeated three times in paraphrased form.

Rating Phase. Immediately after the exposure phase, participants began the rating phase. Participants were correctly informed "some of the headlines you will have seen in the previous section, others will be new." Again, participants were not informed about the truth of the headlines. Participants then saw the 36 key headlines, one at a time, above the question "How true or false is the headline above?" and selected from the options *Definitely False, Probably False, Possibly False, Possibly True, Probably True,* or *Definitely True* (scored from 1 to 6 in our analyses). Twelve headlines were new (i.e., shown for the first time on the rating phase), 12 were repeated verbatim in the exposure phase, and 12 were repeated in paraphrased form in the exposure phase.

Participants then answered some optional demographic questions (gender, race, ethnicity, education level) and were asked a few debriefing questions (what they thought the study was about, whether they noticed the same statement multiple times in the first phase, whether they noticed different versions of the statement in the first phase). Finally, participants were thanked for their time and informed about the purpose of the study.

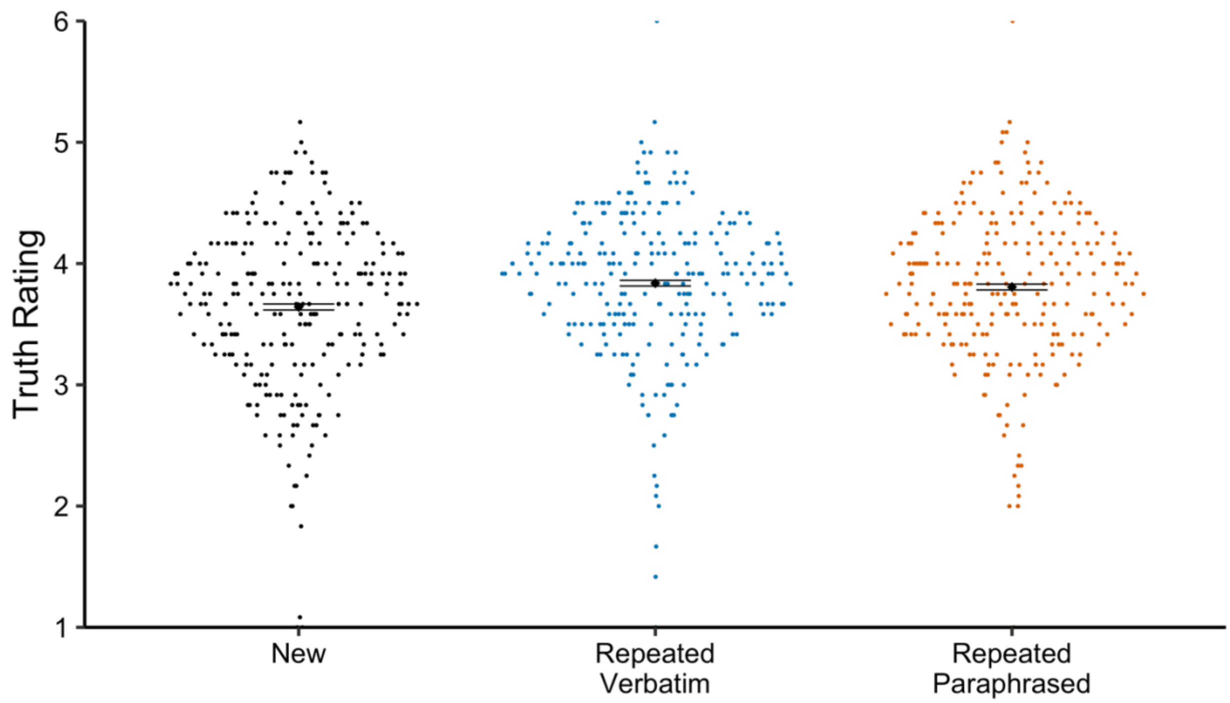
Results

For all experiments, all statistical tests are conducted at the .05 alpha level and are pre-registered unless labelled as exploratory. ANOVA tests were conducted using *rstatix* version 0.7.0 (Kassambara, 2020), frequentist *t*-tests were conducted in base R version 3.6.3 (R Core Team, 2020), and Bayesian *t*-tests were conducted using *BayesFactor* version 0.9.12 (Morey et al., 2015).

We hypothesized that, in line with the illusory truth effect, repetition (in either verbatim or paraphrased form) would increase belief. As shown in Figure 3.1, this was the case. Critically, we also predicted that headlines repeated in verbatim form would be perceived as more true than repeated paraphrased headlines. Contrary to our hypothesis, truth ratings were similar across the verbatim and paraphrased headlines.

Figure 3.1

Mean Truth Ratings by Repetition Status (Experiment 1)



Note. Ratings were provided on a scale from 1 = *Definitely False* to 6 = *Definitely True*. Each dot represents one participant (N = 262) with values horizontally shifted to represent the density distribution. Black diamonds reflect group means and error bars reflect standard errors of the mean.

A one-way ANOVA revealed a significant main effect of repetition type (new, repeated verbatim, repeated paraphrase), $F(2, 522) = 17.58$, $p < .001$, $\eta_p^2 = 0.06$. Relative to new items (M = 3.64), participants gave higher truth ratings to items repeated verbatim (M = 3.84), $t(261) = 5.08$, $p < .001$, 95% confidence interval of the difference (CI) [0.12, 0.27], , d = 0.27 and items repeated in paraphrased form (M = 3.81), $t(261) = 4.41$, $p < .001$, 95% CI [0.09, 0.24], d = 0.31. However, we did not observe a significant difference in ratings for items repeated in verbatim versus in paraphrased form, $t(261) = 1.05$, $p = .295$, 95% CI [-0.03, 0.09], $d = 0.06$

To follow up on this null result, we conducted an exploratory Bayesian *t*-test comparing participant's mean ratings for items repeated in verbatim versus paraphrased form. Using the default Cauchy distribution with width 0.707 for our prior probability distribution, we calculated a Bayes factor of 8.40 in favor of the null hypothesis. That is, our data are 8.40 times more likely under the null hypothesis (that ratings are identical in the repeated verbatim and repeated paraphrased conditions) than under the alternative (that there is a difference). This Bayes Factor is in the range generally considered moderate evidence in favor of the null hypothesis (Held & Ott, 2018).

Discussion

In Experiment 1, repeating an idea using the same or different wording increased belief to a similar degree, consistent with a conceptual fluency account. These results are inconsistent with a perceptual fluency account, which predicts that verbatim repetition should be most powerful. The results are also inconsistent with a social consensus account, which predicts that paraphrased headlines should indicate endorsement by different sources, increasing belief. However, there are two important limitations of Experiment 1. First, our manipulation may have evoked multiple mechanisms. For the paraphrased headlines, social consensus may have increased belief and a lack of perceptual fluency may have decreased belief, with both effects offsetting each other. Second, participants may not have inferred that the different wordings indicated different sources.

Experiment 2

Thus, Experiment 2 examines a different, more direct manipulation of consensus: the number of unique people sharing a headline. To our knowledge, only one prior paper has examined the role of source variability in the effects of repetition on belief (Roggeveen & Johar, 2002), finding no effect in one study (Experiment 2) and a limited effect in another (Experiment 3). Critically, in the former study, sources were present during both the initial exposure and the rating phase. Thus, social consensus was confounded with the level of perceptual overlap between the repeated stimuli. In addition, the latter study manipulated repetition between-subjects, a design that minimizes the effects of repetition on belief (Dechêne et al., 2009).

Addressing these limitations, in Experiment 2, participants read a news headline three times (verbatim), alongside the same source or a different source each time. Then, they rated the perceived truth of these headlines, along with new headlines, without any accompanying source.

By the social consensus account, people should be most likely to believe statements repeated by different sources. By contrast, a conceptual fluency account suggests that repetition should increase belief regardless of who shared it. The perceptual fluency account makes the same prediction: because all statements were rated without a source, they should be similarly perceptually fluent due to prior exposure.

We again predicted that repetition from one or three sources would increase belief. However, we did not make an explicit prediction about whether one kind of repetition would be more impactful. Finally, to examine whether people can keep track of the variation in sources, we implemented a memory check in which participants are asked to report how many different people shared each headline.

Method

Participants.

Statistical Power. Our pre-registered sample size was based on an *a priori* power analysis in G*Power, which indicated that 229 participants were needed to achieve 80% power to detect an effect size of f = .08 in a two-group repeated measures ANOVA (equivalent to a paired t-test comparing items repeated from one or three sources).

This power analysis is identical to that of Experiment 1 except in two ways. First, the minimal effect size of interest increased from f = .055 to f = .08, as we used the observed *SD* of truth rating data in Experiment 1 (*SD* = .62 for repeated verbatim & paraphrased items, versus *SD* = 0.88 used in the power analysis in Experiment 1) . The minimal effect size of interest still corresponds to an absolute difference of 0.1 points on our 6-point scale. Second, we reduced the correlation among repeated measures from 0.8 to 0.63, again based on the observed correlation value for repeated verbatim and repeated paraphrased items in Experiment 1.

Recruitment. 229 adult participants were recruited in the same manner as described in Experiment 1, except participants received $2.42 for completing this longer experiment. Four additional participants were excluded for failing our two attention checks (typing the name of the animal depicted by a black and white cartoon image and selecting a requested responses on a 5-point Likert item) and did not count towards our final sample.

Demographics. The mean age of participants was 39.76 (*SD* = 11.11; Range = 19-77). Our final sample was predominantly White (81 %, 7.4% Asian, 7.0 % Black, 3.1% Multiracial, 0.4% Other, 1.3% not reporting) and non-Hispanic (94 %, 4.4% Hispanic, 1.8% not reporting), and most participants were women (52 %, 46 %, 0.9% nonbinary, 0.9 % not reporting). Most participants (71%) had at least a college degree.

Design. We manipulated repetition type (new, repeated from one source, repeated from three sources) within-subjects. As in Experiment 1, participants were assigned to one of three counterbalancing groups, created by splitting items into three sets and rotating them through the three levels of repetition type.

Materials. We used the 36 key headlines shown in the rating phase of Experiment 1. Note that only one version of each headline was used in this experiment.

During the exposure phase, headlines were paired with sources, which were full-body photographs created by Connor et al. (2021). The full set includes 454 photos of people whose race (Asian, Black, White) and gender (male, female) were noted. We selected 48 photos for this experiment, attempting to evenly sample across all combinations of race and gender present in the database. However, as the set only consisted of three photos of Asian women, we selected all three, and then randomly selected nine photos within each of the remaining five combinations of the photo subject's gender and race.

