Correcting Misinformation—A Challenge for Education and Cognitive Science

Chapter for "Processing Inaccurate Information: Theoretical and Applied Perspectives from Cognitive Science and the Educational Sciences"

Ullrich K. H. Ecker, Briony Swire, and Stephan Lewandowsky

School of Psychology

University of Western Australia

"Words have weight; something once said cannot be unsaid. Meaning is like a stone dropped into a pool; the ripples will spread and you cannot know what back they wash against."

(Philippa Gregory)

Jenny McCarthy is a popular and successful U.S. TV show host, actress, and author. She is also the mother of a child diagnosed with autism, which has inspired her to become an activist and serve on the board-of-directors of *Generation Rescue*, an organisation dedicated to informing the public about the recovery of children with autism spectrum disorders, and the presumed risks of vaccinations.

In March 2012, she wrote: "MMR [i.e., the common measles-mumps-rubella vaccine], by far, has been the vaccine most commonly cited by parents as a trigger for a regression into autism" (McCarthy, 2012). This argument concerning the cause of autism originates from a 1998 publication in the medical journal *The Lancet*, which suggested a relationship between the MMR vaccine and the onset of autism (Wakefield et al., 1998).

Celebrities commonly use their popularity to rally for worthy causes. Angelina Jolie raises awareness of humanitarian crises and refugee issues, Oprah Winfrey promotes education for disadvantaged girls in Africa and the U.S., Salma Hayek condemns violence against women and promotes UNICEF vaccination programs. So why did we choose McCarthy's example to open a chapter on misinformation? What's wrong with McCarthy's activism?

As it turns out, quite a few things. First and foremost, there is a strong consensus in the medical science community that there is no causal link between the MMR vaccine (or any other vaccine) and autism. This consensus is based on exhaustive and widely published global research efforts to replicate the Wakefield et al. (1998) findings. For example, a retrospective cohort study by Madsen et al. (2002) reviewed all children born in Denmark between January 1991 and December 1998, and concluded that there was no increase in the

incidence of autism for vaccinated as opposed to unvaccinated children. In a review paper, DeStefano and Thompson (2004) concluded that "evidence now is convincing that the measles-mumps-rubella vaccine does not cause autism or any particular subtypes of autistic spectrum disorder." (DeStefano & Thompson, 2004, p. 19). In fact, it has become clear that the Wakefield et al. (1998) study was a fraud. *The Lancet* officially retracted the article in 2010 (Editors of the Lancet, 2010), and the UK's General Medical Council found the lead author guilty of professional misconduct and removed him from the medical register. Hence, McCarthy's public claims represent misinformation.

The second issue regards the use of one's celebrity status to comment on scientific questions without being an expert on those issues. This is relevant because the public is demonstrably receptive to non-expert opinions. For example, roughly a quarter of survey respondents in the U.S. placed 'some' (24 %) or even 'a lot' (2 %) of trust in the opinions of celebrities on vaccine safety (Freed, Clark, Butchart, Singer, & Davis, 2011). In the present context, this affirms the likelihood that McCarthy's claims will affect people despite the plethora of evidence that shows those claims to be wrong.

The fact that people receive much of their information from potentially unreliable sources such as celebrities, popular "infotainment" TV shows and non-expert websites poses a problem in particular because it is known that misinformation continues to exert an influence on people's opinions even after it has been retracted. To illustrate, in 2002, when it had become clear and widely publicised that the Wakefield et al. (1998) findings were invalid, 24% of mothers in the UK still erroneously considered the vaccine a greater risk than the disease it was preventing (Smith & Yarwood, 2007). Even after a decade of rectification—efforts by doctors, scientists, health agencies, and some media outlets to correct the misconceptions—as well as events such as the vaccine-preventable mumps epidemic in the UK in 2005, rates of immunisation in the UK still had not returned to the

level of 1996-1997 (Health Protection Agency, 2011). Clearly, erroneous information can have adverse consequences for public health and society at large, despite subsequent corrections and widespread efforts to disseminate the correct information.

In this chapter, we outline how retracted misinformation still influences people's reasoning, and why misinformation can persist in society, being surprisingly difficult to eradicate. We discuss how attempts to correct misinformation can under some circumstances even *worsen* the problem, but we will also explore ways in which misinformation can be useful for educational purposes. We use the term 'misinformation' for any information that is (or might reasonably be) believed to be valid when first acquired, but which is subsequently shown to be incorrect or superseded by updated information.