For each participant, 12 of the 48 sources were randomly assigned to the "repeated from one source" condition and the remaining 36 to the "repeated from three sources" condition. In the one source condition, the 12 headlines were randomly paired with a single source and repeated three times with that same source during the exposure phase. In the three-source condition, each of the three instances of the 12 headlines had a different, unique source, exhausting all 36 remaining sources. No source was paired with more than one headline.

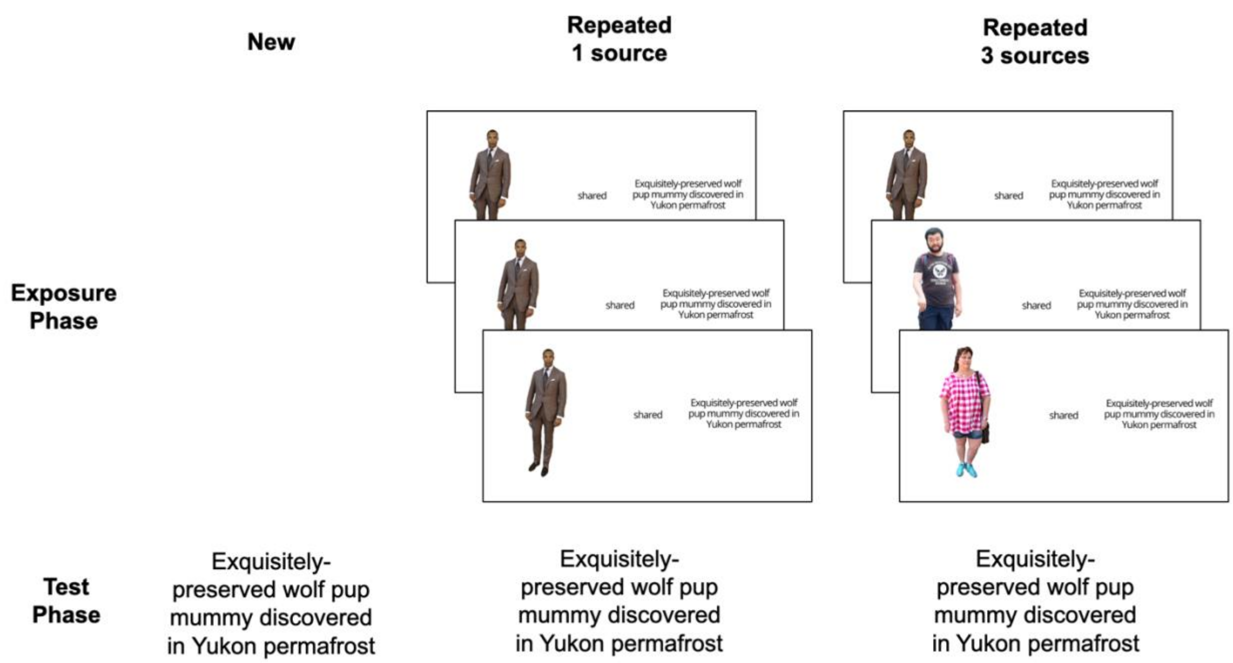
Procedure. This experiment was administered online using the gorilla.sc platform (Anwyl-Irvine et al., 2020b).

Exposure Phase. As in Experiment 1, participants began by reading an information sheet, completing two attention checks, and receiving instructions to rate their interest in "various claims that have been posted online." Unlike in Experiment 1, participants were also told that headlines would appear next to a photo of a person who shared it online and were told "Please pay attention to who shared each headline. We will ask you some questions about who shared each headline at the end of this experiment." We again did not inform participants of the truth of the headlines.

Then, participants saw 72 headlines, one at a time, and were asked “How interesting is the headline above?” (*Very Uninteresting*, *Uninteresting*, *Slightly Uninteresting*, *Slightly Interesting*, *Interesting*, or *Very Interesting*). Unlike in Experiment 1, headlines were placed next to a full body photo of a person, as shown in Figure 3.2. As described above, 12 headlines were repeated three times with the same source and 12 were repeated three times with different sources each time. All headlines were repeated verbatim.

Figure 3.2

Sample Headlines



Note. Participants saw a given headline in one of the three conditions. The exact sources associated with any headline were randomized on a participant-by-participant basis.

Rating Phase. The rating phase was identical to that of Experiment 1, except that we also told participants, “we will not indicate who shared each headline” . All participants rated the truth of 36 headlines by responding to the question “How true or false is the headline above?” (*Definitely*

False, Probably False, Possibly False, Possibly True, Probably True, or *Definitely True*, scored from 1 to 6).

Memory Check. After the rating phase, participants completed an additional, exploratory measure of their memory for the sources of each headline. Participants were shown 12 headlines (randomly selected for each participant), with even numbers of headlines that were new, repeated from one source, and repeated from three sources. For each of the 12 headlines, participants were asked "How many different people shared this headline?" (0, 1, 2, or 3). Participants were instructed that this task referred to the sources that shared each headline during the first part of the experiment (exposure phase).

Finally, participants answered optional demographic questions (gender, race, ethnicity, education), a debrief question (what they thought the purpose of the study was) and were informed about the purpose of the study.

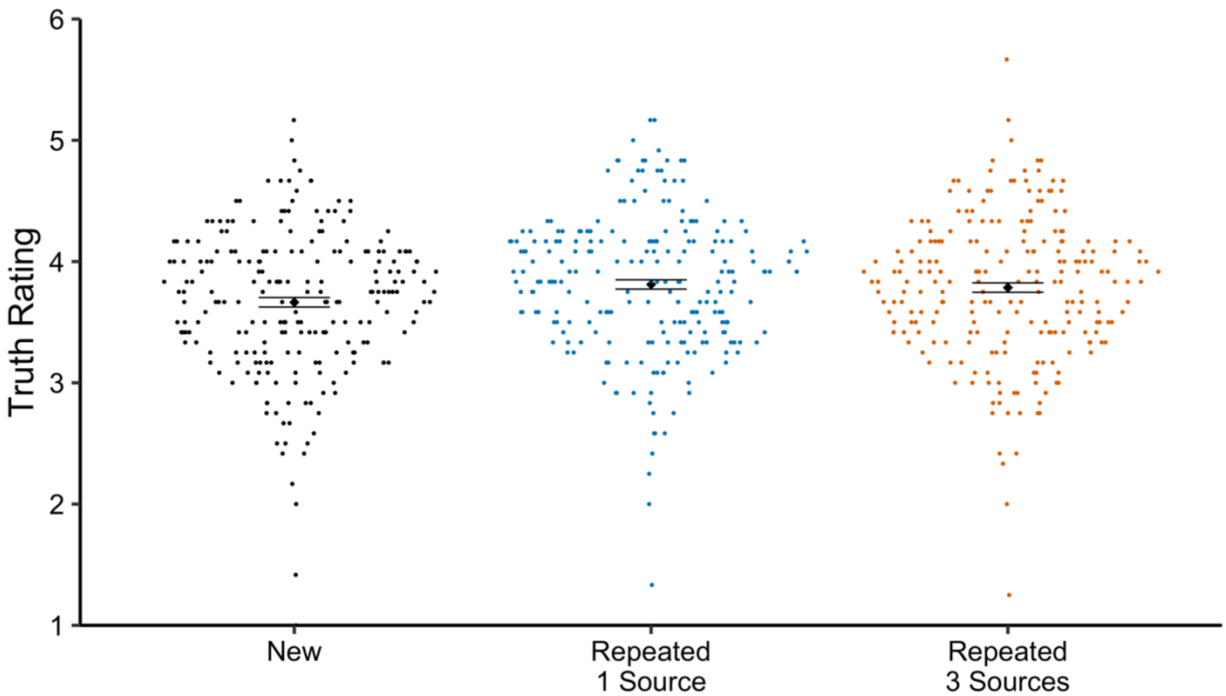
Results

Truth Ratings. We hypothesized that, in line with the illusory truth effect, repetition of headlines from a single source or from multiple sources would increase belief. As shown in Figure 3.3, this was this case. Our main research question was whether headlines repeated from a single source would be perceived as more or less true than headlines repeated from multiple sources. As shown in Figure 3.3, ratings were similar across these two conditions.

A one-way ANOVA revealed a significant main effect of repetition type (new, repeated from one source, repeated from three sources), $F(2, 456) = 11.50$, $p < .001$, $\eta_p^2 = 0.05$. Relative to new items ($M = 3.66$), participants gave higher truth ratings to headlines repeated from one source ($M = 3.81$), $t(228) = 4.41$, $p < .001$, 95% CI [0.08, 0.21], $d = 0.29$ and headlines repeated from three sources ($M = 3.79$), $t(228) = 3.47$, $p < .001$, 95% CI [0.05, 0.19], $d = 0.23$ However, we did not observe a significant difference in ratings for headlines repeated from one versus three sources, $t(228) = 0.87$, $p = .387$, 95% CI [-0.03, 0.08] , $d = 0.06$

Figure 3.3

Mean Truth Ratings by Repetition Status (Experiment 2)



Note. Ratings were provided on a scale from 1 = *Definitely False* to 6 = *Definitely True*. Each dot represents one participant (N = 229) with values horizontally shifted to represent the density distribution. Black diamonds reflect group means and error bars reflect standard errors of the mean.

Like in Experiment 1, we followed up on this null result by conducting an exploratory Bayesian *t*-test comparing ratings for headlines repeated for one versus three sources. The Bayes factor for this *t*-test was 9.33, suggesting our data are 9.33 times more likely under the null hypothesis. Again, we find moderate evidence in favor of the null hypothesis that the two kinds of repetition we tested do not have different effects on belief.

Memory for Source Variability. Finally, to verify that participants were able to remember the extent to which there was social consensus around different claims, we conducted an exploratory analysis on participants' responses to the memory check ("How many different people shared this headline?" (0, 1, 2, or 3)). Overall, while estimates were not very accurate, participants were able to qualitatively distinguish between the new, single-source and three-source headlines. Participants provided higher estimates for headlines that were repeated from a single source (M = 1.92) relative to new headlines (M = 0.68), $t(228) = 20.55$, $p < .001$, 95% CI [1.12, 1.36], $d = 1.36$. In addition, participants provided higher estimates for headlines that were repeated from three sources (M = 2.11) relative to headlines that were repeated from a single source, $t(228) = -5.24$, $p < .001$, 95% CI [0.12, 0.26], $d = 0.35$

While participants were able to qualitatively distinguish between headlines across the different conditions, memory was not particularly accurate. For instance, participants overestimated the number of unique sources for headlines repeated by a single source on average (M = 1.92). Given this, we were interested in examining whether the key pattern of truth ratings reported in Figure 3.3 would differ for participants with more accurate memory for the number of sources. To this end, we first calculated the effect of repetition type (difference score for average rating of statements repeated by three sources minus that of statements repeated by one source) at the participant level. Then, we regressed this measure on participants' relative memory for these statements (difference score for source memory ratings of the same two conditions). Ultimately, we do not find a significant effect of source memory on the effect of repetition type ($b = 0.02$, $t(227) = 0.44$, $p = .658$).

Complementing this approach, we also focused on participants with the largest difference in memory responses between the one-source and three-source conditions (top one third of the sample; *N* = 62)[3]. Note that even among these participants memory was not particularly accurate. The average difference between memory ratings was 0.95 (SD = 0.23), while perfect memory would be a difference of 2 (3 – 1). We then conducted an exploratory *t*-test comparing truth ratings for headlines repeated by one versus three sources among these participants As in our main analyses, we found no difference in truth ratings for headlines repeated by one ($M = 3.88$) versus three ($M = 3.90$) sources in this condition, $t(61) = -0.12$, $p = .825$, 95% CI [-0.12, 0.10], $d = 0.01$, and we observed moderate Bayesian evidence in favor of this null effect ($BF_{01} = 7.02$).

Discussion

Repetition increased belief regardless of whether it came from one person or many, contrary to predictions that social consensus would magnify the effects of repetition on belief. Experiment 2 is again most consistent with fluency-based accounts in which repetition increases belief simply by making statements easier to process.

Still, there remains one important limitation to our consensus manipulation. Even when participants saw headlines from different sources, these different sources repeated the same headline verbatim. In this way, the consensus may not have been perceived as coming from sources who independently endorsed the information. Other work suggests people are less sensitive to social consensus cues when multiple sources co-depend on the same data (Connor Desai et al., 2022; Whalen et al., 2018). Thus, Experiment 3 has different sources convey paraphrased versions of each headline, making the consensus more likely to reflect a process by which multiple people separately decided to share the same idea. This sort of "independent" consensus may seem more

[3]Note that memory difference scores for participants were fairly discrete, taking on values from -1.25 to 2 in increments of 0.25. Thus, in selecting a cutoff for the top tertile of responses, we ended up with slightly less than one third of our full sample size.

informative and thus more likely to affect judgements, providing a stronger test of the social consensus hypothesis.

Experiment 3

Experiment 3 replicates Experiment 2, with a few modifications. Primarily, headlines shared by three different sources are now paraphrased repetitions, rather than verbatim. We also added exploratory measures to examine why there might be a disconnect between memory for the number of sources and accuracy ratings for the headlines. Recall that we noted three steps needed for social consensus to affect belief. First, people must track whether the information comes from a single source or from multiple sources. Second, people must attribute this variation to different levels of social consensus. Finally, people must incorporate this consensus into their beliefs. Data from Experiment 2 suggests that the first process is taking place: people can differentiate claims repeated by many people versus one person. Here, we examine the second process: does this memory translate into perceptions of the level of endorsement around the repeated claims?

To address this question, we added three exploratory measures. First, we directly asked participants to estimate what proportion of social media users would agree with a subset of headlines. We also added two questions asking participants why they thought people shared the news headlines they saw, to see if participants thought the sources shared the news without actually *endorsing* its accuracy.

We again predicted that either kind of repetition would increase belief. However, we did not make any prediction about whether headlines repeated in paraphrased form from three different people would be perceived as more or less true than headlines repeated verbatim from a single source.

Method

Participants.

Statistical Power. Our pre-registered sample size was based on an *a priori* power analysis in G*Power, which indicated that 226 participants were needed to achieve 80% power to detect an effect of $dz = 0.19$ in a matched-pairs t-test. Note that this power calculation is identical to that reported in Experiment 2, except that we directly used the matched-pairs t-test option in G*Power rather than the two-group repeated measures ANOVA option. These two power calculations are mathematically equivalent, but due to rounding differences, produce slightly different sample size estimates (226 versus 229).

Recruitment. 227 participants were recruited via the CloudResearch Connect platform and received $3.02 for completing the experiment. Note that this is one higher than our pre-registered sample size due to an additional participant completing the experiment before the posting formally closed on Connect. As in Experiments 1 & 2, participants were excluded (not counting towards our final sample size) if they failed two attention checks.

Demographics. The mean age of participants was 38.93 ($SD = 12.56$; Range = 18-77). Our final sample was predominantly White (76%, 12% Black, 5.7% Asian, 2.6% Multiracial, 0.88% reporting some other race, 2.2% not reporting) and non-Hispanic (90%, 6.5% Hispanic, 3.5% not reporting), and most participants were men (55%, 40% women, 1.3% nonbinary, 4.4% not reporting). Most participants (63%) had at least a college degree.

Design. We manipulated repetition type (new, repeated verbatim from one source, repeated paraphrased from three sources) within-subjects. As in Experiments 1 and 2, participants were assigned to one of three counterbalancing groups, created by splitting items into three sets and rotating them through the three levels of repetition type.

Materials. Stimuli consisted of the 36 news headlines (each with four paraphrased versions) used in Experiment 1. As in Experiment 1, one key version of each headline was used in the rating phase. Depending on the condition, participants either saw this headline three times, saw three other versions of the headline, or did not see the headline at all during the exposure phase.

In addition, we used the same 48 full-body photographs of people as in Experiment 2.

Procedure. The procedure was identical to that of Experiment 2 with a few exceptions. First, during the exposure phase, the headlines shared by different sources were all shared in paraphrased wording each time.

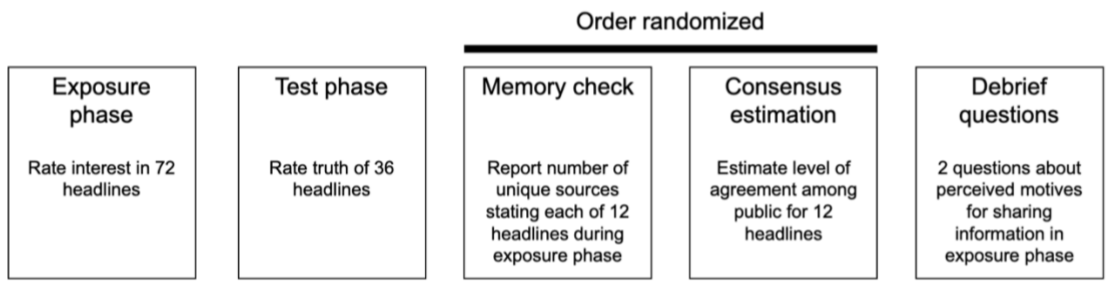
Second, we changed the wording of the memory check, as headlines were repeated not only multiple times, but also in different wordings. Thus, participants answered the question "How many different people shared this headline or a similar headline that contained the same idea" (0, 1, 2, or 3).

Third, in addition to the memory check, we added an additional "social consensus estimation" task. Participants were shown 12 headlines and asked for each "What percent of social media users do you think would believe this headline to be true?" entering their response using a slider from 0 to 100 (default set to 50). To avoid repeating headlines between the memory check and social consensus, we split our 36 key headlines into three sets of 12 (each containing even numbers of headlines per repetition condition), and randomly showed participants one set for the memory check and another for the social consensus estimation task. In addition, the order of these two tasks was randomized across participants.

Finally, after the social estimation and memory tasks, we added two exploratory debrief questions. First, we asked participants the open-ended question "What do you think the main motivation was for people to share the headlines? In other words, why did these social media users decide to share the headlines you saw?" Next, we asked participants "How important do you think it was to these social media users that the information they share is accurate?" (1-10 Likert scale with anchors "*not very important*" and "*very important*"). Figure 3.4 depicts all phases of this experiment.

Figure 3.4

Phases of Experiment 3



Results

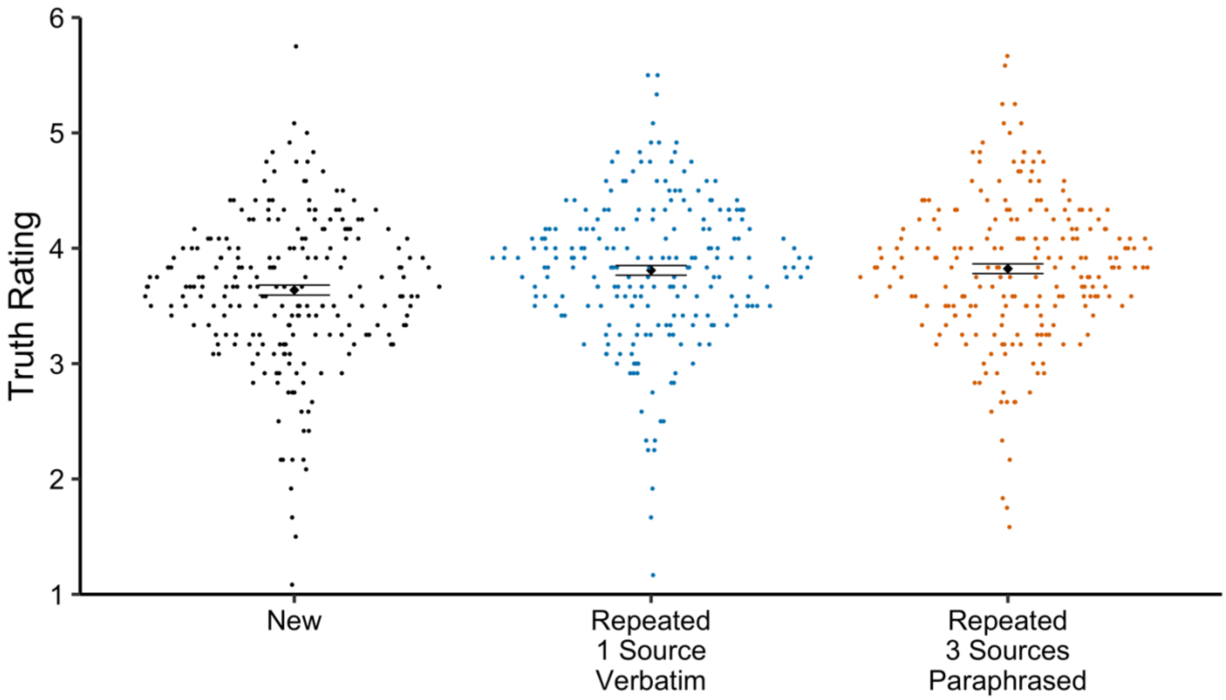
Truth Ratings. We hypothesized that, in line with the illusory truth effect, repetition of headlines from a single source or from multiple sources would increase belief. As shown in Figure 3.5, this was this case. Our main research question was whether headlines repeated verbatim from a single source would be perceived as more or less true than headlines repeated in paraphrased form from multiple sources. As shown in Figure 3.5, ratings were similar across these two conditions.