Misinformation effects and individual-level cognition

In contemporary information-driven societies we are confronted with a myriad of information sources, and it can be difficult to distinguish fact from fiction. Considering how much information we process and how quickly the world can change, people actually do a remarkable job at updating their knowledge and mental models of the world—hardly anyone would think George W. Bush was still U.S. President, and we can generally keep track of the things we have done versus those that are still on our to-do list. However, human memory is faced with the conundrum of maintaining stable memory representations (which is the whole point of having a memory in the first place) while also allowing for flexible modulation of memory representations to keep up-to-date with reality. Memory has evolved to achieve both of these aims, and hence it does not work like a blackboard: Out-dated things are rarely actually wiped out and over-written; instead, they tend to linger in the background, and access to them is only gradually lost (De Beni & Palladino, 2004; Oberauer & Vockenberg, 2009).

Retractions and the provision of alternatives—The continued influence effect

The degree to which out-dated information lingers depends on the likelihood that it will become relevant again in the future. Hence, when people are provided with a plausible alternative to replace the original misinformation, memory updating is usually successful and people no longer rely on the initial, later retracted, information (H. Johnson & Seifert, 1994; Ecker, Lewandowsky, & Tang, 2010). Once you understand that your parents sneakily exchanged the milk teeth under your pillow for dollar coins, you no longer have a reason to believe in the tooth fairy.

However, if the valid alternative is unknown, or difficult to understand, misinformation will prevail. In the case of autism, the retraction of Wakefield et al.'s (1998) claims might have been more successful had the actual causes of autism been better understood and offered to the general public as an alternative account. A retraction creates a gap in a person's mental model of the world, and without a plausible alternative account, people may continue to make use of retracted misinformation, preferring an out-dated, possibly incorrect model of the world to an *incomplete* model of the world. Hence, plain retractions—simply stating that something is incorrect—are usually rather ineffective in correcting misinformation (Ecker, Lewandowsky, Swire, & Chang, 2011a; Ecker et al., 2010; H. Johnson & Seifert, 1994; van Oostendorp & Bonebakker, 1999; cf. also Rapp & Kendeou, 2007, 2009). People's ongoing reliance on corrected misinformation is known in the literature as the 'continued influence effect' (H. Johnson & Seifert, 1994; see Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012, for a review).

Classic laboratory studies by Wilkes and Leatherbarrow (1988) and H. Johnson and Seifert (1994) have demonstrated that people rely on information after it is no longer relevant (see also Ecker, Lewandowsky, & Apai, 2011b; Ecker et al., 2010). In these studies, participants were given a transcript of a fictitious event, such as a warehouse fire. As the

story is being told, negligent storage of volatile materials (oil paint and pressurised gas cylinders) is originally imputed as a likely cause of the fire. Similar to a news story changing as new information comes to light, some participants were then told that actually no such materials had been found. When participants were asked about the presence of corrections in the report, they acknowledged the absence of oil paint and gas cylinders. However, when the same participants were making inferences to more indirect questions such as "What was the cause of the explosions?" people seemed to fall back on the only explanation they knew, namely the pressurized gas cylinders. Explicit knowledge of a correction is therefore no barrier to people using the out-dated information during their reasoning. In these studies, reliance on out-dated misinformation was eliminated only if people were given an alternative account of the fire (i.e., evidence for arson, such as petrol-soaked rags, were found at the scene). Access to this alternative explanation enabled people to (largely) discount the volatile materials as the cause of the fire.

However, providing an alternative account is not always a solution to the problem. Specifically, providing an alternative may be helpful in simple scenarios with a straightforward cause-and-effect chain (e.g., arson, not negligence, caused the fire), but it may be potentially less useful in scenarios with many alternatives and complex interactions. In particular, if the alternative account is more complicated than the original misinformation-based account, the alternative may not be fully embraced. People generally prefer simple explanations to those which require more cognitive effort to understand and remember (Lombrozo, 2007). For example, Heit and Rotello (2012) reported that valid explanations are considered stronger when they are short, whereas, ironically, illogical explanations are considered more convincing when they are long.Those results mesh well with Chater's (1999) proposal that finding the simplest explanatory pattern from complex data is a fundamental principle of cognition that guides perception, learning, memory, and reasoning.