A one-way ANOVA revealed a significant main effect of repetition type (new, repeated verbatim from one source, repeated paraphrased from three sources), $F(2, 452) = 18.13$, $p < .001$, $\eta_p^2 = 0.07$. Relative to new items ($M = 3.64$), participants gave higher truth ratings to headlines repeated verbatim from one source ($M = 3.81$), $t(226) = 4.91$, $p < .001$, 95% CI [0.10, 0.24], $d = 0.33$ and headlines repeated in paraphrased form from three sources ($M = 3.82$), $t(226) = 5.12$, $p <$

.001, 95% CI [0.11, 0.26], $d = 0.34$. However, we did not observe a significant difference in ratings across the two kinds of repeated headlines, $t(226) = -0.45$, $p = .650$, 95% CI [-0.08, 0.05], $d = 0.03$.

Figure 3.5

Mean Truth Ratings by Repetition Status (Experiment 3)



Note. Ratings were provided on a scale from 1 = *Definitely False* to 6 = *Definitely True*. Each dot represents one participant (N = 227) with values horizontally shifted to represent the density distribution. Black diamonds reflect group means and error bars reflect standard errors of the mean.

As in Experiments 1 & 2, we followed up on this null result by conducting a Bayesian *t*-test comparing ratings for headlines in the two repeated conditions. The Bayes factor for this *t*-test was 12.16 in favor of the null hypothesis, suggesting our data are 12.16 times more likely under the null hypothesis. Again, we find moderate evidence that the two kinds of repetition we tested do not have different effects on belief.

Memory for Source Variability. Like in Experiment 2, we sought to verify that participants were able to remember whether there was social consensus around different claims. Thus, we conducted exploratory analyses on participants' responses to the memory check ("How many different people shared this headline or a similar headline that contained the same idea?" (0, 1, 2, or 3)). Again, while estimates were not very accurate, participants were able to qualitatively distinguish between the new, single-source and three-source headlines. Participants provided higher estimates for headlines that were repeated from a single source ($M = 1.84$) relative to new headlines ($M = 0.73$), $t(226) = 18.94$, $p < .001$, 95% CI [0.99, 1.23], $d = 1.26$. In addition, participants provided higher estimates for headlines that were repeated from three sources ($M = 2.11$) relative to headlines that were repeated from a single source, $t(226) = 6.61$, $p < .001$, 95% CI [0.19, 0.35], $d = 0.44$.

We again observed that while participants qualitatively distinguished between different kinds of headlines, memory was not particularly accurate. Again, we regressed participants relative truth rating for statements repeated by three versus one source on their difference in source memory ratings across the two conditions. Again, we do not find a significant effect of source memory on the effect of repetition type ($b = 0.02$, $t(225) = 0.48$, $p = .635$). Then, like in Experiment 2, we conducted an exploratory *t*-test comparing truth ratings for headlines repeated by one versus three sources among participants with the most accurate memories (the top third of participants in terms of the difference between estimated number of sources for one vs three source items; mean difference = 1.09, SD = 0.32; $N = 59$). As in our main analyses, we found no difference in truth ratings for headlines repeated by one ($M = 3.83$) versus three ($M = 3.81$) sources among these participants, $t(61) = 0.24$, $p = .811$, 95% CI [-0.11, 0.14], $d = 0.03$, and we observed moderate Bayesian evidence in favor of this null effect ($BF_{01} = 6.83$).

Perceptions of Social Consensus. We next conducted a series of exploratory analyses directly examining participants estimates of the level of social consensus around different claims. Recall that we asked participants "What percent of social media users do you think would believe this headline to be true?" For "new" headlines, participants estimated that 56.2% of social media users would believe the headline. Participants thought more social media users would believe headlines repeated verbatim from one source (60.0%), $t(226) = 3.60$, $p < .001$, 95% CI [1.72, 5.86], $d = 0.24$. However, participants were no more likely to think social media users would believe a headline when it was repeated in paraphrased form by three sources (59.9%) than when it was repeated verbatim from one source, $t(226) = -0.09$, $p = .924$, 95% CI [-2.09, 1.89], $d = 0.01$.

These results suggest a disconnect between memory and perceptions of social consensus—even though participants can distinguish headlines repeated by multiple versus a single person, they do not attribute these differences to different levels of endorsement. One possible explanation is that participants simply thought that the sources in the exposure phase were not sharing information they believed. Indeed, when asked an open-ended question about the sources' main motive for the sharing the headlines, only 11.5% of participants spontaneously reported that the sources shared the headlines because they believed them or thought headlines were accurate.[4] However, this is not to say that participants thought the sources were attempting to share information they disagreed with. When directly asked, participants reported thinking that the

[4] We had two research assistants code responses to the open-ended question "What do you think the main motivation was for people to share the headlines? In other words, why did these social media users decide to share the headlines you saw?" Responses were coded as 1 = accuracy or belief-based or 0 = other (Cohen's kappa = 0.834). All discrepancies were resolved by the first author. The full coding scheme is available at the OSF site.

sources placed a moderate level of importance on sharing information they thought was accurate (M = 5.35, SD = 2.79; scale from 0 = not very important to 10 = very important).

Discussion

Repetition again increased belief to a similar degree whether it came from distinct sources or a single source—even though the distinct sources were made to seem independent (i.e., sharing different versions of the same information). These results are again most consistent with fluency-based theories of the illusory truth effect.

Analyses of our exploratory measures shed some insight into the disconnect between our manipulation of consensus and belief. While participants could distinguish between news headlines repeated by one person versus many, this memory did not translate into perceptions about the level of social consensus around a claim. Instead, participants thought claims were more likely to be agreed upon simply due to repetition—regardless of the actual number of unique sources (see Weaver et al., 2007 for similar findings). Thus, people seem to represent consensus around repeated claims poorly.

Interim Discussion

Across three experiments, we examined whether variations in the wording or source of repeatedly-encountered news items would moderate the effects of repetition on belief. In contrast to theories suggesting that repetition should be most influential when it reflects a consensus among varied sources, we found that none of these manipulations affected belief. These findings also go against a perceptual fluency explanation of the illusory truth effect, which predicts that repeating a statement using identical wording should increase belief more than paraphrased repetition. Instead, our results were most consistent with theories highlighting the role of low-level semantic cognitive processes like conceptual fluency in shaping judgements of truth. Repetition increases belief by making statements conceptually easier to process, regardless of the precise wording or source.

Why did information about consensus not bear on participants' judgements of truth? Recall that we outlined three processes needed for this relationship to hold. First, participants must track, in memory, whether repeated statements came from one source or many. Second, participants must interpret this variation in sources as indicating different degrees of consensus. Finally, participants must use this perceived consensus as a cue for truth. In Experiments 2 & 3, exploratory source memory data suggested the first process likely held—participants were able to track which claims were repeated by more distinct sources than others (although not particularly well). Instead, exploratory data from Experiment 3 suggests participants did not reliably engage in the second process—inferring broad agreement around a claim based on their memory for the sources.

Interestingly, the news headlines that participants saw did affect perceptions of consensus: repeated headlines were judged as more likely to be widely agreed upon than new headlines. However, it did not matter whether the repetition came from one person or many—both kinds of repetition similarly inflated perceptions of consensus. These findings mirror those of Weaver et al. (2007), who asked participants to estimate the prevalence of opinions they had seen once, three times from one source, or three times from three sources. Repetition made opinions seem more prevalent, even when only one source repeated the opinion. These results, along with ours, suggest that simply hearing information multiple times can make it seem more widely agreed-on, even if the repetition comes from only one source.

This insight also helps reconcile the present experiments with past work on how social consensus affects belief. In past work, information about consensus could readily affect judgements because it was directly presented (e.g., through summary statistics; Lewandowsky et al., 2013) or easily inferred (e.g., through sequential presentation of similar or differing views; Connor Desai et al., 2022). By contrast, in the present experiment, participants would have had to retroactively infer the level of consensus around a repeated claim based on the news items they had seen. Thus, consensus might not have affected beliefs because people relied on coarse cues like fluency, rather than effortfully tracking how many different people shared each news item.

It is possible that in other contexts, people would be better able to track social consensus from the information they encounter. For instance, people may be more attentive to sources if they are more personally salient (e.g., friends, coworkers, family) than the strangers shown in this experiment. Similarly, people might pay more attention when the source's credibility is obvious (e.g., multiple news outlets breaking a news story versus an irreputable blog repeating a story). Finally, while we examined news item repeated three times, it may be easier to track differences in consensus for statements repeated a greater number of times. In sum, it is still possible that under certain circumstances, repetition is more influential when it reflects a broad consensus.

Despite these limitations, our data provide valuable insights into how people form beliefs in real-world settings. In our digitized world, information moves quickly and reaches us repeatedly. Our results suggest that these repetitions can make information seem more true simply by making the information easier to process, even if that repetition comes from a single source. Across our three experiments, repetition of an idea increased belief—regardless of who said it or how it was said.

CHAPTER 4

General Discussion

Numerous studies have demonstrated that people are more likely to judge information as true after simply seeing it repeatedly (Dechêne et al., 2010). However, these studies have tended to focus on the effects of a single, verbatim exposure to a statement without any identifiable social context (beyond reading statements for an experiment). In real life, by contrast, statements come to us from speakers whom we may know or not know and trust or distrust. Further, we often see statements repeated many times, potentially in different wordings and from different sources. The goal of this dissertation, then, has been to examine whether and how these aspects of the social context in which people encounter statements affects belief. Eight experiments examined the impacts of source trustworthiness (Chapter 2) and cues to social consensus (Chapter 3) on the illusory truth effect. What has been learned from this work?