Chater argued that simplicity was not only a crucial factor in describing people's behavior across a range of cognitive domains, but that the application of a simplicity principle was also normatively justified (meaning essentially that all other variables being equal, simple explanations are more likely to be correct and therefore helpful in making predictions and decisions). On that basis, a simple mental model of an event is not only easy to understand, efficiently stored and remembered, but it is also likely to be adequate. Hence, plausible and simple explanations will be more readily accepted, remembered, and utilised than a correct but over-complicated account.

To summarize, misinformation effects are usually not an issue when there is a simple and plausible alternative to fill the gap created by a retraction in a person's mental model of an event or links between concepts. When there is no simple alternative available, however, people often continue to rely on retracted misinformation. To better understand this reliance on out-dated or incorrect information, it is useful to consider the relevant memory processes in more detail. We frame our discussion within a view that proposes two distinct sets of memory processes: One based on familiarity and one based on a more targeted, strategic memory search to recollect contextual details.

A dual-process account of misinformation effects

Familiarity is considered a quick and automatic appraisal process. When people encounter a statement, they judge its familiarity automatically and independently of the recollection of its source and its validity assessment (Begg, Anas, & Farinacci, 1992; Hintzman, 2010; Jacoby, 1991; Zimmer & Ecker, 2010). For example, unfounded statements such as 'the Great Wall of China can be seen from the moon,' 'sugar causes hyperactivity in children,' or 'humans use only 10 % of their brain' will be immediately familiar to many people even if they can also recollect having learned that these statements are, in fact, myths. Yet, familiarity can lead to errors of judgment because it is often difficult to determine *why* a

piece of information is familiar. Hence, the simple fact that an item is familiar can lead to its acceptance as true (Schwarz, Sanna, Skurnik, & Yoon, 2007) or the impression that the item had been encountered just recently. An example of the latter can be seen in the 'mirror effect' in episodic recognition studies, where unstudied but highly familiar items (e.g., high-frequency words) typically attract more false alarms than less familiar items; cf. Reder, Angstadt, Cary, Erickson, & Ayers, 2002; Reder, Nhouyvanisvong, Schunn, Ayers, Angstadt, & Hiraki, 2000). In this case, people misattribute a sense of familiarity that arises from an item's high natural-language frequency (i.e., its general long-term familiarity) to its presumed occurrence in an experiment's study phase (i.e., a specific short-term familiarity or recency).

Likewise, a question about an event or the relationship among concepts will automatically activate related memory representations, some of which will be relevant and valid, but some of which may supply irrelevant, incorrect, or subsequently invalidated information (Ayers & Reder, 1998). For example, in the well-studied *post-event misinformation* literature (Loftus & Hoffman, 1989; Loftus & Palmer, 1974), people may retrieve information merely suggested to them after witnessing an event because that suggestive misinformation is more recent and hence potentially more familiar at retrieval (Ayers & Reder, 1998; Lindsay & Johnson, 1989). It might be suggested to a crime witness, for instance, that the victim's arm was hurt during a theft incidence, when in actual fact the witness saw the victim's neck being hurt. Because of the suggestion, the witness may later remember seeing the victim's arm being hurt rather than their neck (Okado & Stark, 2003).

The reliance on familiarity alone is therefore insufficient for accurate information processing. This implies the need for strategic memory processes: there needs to be a filter or monitoring process that can scrutinize the retrieved memory output, and that can direct a potentially more targeted memory search process with the aim of recollecting specific contextual details (M. Johnson, Hashtroudi, & Lindsay, 1993; Schacter, Norman, &

Koutstaal, 1998; Zimmer & Ecker, 2010). This filter and directed search is the domain of strategic memory processes. For example, recollecting how your PSYC101 lecturer showed you the evidence that people (mostly) use more than 10 % of their brain will prevent you from accepting the myth as true even though it is familiar. Likewise, a specific search for information regarding an item's source (e.g., whether it was actually encountered within the context of an experiment's study list, or whether it was actually encountered during a witnessed event) will—if successful—preclude its false acceptance as valid. In other words, if strategic memory processes are utilised and functional, misinformation effects based on familiarity will not occur. Alas, strategic memory processes often fail; they require more executive control and mental effort than automatic memory processes (Herron & Rugg, 2003; Jacoby, 1991), and hence they are more readily compromised by factors such as divided attention (Jennings & Jacoby, 1993; Troyer & Craik, 2000), old age (Jennings & Jacoby, 1993; Spencer & Raz, 1995), alcohol intoxication (Bisby, Leitz, Morgan, & Curran, 2010), or simply a lack of motivation.