Theoretical Implications

Source Trustworthiness

Chapter 2 demonstrated that source trustworthiness can moderate and even eliminate the effects of repetition on belief, but only under certain conditions. Once a statement has been encountered, the trustworthiness of the source that repeats it does not seem to matter. While people are more likely to believe statements coming from trustworthy (versus untrustworthy) sources, repetition of statements by either source increases belief to a comparable degree (Experiment 1). This finding is consistent with a number of studies showing that repetition affects belief additively—not interactively—with other cues people may use to evaluate truth, such as prior knowledge (Fazio, 2020; Fazio et al., 2015), political orientations (Pennycook et al., 2018), or advice from an external source (Unkelbach & Greifeneder, 2018). Theoretically, the finding is consistent with dominant cognitive theories suggesting that repetition elicits low-level cues that are readily incorporated into judgements of truth.

In contrast to the above data, Chapter 2 also revealed that the trustworthiness of the source of a statement from which people *initially* hear a statement is likely to matter. After hearing a statement from an untrustworthy source, later repetition of that statement is less likely to increase belief than if the statement had originally been encountered by a trustworthy source (Experiment 3-4). In fact, in a situation where the source is *completely* untrustworthy, repetition of statements made by that source no longer increases belief at all (Experiment 5). These findings challenge current accounts, which generally predict uniform effects of repetition regardless of the source of the statement. Further, these results pose a novel boundary condition to the illusory truth effect, which, as reviewed in Chapter 1, is generally a robust effect.

To explain these findings, we invoked a process of source memory (see, e.g., Begg et al., 1992). Hearing a statement from an untrustworthy source may evoke a sense of doubt or skepticism (Mayo, 2015) which may be encoded into memory in association with the statement. Then, upon re-exposure to the statement, if people are able to recall that the original source was untrustworthy, they may be able to discount the effects of fluency from repetition on belief. Two converging lines of evidence support this account. First, the effects of source on belief were larger in experiments with tasks designed to increase source memory (Experiments 4 & 5 versus 3). Second, correlational evidence suggested that participants who were better able to remember which source said what statement were more likely to believe statements repeated by the trustworthy versus untrustworthy source.

Social Consensus

In contrast to Chapter 2, Chapter 3 reported results that were much more consistent with current cognitive theories of the illusory truth effect. Hearing a statement three times increased belief to a comparable degree regardless of whether or not those repetitions varied in their wording (Experiment 6), source (Experiment 7), or both (Experiment 8). These results are consistent with a pattern in which repetition makes statements seem truer simply by making their core idea easier to process (Unkelbach & Rom, 2017; Vogel et al., 2020), irrespective of who all says it or how exactly it is said.

By contrast these results run counter to predictions from evidence in the literature on social influence that highlights' people's propensity to think in line with others (e.g., Asch, 1956; Ransom et al., 2021). We proposed that one reason we failed to find such an effect is that social consensus may be more difficult to track across repeated instances of exposure to information (versus in past research where information about consensus was presented all at once). Indeed, this idea is supported by exploratory data from Experiment 8, where we directly asked participants to estimate the level of social consensus around repeated information. We found that repetition generally increased the perception that a statement was widely agreed-upon, but the exact nature of that repetition (i.e., verbatim repetition by a single source versus paraphrased repetition by multiple sources) did not matter. Thus, one reason consensus cues may not have mattered here, is because participants did not successfully track or infer these cues from across repeated instances of exposure to the statements.

Integrating Findings on Source Trustworthiness and Social Consensus

Considered together, the findings from Chapters 2 & 3 paint somewhat divergent views of how humans judge truth. Where social factors shaped the relationship between repetition and truth judgements in Chapter 2, social factors did not have such an effect in Chapter 3. In this section, I sketch out a framework that can help integrating these findings.

The starting point of this view is the assumption that low-level cognitive cues such as fluency are relatively robust determinant of truth judgements. Existing theories suggest these cues are readily accessible to our cognitive systems (e.g., Alter & Oppenheimer, 2009). Further, the interpretation and application of these cues as a signal for truth is learned implicitly starting from a very young age (Fazio & Sherry, 2020). Thus, once such a signal accrues (i.e., a statement has been encountered), it is likely to affect belief to a similar degree regardless of the social context in which it is later encountered. This is not to say that the social context will not affect *belief.* For instance, in Experiment 1, participants gave higher truth ratings to statements being made by a trustworthy versus untrustworthy source. However, this social context is not enough to moderate the effects of repetition on belief: in that same experiment, prior exposure increased belief in statements made by either source to a similar degree.

Of course, this is not to say that social factors do not affect the relationship between repetition and belief. Instead, the current experiments point to the possibility that the social context in which repeated information is *initially encountered* may be particularly influential. For instance, in Experiments 3-5, initially hearing a statement from an untrustworthy (versus trustworthy) source reduced belief when the statement was later repeated. Why, then, did social consensus cues—which were conveyed in the initial exposures to a statement—not affect belief in Experiments 6-8?

I argue that one explanation for this pattern is that the influence of social information on later belief relies on two critical processes, previewed in Chapter 3. First, people must reconstruct the relevant social information when making the judgement of the truth of a statement. This may occur through a process of recollection (e.g., explicitly remembering who said a given statement) or through inferences based on other signals (e.g., inferring that many different people repeated a statement because it seems familiar; Weaver et al., 2007) Second, people must interpret this social information in terms of its relevance to truth and the relationship between fluency and truth.

In the case of source trustworthiness, it is clear that both of these assumptions likely held. Participants in Experiments 3-5 were generally able to remember who said which statement, and knowing who said what can be readily interpreted as diagnostic of truth, as one source was regarded as much more trustworthy than the other. Cross-experimental comparisons also support the importance of each assumption. Between Experiment 3 and 4, the initial encoding task was changed to encourage better source memory, and this change was accompanied by a larger effect of initial source on belief, in line with the first assumption. Then, between Experiments 4 and 5, the difference in trustworthiness between the two sources was heightened, further enhancing the effects of initial source on belief. This may have occurred because, in Experiment 5, participants regarded the information about the sources as more diagnostic of the truth of what they were saying, in line with the second assumption.

Turning to the experiments on social consensus, however, we see that only the first of these assumptions likely held. In Experiments 7 and 8, participants were able to correctly respond that more unique people had shared the statements that were, in fact, shared by three people versus one person repeatedly (though their estimates were not incredibly accurate overall). Thus, participants held the information needed to distinguish high- and low-consensus statements in memory. However, when directly asked to make estimations about how many other people would believe the statement, participants gave comparable estimates for statements repeated by different sources versus those repeated by one source. Thus, it seems it seems that participants were not able to interpret the information they had in memory as a cue for consensus and, thereby, truth.

In sum, one way of integrating the theoretical insights from the present research is as follows. First, low-level cognitive cues like fluency, which may arise through repetition, will increase perceptions of truth and will do so in a manner invariant to other cues that may be available at the time of judgement. Second, social information about the initial context in which repeated information was encountered will be integrated into these judgements to the degree that such information is a) available in memory and b) interpreted as relevant for judgements of truth.

Practical Implications

At the outset of this dissertation, I noted that current concerns about the spread of false and misleading information provide a real-world urgency to the basic investigation of how humans judge truth. As such, I would be remiss not to mention some practical implications of the present work.

Researchers have suggested, based on experiments on the illusory truth effect, that repetition of false information may be one factor by which people come to form false beliefs about key topics (e.g., Pennycook et al., 2018). As a result, researchers have put effort into investigating potential ways to reduce these adverse effects of repetition, such as through warnings (Nadarevic & Aßfalg, 2017), or asking people to evaluate the truth of incoming claims (Brashier et al., 2020). Chapter 2 provides two insights relevant to this research agenda. First, these results suggest that people may be able to rely on source cues to naturally discount the effects of repetition (at least partially) for information originally coming from untrustworthy sources. As a result, repetition of falsehoods by *trusted* individuals or organizations may be a particularly pernicious form of repetition that may deserve particular attention to address. Second, and relatedly, these data underscore the importance of paying attention to source cues while people navigate their informational environments. This is consistent with recent intervention efforts (e.g., "digital media literacy tips"; Guess et al., 2020) designed to encourage attention to the source of information and facilitate accurate perceptions of reliable versus unreliable news sources.

Chapter 3 provides further insights into the issues surrounding repetition and false beliefs.[5] Relying on repetition as a cue for belief can be problematic because it is illogical: merely hearing a statement again does not actually affect its truth value. In his *Philosophical Investigations*, Wittgenstein (1953, as cited in Unkelbach et al., 2019) characterized the inference that repetition connotes truth as being as absurd as “buying several copies of the morning paper to ensure that the content is true.” However, to extend the metaphor, one might argue that checking *different* copies of *different* newspapers may not be so bad a basis for judging truth, as these varied sources may provide a form of *convergent validity* (see Arkes et al., 1991). Thus, even if people’s beliefs are swayed by a single, repetitive voice, perhaps they are more swayed by disparate array of independent sources. However, Experiments 6-8 suggest that people might not make such a distinction when evaluating the truth of repeated information. Repeatedly seeing a statement from a single source may influence belief just as much as repetition from many different sources. This may be problematic as, the latter type of repetition may, under certain conditions, furnish people with more accurate information.

Future Directions

[5] Of course, Experiments 6-8 did not actually use false information as stimuli. However, the illusory truth generally replicates across true and false information alike, even in the face of relevant differences, such as knowledge that contradicts a statement (Fazio et al., 2015). Further, ratings for the stimuli in these experiments were at about the midpoint, providing justification for generalizing the results at least to other similar, moderately plausible statements.

This dissertation leaves open a number of avenues for future research. The most straightforward open question is how belief is affected by other aspects of the social context in which people repeatedly encounter information. For instance, related to—yet distinct from—the concept of source trustworthiness is the concept of *source expertise*, or a source's capability to state valid information about a topic at hand (Pornpitakpan, 2004). Future work may examine whether source expertise may attenuate the effects of repetition on belief in a similar manner to source trustworthiness. Another relevant social factor is the speaker's *group identity* (Tajfel et al., 1971), such as their political orientation or nationality. These group identities can affect what people believe (Van Bavel et al., in press), but it is unclear how exactly identity may shape the effects of repeated exposure to information.[6] Thus, one open question is how repetition affects belief when a statement is repeated by a speaker who does or does not share a social identity. For these and other factors, a key question is whether the framework introduced in this chapter can account for the findings.

Another open question is how the factors investigated in the chapters of this dissertation interact with one another. Chapter 2 considered the trustworthiness of a single source, and Chapter 3 considered variability in sources irrespective the trustworthiness of each one. However, it is easy to conceive of a context in which these factors interact. For instance, one can hear a statement from three trustworthy sources, three untrustworthy sources, or some combination thereof. How might this variation affect belief? One straightforward possibility is that people may be sensitive to the average quality of the sources that repeat a statement. Thus, belief may increase linearly with the

---

[6] In one set of studies, repetition was found to increase belief to similar degree whether the information favored or disfavored one's political identity (i.e., was pro-Democratic or pro-Republican; Pennycook et al., 2018). However, it remains an open question whether similar patterns would hold if the *source* of a statement, rather in its content was aligned or unaligned with one's identity.

proportion of trustworthy sources that repeat a statement. Another possibility is that people may overweight the impact of untrustworthy (versus trustworthy) sources, drawing on a general cognitive bias to weight negative stimuli and experiences more than positive ones (Rozin & Royzman, 2001). Thus, replacing one of four trustworthy sources with an untrustworthy one may have a larger effect on belief than replacing one of four *untrustworthy* sources with a *trustworthy* one. Future research may test these alternate possibilities to further examine how different kinds of repetition affect belief.

Finally, the present studies all examined the effects of repetition on belief in a single-session study. Of course, in real life we often see information repeated with longer delays between repetitions, such as hours or weeks. The effects of repetition on belief generally grow weaker as longer delays separates exposures (Henderson, Simons, et al., 2021), likely because the underlying cognitive cues like fluency or familiarity fade over time. However, recollection for details about how the information was encountered may fade even more quickly (Silva et al., 2017), taking with it memory for details of the social context of the statement. Thus, future work may reexamine whether the present effects hold at longer delays, or whether the interplay of these two kinds of forgetting produce different patterns.

Conclusion

Every day, we are faced with a constant stream of claims about the world, not all of them true. How do we decide what is fact and what is fiction? The present dissertation shows that these judgements of truth are persistently affected by low-level cognitive inputs (e.g., how easy it is process a statement) that can, under some circumstances, be moderated by higher-level social information. Repetition makes statements easier to comprehend, and people interpret this ease of processing as a cue to a statements' truth. While this illusory truth effect occurs just the same whether one source repeats a claim or many, this effect wanes when an untrustworthy source repeats the statement.

REFERENCES

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

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020a). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*, 388–407.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020b). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407.

Arkes, H. R., Boehm, L. E., & Xu, G. (1991). Determinants of judged validity. *Journal of Experimental Social Psychology*, *27*(6), 576–605. https://doi.org/10.1016/0022-1031(91)90026-3

Arkes, H. R., Hackett, C., & Boehm, L. (1989). The generality of the relation between familiarity and judged validity. *Journal of Behavioral Decision Making*, *2*(2), 81–94. https://doi.org/10.1002/bdm.3960020203

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70. https://doi.org/10.1037/h0093718

Aslett, K., Sanderson, Z., Godel, W., Persily, N., Nagler, J., & Tucker, J. A. (2023). Online searches to evaluate misinformation can increase its perceived veracity. *Nature*. https://doi.org/10.1038/s41586-023-06883-y

Bailey, G. (Director). (2005, April 7). Buster the Myth Maker Arthur (Season 9, Episode 9b) [TV series episode]. In *Arthur*. PBS.

Bak-Coleman, J. B., Alfano, M., Barfuss, W., Bergstrom, C. T., Centeno, M. A., Couzin, I. D., Donges, J. F., Galesic, M., Gersick, A. S., Jacquet, J., Kao, A. B., Moran, R. E., Romanczuk, P., Rubenstein, D. I., Tombak, K. J., Van Bavel, J. J., & Weber, E. U. (2021). Stewardship of global collective behavior. *Proceedings of the National Academy of Sciences*, *118*(27), e2025764118. https://doi.org/10.1073/pnas.2025764118

Begg, I. M., Anas, A., & Farinacci, S. (1992). *Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth*. *121*(4), 446–458. https://doi.org/10.1037/0096-3445.121.4.446

Brashier, N. M., Eliseev, E. D., & Marsh, E. J. (2020). An initial accuracy focus prevents illusory truth. *Cognition*, *194*, 104054. https://doi.org/10.1016/j.cognition.2019.104054

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

Brashier, N. M., Umanath, S., Cabeza, R., & Marsh, E. J. (2017). Competing cues: Older adults rely on knowledge in the face of fluency. *Psychology and Aging*, *32*(4), 331–337. https://doi.org/10.1037/pag0000156

Brown, A. S., & Nix, L. A. (1996). Turning lies into truths: Referential validation of falsehoods. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1088–1100. https://doi.org/10.1037/0278-7393.22.5.1088

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116.

Coelho, G. L. D. H., Hanel, P. H. P., & Wolf, L. J. (2020). The very efficient assessment of need for cognition: Developing a six-item version. *Assessment*, *27*(8), 1870–1885. https://doi.org/10.1177/1073191118793208

Connor Desai, S., Xie, B., & Hayes, B. K. (2022). Getting to the source of the illusion of consensus. *Cognition*, *223*, 105023. https://doi.org/10.1016/j.cognition.2022.105023

Connor, P., Varney, J., Keltner, D., & Chen, S. (2021). Social class competence stereotypes are amplified by socially signaled economic inequality. *Personality and Social Psychology Bulletin*, *47*(1), 89–105.

Cummings, K. M., Brown, A., & O'Connor, R. (2007). The cigarette controversy. *Cancer Epidemiology, Biomarkers & Prevention*, *16*(6), 1070–1076. https://doi.org/10.1158/1055-9965.EPI-06-0912

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2009). Mix me a list: Context moderates the truth effect and the mere-exposure effect. *Journal of Experimental Social Psychology*, *45*(5), 1117–1122. https://doi.org/10.1016/j.jesp.2009.06.019

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

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

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

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*(5), 993–1002. https://doi.org/10.1037/xge0000098

Fazio, L. K., Rand, D. G., & Pennycook, G. (2019). Repetition increases perceived truth equally for plausible and implausible statements. *Psychonomic Bulletin & Review*, *26*(5), 1705–1710. https://doi.org/10.3758/s13423-019-01651-4

Fazio, L. K., & Sherry, C. L. (2020). The effect of repetition on truth judgments across development. *Psychological Science*, *31*(9), 1150–1160. https://doi.org/10.1177/0956797620939534

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117–140.

Garcia-Marques, T., Silva, R. R., Reber, R., & Unkelbach, C. (2015). Hearing a statement now and believing the opposite later. *Journal of Experimental Social Psychology*, *56*, 126–129. https://doi.org/10.1016/j.jesp.2014.09.015

Garrett, R. K. (2011). Troubling consequences of online political rumoring. *Human Communication Research*, *37*(2), 255–274. https://doi.org/10.1111/j.1468-2958.2010.01401.x

Ghetti, S., & Angelini, L. (2008). The development of recollection and familiarity in childhood and adolescence: Evidence from the dual-process signal detection model. *Child Development*, *79*(2), 339–358. https://doi.org/10.1111/j.1467-8624.2007.01129.x

Greifeneder, R., & Unkelbach, C. (2013). Experiencing thinking. In C. Unkelbach & R. Greifeneder, *The experience of thinking: How the fluency of mental processes influences cognition and behaviour* (pp. 11–18). Psychology Press.

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences*, *117*(27), 15536–15545. https://doi.org/10.1073/pnas.1920498117

Hartman, R., Moss, A. J., Jaffe, S. N., Rosenzweig, C., Litman, L., & Robinson, J. (2023). *Introducing Connect by CloudResearch: Advancing Online Participant Recruitment in the Digital Age* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/ksgyr

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

Hawkins, S. A., & Hoch, S. J. (1992). Low-involvement learning: Memory without evaluation. *Journal of Consumer Research*, *19*(2), 212–225. https://doi.org/10.1086/209297

Held, L., & Ott, M. (2018). On *p* -values and bayes factors. *Annual Review of Statistics and Its Application*, *5*(1), 393–419. https://doi.org/10.1146/annurev-statistics-031017-100307

Henderson, E. L., Simons, D. J., & Barr, D. J. (2021). The trajectory of truth: A longitudinal study of the illusory truth effect. *Journal of Cognition*, *4*(1), 1–23. https://doi.org/10.5334/joc.161

Henderson, E. L., Westwood, S. J., & Simons, D. J. (2021). A reproducible systematic map of research on the illusory truth effect. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-021-01995-w

Ishihara, S. (1917). *Tests for color blindness*.

Jacoby, L. L., Kelley, C., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology*, *56*(3), 326–338.

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

Jacoby, L. L., Woloshyn, V., & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General*, *118*(2), 115–125.

Jalbert, M., Newman, E., & Schwarz, N. (2020). Only half of what I'll tell you is true: Expecting to encounter falsehoods reduces illusory truth. *Journal of Applied Research in Memory and Cognition*, *9*(4), 602–613. https://doi.org/10.1016/j.jarmac.2020.08.010

Kassambara, A. (2020). *rstatix: Pipe-friendly framework for basic statistical tests* (0.7.0) [Computer software].

Kumar, S., West, R., & Leskovec, J. (2016). Disinformation on the web: Impact, characteristics, and detection of Wikipedia hoaxes. *Proceedings of the 25th International Conference on World Wide Web*, 591–602. https://doi.org/10.1145/2872427.2883085

Lacassagne, D., Béna, J., & Corneille, O. (2022). Is Earth a perfect square? Repetition increases the perceived truth of highly implausible statements. *Cognition*, *223*, 105052. https://doi.org/10.1016/j.cognition.2022.105052

Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science*, *4*(1), 251524592095150. https://doi.org/10.1177/2515245920951503

Lewandowsky, S., Cook, J., Fay, N., & Gignac, G. E. (2019). Science by social media: Attitudes towards climate change are mediated by perceived social consensus. *Memory & Cognition*, *47*(8), 1445–1456. https://doi.org/10.3758/s13421-019-00948-y

Lewandowsky, S., Gignac, G. E., & Vaughan, S. (2013). The pivotal role of perceived scientific consensus in acceptance of science. *Nature Climate Change*, *3*(4), 399–404. https://doi.org/10.1038/nclimate1720

Litman, L., Robinson, J., & Abberbock, T. (2017a). TurkPrime. Com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442.

Litman, L., Robinson, J., & Abberbock, T. (2017b). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442.

Lyons, B. A. (2023). Older Americans are more vulnerable to prior exposure effects in news evaluation. *Harvard Kennedy School Misinformation Review*, *4*(4). https://doi.org/10.37016/mr-2020-118

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252–271.