It follows that people will utilise misinformation when (a) an invalid memory representation is activated by a cue, and (b) strategic memory processes fail (M. Johnson et al., 1993; Okado & Stark, 2003; Roediger & McDermott, 1995; Schacter et al., 1998).

This basic idea of a failed strategic memory process contributing to misinformation effects has been applied in various forms. Sometimes, one correctly remembers an aspect of an event but misremembers important details, such as the source of the information. Incorrect recollection of the source may contribute to misinformation effects, for example, when one erroneously remembers receiving a piece of information from a reliable source (e.g., reading it in a scientific review paper) when in fact one received it from an unreliable source such as a blog, or, as discussed above, when one remembers witnessing something (e.g., an injury of a crime-victim's arm) that was in fact merely suggested (M. Johnson et al., 1993; Ost,

Granhag, Udell, & af Hjelmsäter, 2008; Ruffman, Rustin, Garnham, & Parkin, 2001; Zaragoza & Lane, 1994).

Sometimes misinformation effects may arise when people remember the misinformation but do not recollect the retraction. Considering the earlier emphasis on people *being able to remember* the retraction but not *using* it during inferencing, this may seem like a contradiction.However, because recollection is a controlled process, there is a difference between *being able* to recollect something (in principle) and *actually* recollecting it (cf. Ecker, Zimmer, & Groh-Bordin, 2007). This is why people may use misinformation in their reasoning in response to indirect questions about the cause of an event, while still recollecting the correction when questioned about the event's cause, or asked even more directly whether a presented report contained any corrections (as discussed above). The assumption that misinformation effects may occur when people remember the misinformation but fail to recollect the retraction is supported by research on negations. These studies assume that representations of information that is retracted or negated remain intact but the information's impact is usually offset by concurrent retrieval of a retraction/negation "tag" affixed to the misinformation representation in memory (Ecker et al., 2011a; Gilbert, Krull, & Malone, 1990; Wilson & Park, 2008)—when the tag is lost or is not retrieved alongside 'its' misinformation, misinformation effects occur.

Some evidence for the importance of strategic monitoring processes for the continued-influence effects of misinformation comes from a study by Ecker et al. (2010). In that study, people read news reports containing corrections. Participants relied less on retracted misinformation in their reasoning when they were informed about the basic notion of misinformation effects, and were warned about the possible presence of misinformation, before reading. The authors suggested that warnings put participants into a more strategic

processing mode, allowing for both more efficient "tagging" of misinformation at encoding and better strategic monitoring at retrieval (cf. also Echterhoff, Hirst, & Hussy, 2005).

As mentioned above, the efficiency of strategic memory and monitoring processes is known to depend on age, with both children and the elderly having less efficient strategic processes. It is therefore unsurprising that misinformation effects tend to be particularly pronounced in children (Ruffman et al., 2001) and in the elderly (Ansburg & Heiss, 2012; Dehon, 2006; Skurnik, Yoon, Park, & Schwarz, 2005; cf. also Spencer & Raz, 1995).

By the same token, because strategic monitoring is an effortful process, misinformation effects are generally more likely to occur when people operate under high cognitive load or with few cognitive resources available (Ansburg & Heiss, 2012; Dehon, 2006; Gilbert et al., 1990; Wilson & Wolf, 2009). For example, when participants are sleep-deprived at retrieval, they are more likely to produce false memories in the Deese-Roediger-McDermott (DRM) paradigm (Diekelmann, Landolt, Lahl, Born, & Wagner, 2008; note: in the DRM paradigm, participants study the associates of a focal word [e.g., cloud, water, etc.] but not the focal word itself [e.g., rain]; yet participants reliably misremember studying the focal word). In Diekelmann et al.'s study, sleep-deprivation (i.e., fatigue, not wakefulness after study) led to higher levels of false recognition of the non-studied focal word.

In summary, people will tend to rely on automatically retrieved, familiar information when they do not have the cognitive resources to invest in monitoring their memory and reasoning. Resources might be compromised by many factors including age and the demands of the situation. Whenever people rely on automatic memory processes alone, the potential for inadvertent reliance on information that is known to be false is greatest.

Why retractions fail: The effects of repetition and familiarity

Because of the impact of automatic memory processes, making (mis-)information more familiar is one of the most powerful ways to make people believe and use it. It is for

this reason that advertisements are put on high rotation, and this also explains why politicians repeat the same talking points ad infinitum (Singh, Kristensen, & Villaseñor, 2009). Moons, Mackie, and Garcia-Marques (2009) reported that repeating an argument led to stronger agreement with it; if the argument was a weak one, however, repetition led to stronger agreement only in participants with low induced motivation for analytic processing—a finding that supports the link between low cognitive effort / strategic memory processing and familiarity-based misinformation effects.

It may seem rational to more strongly believe in statements one has heard repeatedly from various sources—in fact, a statement encountered in different contexts will be particularly memorable (e.g., Verkoeijen, Rikers, & Schmidt, 2004), and true information is more likely to be encountered in a large number of contexts than a fabrication. However, even repetition from a single source will make a statement more believable (Weaver, Garcia, Schwarz, & Miller, 2007), and internally repeating a statement by retrieving it from memory may be even more powerful in generating perceived validity than actual (external) repetition (Ozubko & Fugelsang, 2011). It is particularly remarkable that even when people accurately recognize that a statement is familiar simply because it was just encountered in a lab experiment, or when the source is known to be unreliable, the statement's familiarity will still influence people's belief in it (Begg et al., 1992; Henkel & Mattson, 2011).

Given how generally effective repetition seems to be as a memorial agent, it is tempting to assume that the same benefits of repetition and familiarity might also arise for retractions: Perhaps three corrections are better than one? Contrary to that expectation, the repetition of retractions appears to be less powerful than the repetition of initial information. In a study by Ecker et al. (2011a), repeating a piece of misinformation strongly increased people's reliance on it, but repeating the retraction only had a limited effect. In their study, three retractions were more effective than a single retraction after misinformation was

encoded three times, but three retractions were as (in-)effective as one retraction when a piece of misinformation was encoded only once. The authors explained this in the framework of "retraction-tags" that was introduced earlier. Ecker et al. argued that each repetition of misinformation lays down a separate memory representation, and repetition thus leads to higher overall availability of misinformation in memory (because there are multiple traces). However, each representation can only be offset by one retraction, which "tags" the representation as being incorrect. Repeated retractions therefore have only limited efficacy because once a representation has been tagged, additional repetitions of the retraction have no effect on that particular representation.

Another explanation for why retractions, and repeated retractions in particular, are so ineffective can be derived from the dual-process perspective. We have already discussed how retractions are often ineffective: It turns out that, paradoxically, retractions may even *increase* an individual's acceptance of incorrect information, a class of phenomena known as "backfire" or "boomerang" effects. One example of this is the "familiarity backfire effect." This counterintuitive phenomenon occurs when a retraction repeats the misinformation, thereby inadvertently increasing its familiarity. For example, the statement "the belief that the MMR vaccine causes autism is false" repeats the association between MMR and autism and may make the statement "MMR causes autism" seem more familiar. When recollection of specific contextual information fails—memory for details such as the source of the information or its "tagging" as false can be expected to deteriorate over time—only the enhanced familiarity of the statement may remain (Skurnik et al., 2005; also see Schwarz et al., 2007). The marketing industry is well aware of this effect: Even negative advertisements and negative publicity can have a positive long-term effect on sales when it serves to make people more familiar with a product or brand (Berger, Sorensen, & Rasmussen, 2010).

The familiarity backfire effect has been demonstrated most clearly by Skurnik et al. (2005). These authors presented participants with health-related claims such as "Aspirin destroys tooth enamel." Claims were either labeled as valid or invalid (because of ethical concerns with presenting misleading health claims, all claims were actually valid but were not obviously true or false to the non-expert participants), and were presented once or three times. Thus, an invalid item presented three times would have been identified as invalid three times, as opposed to just once for the once-presented counterpart. The results were intriguing. On the one hand, Skurnik et al. found that repeating claims that were labeled invalid led older adults to misremember them as valid *less* often after a short delay of 30 minutes (presented three times: 17 % vs. once: 28%), thus attesting to the success of repeated retractions (young adults had much lower rates, misremembering 7 % of invalid claims presented three times and 10 % of invalid claims presented once). On the other hand, after a delay of three days, the effect was reversed, with older people misremembering the thrice-repeated invalid items as valid more often (40 %) than their once-presented counterparts (28 %). (Young adults did not show this effect, with rates for invalid claims presented thrice vs. once at 14 and 24 %, respectively.) Skurnik et al. argued that although repetition served to highlight that a claim was invalid, thereby facilitating its immediate rejection, the repetition also increased the claim's familiarity and thus led to higher levels of misinformation acceptance—at least in older participants—after a delay. After a delay, strategic recollection processes were particularly important, lest they be overwhelmed by relatively intact familiarity-based memory. Because strategic processes are generally weaker in the elderly, they were insufficient to counteract the (misleading) familiarity signal arising from the repetition of the invalid information (cf. Prull, Dawes, Martin, Rosenberg, & Light, 2006; Spencer & Raz, 1995).