Mayo, R. (2015). Cognition is a matter of trust: Distrust tunes cognitive processes. *European Review of Social Psychology*, *26*(1), 283–327. https://doi.org/10.1080/10463283.2015.1117249

McGlone, M. S., & Tofighbakhsh, J. (2000). Birds of a Feather Flock Conjointly (?): Rhyme as Reason in Aphorisms. *Psychological Science*, *11*(5), 424–428. https://doi.org/10.1111/1467-9280.00282

Morey, R. D., Rouder, J. N., Jamil, T., & Morey, M. R. D. (2015). *Package 'bayesfactor'* (0.9.12) [Computer software].

Myers, S. J., Rhodes, M. G., & Hausman, H. E. (2020). Judgments of learning (JOLs) selectively improve memory depending on the type of test. *Memory & Cognition*, *48*(5), 745–758. https://doi.org/10.3758/s13421-020-01025-5

Nadarevic, L., & Aßfalg, A. (2017). Unveiling the truth: Warnings reduce the repetition-based truth effect. *Psychological Research*, *81*(4), 814–826. https://doi.org/10.1007/s00426-016-0777-y

Nadarevic, L., & Erdfelder, E. (2014). Initial judgment task and delay of the final validity-rating task moderate the truth effect. *Consciousness and Cognition*, *23*, 74–84. https://doi.org/10.1016/j.concog.2013.12.002

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

Parkinson, R. G. (2016, November 25). Opinion | Fake news? That's a very old story. *Washington Post*. https://www.washingtonpost.com/opinions/fake-news-thats-a-very-old-story/2016/11/25/c8b1f3d4-b330-11e6-8616-52b15787add0_story.html

Peer, E., David, R., Andrew, G., Zak, E., & Ekaterina, D. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01694-3

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

Pillai, R. M., Yang, S., Jiang, Q., & Fazio, L. K. (under review). *Repetition increases belief more for trivia than for news headlines*.

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

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ransom, K. J., Perfors, A., & Stephens, R. (2021). Social meta-inference and the evidentiary value of consensus. In T. Fitch, C. Lamm, H. Leder, & K. Tessmar (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, Issue 43, pp. 833–839). Cognitive Science Society.

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

Reber, R., & Unkelbach, C. (2010). The epistemic status of processing fluency as source for judgments of truth. *Review of Philosophy and Psychology*, *1*(4), 563–581. https://doi.org/10.1007/s13164-010-0039-7

Rivers, M. L., Janes, J. L., & Dunlosky, J. (2021). Investigating memory reactivity with a within-participant manipulation of judgments of learning: Support for the cue-strengthening hypothesis. *Memory*, *29*(10), 1342–1353. https://doi.org/10.1080/09658211.2021.1985143

Roggeveen, A. L., & Johar, G. V. (2002). Perceived source variability versus familiarity: Testing competing explanations for the truth effect. *Journal of Consumer Psychology*, *12*(2), 81–91. https://doi.org/10.1207/S15327663JCP1202_02

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320.

Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, *14*(4), 332–348.

Silva, R. R., Garcia-Marques, T., & Reber, R. (2017). The informative value of type of repetition: Perceptual and conceptual fluency influences on judgments of truth. *Consciousness and Cognition*, *51*, 53–67. https://doi.org/10.1016/j.concog.2017.02.016

Simmonds, B. P., Stephens, R., Searston, R. A., Asad, N., & Ransom, K. J. (2023). The Influence of Cues to Consensus Quantity and Quality on Belief in Health Claims. In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, Issue 45).

Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 553–558. https://doi.org/10.1037/a0038388

Soll, J. (2016, December 18). The Long and Brutal History of Fake News. *POLITICO Magazine*. https://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535/

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178. https://doi.org/10.1002/ejsp.2420010202

Tauber, S. K., Dunlosky, J., Rawson, K. A., Rhodes, M. G., & Sitzman, D. M. (2013). General knowledge norms: Updated and expanded from the Nelson and Narens (1980) norms.

Behavior Research Methods, *45*(4), 1115–1143. https://doi.org/10.3758/s13428-012-0307-9

Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 219–230. https://doi.org/10.1037/0278-7393.33.1.219

Unkelbach, C., & Greifeneder, R. (2013). A general model of fluency effects in judgment and decision making. In C. Unkelbach & R. Greifeneder, *The experience of thinking: How the fluency of mental processes influences cognition and behaviour* (pp. 11–32). Psychology Press.

Unkelbach, C., & Greifeneder, R. (2018). Experiential fluency and declarative advice jointly inform judgments of truth. *Journal of Experimental Social Psychology*, *79*, 78–86. https://doi.org/10.1016/j.jesp.2018.06.010

Unkelbach, C., Koch, A., Silva, R. R., & Garcia-Marques, T. (2019). Truth by repetition: Explanations and implications. *Current Directions in Psychological Science*, *28*(3), 247–253. https://doi.org/10.1177/0963721419827854

Unkelbach, C., & Rom, S. C. (2017). A referential theory of the repetition-induced truth effect. *Cognition*, *160*, 110–126. https://doi.org/10.1016/j.cognition.2016.12.016

Unkelbach, C., & Stahl, C. (2009). A multinomial modeling approach to dissociate different components of the truth effect. *Consciousness and Cognition*, *18*(1), 22–38. https://doi.org/10.1016/j.concog.2008.09.006

Van Bavel, J. J., Rathje, S., Vlasceanu, M., & Pretus, C. (in press). Updating the identity-based model of belief: From false belief to the spread of misinformation. *Current Opinion in Psychology*.

Vogel, T., Silva, R. R., Thomas, A., & Wänke, M. (2020). Truth is in the mind, but beauty is in the eye: Fluency effects are moderated by a match between fluency source and judgment dimension. *Journal of Experimental Psychology: General*, *149*(8), 1587–1596. https://doi.org/10.1037/xge0000731

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Weaver, K., Garcia, S. M., Schwarz, N., & Miller, D. T. (2007). Inferring the popularity of an opinion from its familiarity: A repetitive voice can sound like a chorus. *Journal of Personality and Social Psychology*, *92*(5), 821–833. https://doi.org/10.1037/0022-3514.92.5.821

Whalen, A., Griffiths, T. L., & Buchsbaum, D. (2018). Sensitivity to shared information in social learning. *Cognitive Science*, *42*(1), 168–187. https://doi.org/10.1111/cogs.12485

Winkielman, P., & Schwarz, N. (2001). How pleasant was your childhood? Beliefs about memory shape inferences from experienced difficulty of recall. *Psychological Science*, *12*(2), 176–179. https://doi.org/10.1111/1467-9280.00330

Wittgenstein, L. (1953). *Philosophical investigations*. *Philosophische Untersuchungen.* (pp. x, 232). Macmillan.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Yousif, S. R., Aboody, R., & Keil, F. C. (2019). The illusion of consensus: A failure to distinguish between true and false consensus. *Psychological Science*, *30*(8), 1195–1204. https://doi.org/10.1177/0956797619856844

APPENDIX A

Chapter 2 Supplementary Analyses

This Appendix contains supplementary results for the experiments reported in Chapter 2.

Pilot

Method

Participants. Sixty participants with full-color vision were recruited from Amazon's Mechanical Turk platform (MTurk) using CloudResearch (Litman et al., 2017), in the same manner as described in Experiment 1 in the main text.

Design. This experiment employed a mixed design manipulating source (Mr. Green, Mr. Red) and statement truth (true, false) within subjects and the presence of training (training, no training) between subjects.

Materials. The key stimuli were 40 true and false trivia statements that were likely unknown to participants, adapted from Fazio (2020). These items are a subset of the 42 items described in Experiment 1 of the main text.

In addition, during the training phase, we used 40 general knowledge trivia statements about which participants likely held relevant prior knowledge, also adapted from Fazio (2020). Similar to the key stimuli, these items are a subset of the 42 items used for the training phase in Experiment 1 of the main text.

Procedure. This experiment was administered online via gorilla.sc (Anwyl-Irvine et al., 2020a). Participants began the experiment by reading the study information sheet, completing two attention checks, and responding to five Ishihara plate test trials to check for possible colorblindness. Then, the study proceeded in three phases: training, test, and follow-up.

Training Phase. The training phase was identical to that of the training phase in Experiment 1 of the main text. Only half of the participants were assigned to complete this phase.

Test Phase. After the training phase, participants proceeded directly to the test phase. In this phase, participants were shown the key stimuli: 40 true or false, obscure general trivia statements. These statements were attributed to one of the two sources (Mr. Red or Mr. Green), as depicted in Figure 2.1 of the main text. In this phase, Mr. Red and Mr. Green conveyed equal numbers of true and false statements, and we counterbalanced which statements were said by which source by randomly assigning participants to one of two sources. Participants rated the truthfulness of each statement on a 6-point scale from "Definitely True" to "Definitely False" without any feedback.

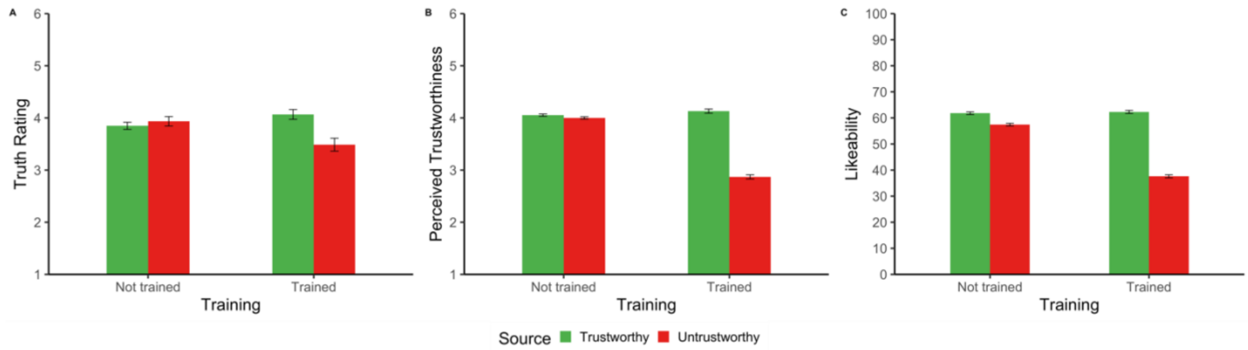
Follow-up. Finally, participants completed the five follow-up questions described in the main text of Experiment 1, asking them how much they trusted and liked each source. Finally, participants were thanked for their time and informed about the purpose of the study.

Results

All statistical tests reported are conducted at the .05 alpha level. This pilot study was not pre-reregistered. Figure S1 depicts the results across three of the main dependent variables.