Wilson and Park (2008) subsequently adapted the statements used by Skurnik et al. (2005). However, rather than labeling statements as valid or invalid, Wilson and Park simply negated some of the statements (e.g., "Aspirin does not destroy tooth enamel"). Again, elderly people were found to be more likely to misremember negated statements as valid, this time after a mere 45 minutes.

Turning to the other end of the adult lifespan, research on misinformation in younger adults after shorter delays has hitherto yielded inconclusive evidence. Recall that Skurnik et al. (2005) found no evidence for a familiarity backfire effect in young adults after 30 minutes. Likewise, Ecker, Lewandowsky, and Hogan (2012a) presented participants with fictional news reports similar to the "warehouse fire" scenario, in which an initially-presented cause of an event subsequently turned out to be false, and was replaced by a valid cause. Ecker et al. found that repeating the initial, outdated cause during the retraction made the retraction *more*, not less, effective (cf. also Johar & Roggeveen, 2007).

In contrast, Skurnik, Yoon, and Schwarz (2007; as discussed in Schwarz et al., 2007) reported evidence for a familiarity backfire effect even in young adults after only 30 minutes of retention. They presented their participants with a flyer regarding the common flu vaccine, which employed the popular "myths vs. facts" format. For example, it stated that "The side effects [of the vaccine] are worse than the flu.—FALSE. The worst side effect you're likely to get […] is a sore arm […]." Skurnik et al. found that retracting (and thereby repeating) common myths led to a marked increase in accepting those myths as true. Interestingly, attitudes towards the vaccine had also changed—participants who had read the flyer rated the vaccine less favourably than those who had not. Berinsky (2012) reported that corrections of the rumor that U.S. health care changes would promote euthanasia tended to be more effective when the correction did not repeat the rumor.

In summary, with young adults the balance of evidence does not warrant undue concerns about the familiarity backfire effect. By contrast, in old adults, the presence of familiarity backfire effects seems to be well established and must give rise to concern.

We have identified the reliance on automatic memory processes, in the absence of effortful strategic processing, as a key contributing factor to the persistence of misinformation. That is, in the absence of a particular motivation to believe one version of an event, or one set of concept-relations, over another, and when there is no reason to doubt the validity of the retraction, people will rely on misinformation to the extent that strategic memory processes fail to retrieve the retraction. This could be due to a genuine failure, but could also happen when the retraction-information is available in principle, but is not retrieved because the retrieval cue is too indirect or people simply do not invest the effort required.The latter variable is particularly relevant because in many circumstances, people *will* be motivated to believe some pieces of information but not others. We therefore next consider the heuristics involved in people's judgment of the believability of incoming information, and the role of people's existing beliefs, attitudes, and worldviews in shaping their beliefs and memory.

Misinformation effects and social cognition

It is impossible to critically assess the veracity of each and every piece of information one encounters, which makes the use of heuristics that can yield quick judgments of credibility particularly attractive. Thus people will often assess the validity of a claim by assessing how it fits in with what they (or their peers) already know or believe. This can be a very efficient strategy and is usually adaptive (e.g., anyone would be skeptical about claims of a newly discovered Pacific island on which there is no gravity). In support, Richter, Schroeder, and Wöhrmann (2009) demonstrated how relevant pre-existing knowledge is used

by people to reject false assertions quickly and automatically, provided that the pre-existing knowledge is accessible and people are certain that the knowledge is accurate.