Figure S1

Results from Pilot Study Testing Training Manipulation



Note. Figure is a reproduction of Figure 2.2. Figure shows results from the following measures: A) truth ratings (1 = definitely false; 6 = definitely true) for new statements attributed to Mr. Green or Mr. Red B) perceived trustworthiness (1 = very untrustworthy, 6 = very trustworthy) for Mr. Green and Mr. Red and C) likeability ratings (0 = don't like, 100 = like very much) for Mr. Green and Mr. Red.

Truth Ratings. We first examined participants' ratings for the new statements in the test phase. As shown panel A of Figure S1, participants who received training believed statements from Mr. Green more than those from Mr. Red, a difference that was not apparent for participants who did not receive training.

To analyze these data statistically, we conducted a 2 (truth: true, false) × 2 (source: Mr. Green, Mr. Red) × 2 (training: present, absent) ANOVA on participants' mean truth ratings. We observe a main effect of truth such that participants gave higher ratings to true (M = 4.09) than false statements (M = 3.59), $F(1,58) = 16.95$, $p < .001$, $\eta_p^2 = .23$. We also observe a main effect of source such that statements from Mr. Green (M = 3.96) received higher ratings than statements from Mr. Red (M = 3.71), $F(1,58) = 6.09$, $p = .017$, $\eta_p^2 = .10$. Critically, this main effect of source was qualified by a significant interaction with the presence of training, $F(1,58) = 11.11$, $p < .001$, $\eta_p^2 = 0.16$. Follow-up t-tests revealed that participants who received training gave higher truth ratings to statements from Mr. Green (M = 4.07) than Mr. Red (M = 3.49), $t(22) = 3.24$, $p = .004$, *d* = 0.68. By contrast, there was no such difference for participants who did not receive this training (M_{Green} = 3.85, M_{Red} = 3.94), *t*(36) = -0.77, *p* = .444, *d* = 0.13. Thus, the main effect of statement source is attributable to differences among the participants who received training.

Follow-Up Questions. Next, we analyzed participants responses to our 3 sets of follow-up questions. First, participants were asked "In the first set of statements you rated, which of the following was most accurate regarding the background colors?" (*Statements from Mr. Red were much more likely to be true, Statements from Mr. Red were somewhat more likely to be true, Statements from Mr. Red or Mr. Green were equally likely to be true, Statements from Mr. Green were somewhat more likely to be true, Statements from Mr. Green were much more likely to be true,* coded from 1 to 5). To analyze these data, we conducted a *t*-test comparing responses on this item from participants who did versus did not receive training. While participants who did not receive training were at the midpoint on this measure (*M* = 3.00), participants who received training gave higher ratings (*M* = 3.78), *t*(48.11) = 3.42, *p* = .001, *d* = 0.90, indicating a greater relative trust in Mr. Green over Mr. Red.

Next, participants were asked about the trustworthiness of each source across two questions ("How trustworthy is [Mr. Red/Mr. Green]?": (*Very Untrustworthy, Untrustworthy, Slightly Untrustworthy, Slightly Trustworthy, Trustworthy, Very Trustworthy*, coded from 1 to 6). Results are shown in Figure S1, panel B. A 2 (source: Mr. Green, Mr. Red) × 2 (training: present, absent) ANOVA revealed significant main effects of source (M_{Green} = 4.09, M_{Red} = 3.44), *F*(1,58) = 20.82, $p < .001$, $\eta_p^2 = .26$, and training ($M_{trained}$ = 3.50, $M_{not\ trained}$ = 4.25), *F*(1,58) = 10.17, $p < .001$, $\eta_p^2 = .15$. Critically, both of these effects were qualified by an interaction between training and source, *F*(1,58) = 17.54, $p < .001$, $\eta_p^2 = .23$. Follow-up *t*-tests revealed that participants who received training trusted Mr. Green (*M* = 4.13) more than Mr. Red (*M* = 2.87), *t*(22) = 5.14, $p < .001$, *d* =

1.07, but that participants who did not receive training showed no such effect (M_{Green} = 4.05, M_{Red} = 4.00), t(36) = 0.32, p = .751, d = 0.05.

Finally, we asked participants about the likeability of each source ("How much do think you would like [Mr. Red/Mr. Green]?": numeric slider entry from *0 = don't like,* to *100 = like very much*, with default of 50). Results are shown in Figure S1, panel C. A 2 (source: Mr. Green, Mr. Red) × 2 (training: present, absent) ANOVA revealed significant main effects of source (M_{Green} = 62.1, M_{Red} = 47.5), F(1,58) = 20.36, p < .001, η_p^2 = .26, and training ($M_{trained}$ = 50.0, $M_{not\ trained}$ = 59.6), F(1,58) = 6.94, p = .011, η_p^2 = .11. Critically, both of these effects were qualified by an interaction between training and source, F(1,58) = 9.94, p = .003, η_p^2 = .15. Follow-up t-tests revealed that participants who received training liked Mr. Green (M = 62.3) more than Mr. Red (M = 37.6), t(22) = 4.33, p < .001, d = 0.90, but that participants who did not receive training showed no such effect (M_{Green} = 61.8, M_{Red} = 57.4), t(36) = 1.22, p = .238, d = 0.19.

Experiment 2

Supplementary Results

Table S1 below reports the full results of the regression model predicting participant-level difference scores in truth ratings for statements repeated by the trustworthy minus untrustworthy source. Predictors include a) participants' difference scores in perceived trust for the trustworthy minus untrustworthy source and b) participants' source memory ability, as indexed by the proportion of old items whose source was correctly classified.

Table S1

Regression Predicting Participants' Relative Belief in Statements from the Trustworthy versus Untrustworthy Sources (Experiment 2)

Value	Estimate	SE	*t* value	*p* value

Intercept	-0.12	0.08	-1.59	.114
Relative perceived trust	0.03	0.02	1.30	.195
Source memory	**0.30**	**0.13**	**2.22**	**.028**

Note: Regression was fit to 274 observations, each corresponding to an individual participant. 7 participants were excluded from these analyses due to missing source memory scores (i.e., participants who did not correctly classify any items as old, leaving a 0 in the denominator for their source memory score). Bold values indicate significant effects.

Experiment 3

Supplementary Results

Table S2 below reports the full results of the regression model predicting participant-level difference scores in truth ratings for statements repeated by the trustworthy minus untrustworthy source. Predictors are the same as those in the model reported in Table S1 above.

Table S2

Regression Predicting Participants' Relative Belief in Statements from the Trustworthy versus Untrustworthy Sources (Experiment 3)

Value	Estimate	SE	*t* value	*p* value
Intercept	-0.05	0.06	-0.83	.409
Relative perceived trust	0.00	0.01	0.29	.770
Source memory	0.16	0.10	1.59	.113

Note: Regression was fit to 491 observations, each corresponding to an individual participant. 7 participants were excluded from these analyses due to missing source memory scores (i.e., participants who did not correctly classify any items as old, leaving a 0 in the denominator for their source memory score).

Experiment 4

Supplementary Results

Source Memory. Here we report full statistics for our comparisons of source memory measures in Experiment 4 versus 2 and 3. First, discriminability (*d'*) was higher in Experiment 4 (M = 0.90) than in Experiment 2 (M = 0.76), $t(592.89) = 2.11$, $p = .035$, $d = 0.16$ and Experiment 3 (M = 0.69), $t(982.70) = 3.91$, $p < .001$, $d = 0.25$. Second, source memory (i.e., the proportion of old items for which participant correctly classified the source) was higher in Experiment 4 (M = 60.56%) than in Experiment 2 (M = 52.25%), $t(590.16) = 5.24$, $p < .001$, $d = 0.39$ and Experiment 3 (M = 54.23%), $t(973.82) = 4.82$, $p < .001$, $d = 0.31$.

Individual Differences in Truth Ratings. Table S3 below reports the full results of the regression model predicting participant-level difference scores in truth ratings for statements repeated by the trustworthy minus untrustworthy source. Predictors are the same as those in the model reported in Tables S1 & S2 above, and also includes participants' scores on the Need for Cognition scale.

Table S3

Regression Predicting Participants' Relative Belief in Statements from the Trustworthy versus Untrustworthy Sources (Experiment 4)

Value	Estimate	SE	*t* value	*p* value
Intercept	-0.18	0.10	-1.77	.077
Relative perceived trust	**0.05**	**0.01**	**3.62**	**<.001**
Source memory	**0.20**	**0.10**	**1.97**	**.049**
Need for Cognition	0.00	0.00	0.91	.361

Note: Regression was fit to 496 observations, corresponding to each of the participants. Bold values indicate significant effects.

Experiment 5

Supplementary Results

Source Memory. Here we report full statistics for our comparisons of source memory measures in Experiment 5 versus Experiments 2-4. First, discriminability (*d'*) was higher in Experiment 5 (M = 1.15) than in Experiment 2 (M = 0.76), $t(627.40) = 5.77$, $p < .001$, $d = 0.42$, Experiment 3 (M = 0.69), $t(962.48) = 8.21$, $p < .001$, $d = 0.52$, and Experiment 4 (M = 0.90), $t(982.77) = 4.27$, $p < .001$, $d = 0.27$. Second, source memory (i.e., the proportion of old items for which participant correctly classified the source) was higher in Experiment 5 (M = 63.66%) than in Experiment 2 (M = 52.25%), $t(598.49) = 7.15$, $p < .001$, $d = 0.53$, Experiment 3 (M = 54.23%), $t(967.46) = 7.11$, $p < .001$, $d = 0.45$, and Experiment 4 (M = 60.56%), $t(987.61) = 2.22$, $p = .026$, $d = 0.14$.

Individual Differences in Truth Ratings. Table S4 below reports the full results of the regression model predicting participant-level difference scores in truth ratings for statements repeated by the trustworthy minus untrustworthy source. Predictors are the same as those in the model reported in Table S3.

Table S4

Regression Predicting Participants' Relative Belief in Statements from the Trustworthy versus Untrustworthy Sources (Experiment 5)

Value	Estimate	SE	*t* value	*p* value
Intercept	-0.36	0.20	-1.80	.073
Relative perceived trust	**0.22**	**0.02**	**10.75**	**< .001**
Source memory	**0.95**	**0.18**	**5.16**	**< .001**
Need for Cognition	-0.01	0.01	-1.59	.112

Note: Regression was fit to 494 observations, corresponding to each of the participants. Bold values indicate significant effects.