The role of attitudes for misinformation effects

Potential problems arise from use of this heuristic when the existing knowledge is false, beliefs are biased, or people hold very strong attitudes that run counter to the evidence. Ecker, Lewandowsky, Fenton, and Martin (2012b) showed that pre-existing attitudes determine how often a piece of attitude-relevant (mis-)information is used in reasoning, but they also demonstrated that pre-existing attitudes do not necessarily compromise the effectiveness of retractions. In their study, people scoring high or low on a racial-prejudice scale read a news report involving either an Aboriginal robber or an Aboriginal hero; the involvement of the Aboriginal was retracted in some conditions. Ecker et al. found that references to retracted racial misinformation co-varied with people's racial prejudice (e.g., people with high racial prejudice referred to the Aboriginal robber more often, but to the Aboriginal hero less often, than people lower in prejudice). Ecker et al. also found that people in both high- and low-prejudice groups reduced their reliance on the racial information after a retraction to equivalent degrees, meaning that there was no evidence that people ignored the retraction when it ran counter to their attitudes.

However, this outcome stands in striking contrast to other research which has found that people holding a strong belief may reject *any* information that runs counter to that belief. Thus, attempts to retract attitude-congruent misinformation can be futile with strongly partisan individuals. Indeed, it can even backfire and strengthen the misconception ("the worldview backfire effect;" see Lewandowsky et al., 2012, for a review). For example, Nyhan and Reifler (2010) reported that attempts to refute non-fictitious misconceptions (e.g., that weapons of mass destruction had been found in Iraq during the 2003 invasion) were successful when the retractions were in line with people's worldview, but that worldview-

incongruent retractions were ineffective or ironically strengthened the misconceptions. It is presently unclear under what circumstances such worldview backfire effects can be observed, but Ecker et al. (2012b) argued that it will depend on whether acknowledging the retraction will require a change of attitudes: In the case of an arbitrary incident such as a robbery, one can uphold one's belief (e.g., that most Aboriginals are criminal) even when accepting that the particular incident (e.g., that a particular crime was not committed by an Aboriginal) did not support the belief. In contrast, a refutation will likely not be accepted when it necessitates a change in belief, and potentially entails a loss of face. For example, for a person who supported the 2003 invasion of Iraq, acknowledging that no weapons of mass destruction were found in Iraq may require a re-assessment of political decisions made at the time, and re-assessment of one's own support.

Motivated reasoning and misinformation effects

The rejection of attitude-incongruent information can be seen as a case of motivated reasoning—biased information processing that seeks to confirm existing beliefs rather than update one's knowledge based on an unbiased assessment of the evidence (for reviews of motivated cognition, see Kunda, 1990; Nickerson, 1998; Redlawsk, 2002). Motivated reasoning behavior is particularly prevalent when the belief or attitude in question is held with confidence and conviction because it is central to a person's value system and defines their identity (Brannon, Tagler, & Eagly, 2007; Hart, Albarracín, Eagly, Brechan, Lindberg, & Merrill, 2009). This means that efforts to correct common misperceptions may have to focus on people with moderate—rather than extremely strong—convictions. Even in people with moderate attitudes, however, motivated reasoning seems sufficiently common (cf. Prasad et al., 2009) to warrant a more detailed examination.

The confirmation bias behind motivated reasoning can lead people to selectively expose themselves only to ("cherry-picked") evidence that supports their prior beliefs. An

example of such behavior is the tendency of consumers to choose newspapers that are known to be biased in their reporting, or internet users to frequent websites that they know to support their worldviews (T. Johnson, Bichard, & Zhang, 2009). Other strategies include denying the existence of contrasting evidence or—when presented with such evidence as in the case of misinformation retractions discussed here—deflecting or discrediting the evidence or its source. For example, when a psychologist corrects the common myth that there is a debate amongst climate scientists as to whether climate change is anthropogenic, explaining how there is in fact a strong consensus on this (cf. Anderegg, Prall, Harold, & Schneider, 2010; Doran & Zimmerman, 2009), a frequent strategy of people who oppose the findings from climate science is to deflect and discredit the source by questioning the psychologist's expertise to speak about climatology. Supporting this anecdotal evidence, derogating the source of the evidence was one of the strategies identified by Jacks and Cameron (2003) in their study on motivated reasoning. The authors presented counter-attitudinal messages (arguing against the death penalty) to participants in favor of the death penalty, and measured people's responses. The strategy of derogating the source was used by their participants in particular when the source of the counter-attitudinal message was perceived as lacking expertise, and when participants' pro death penalty attitudes were particularly strong.

The two most frequently used strategies in the study by Jacks and Cameron (2003), however, were counter-arguing and attitude bolstering. Counter-arguing refers to the covert or overt generation of arguments to counter the corrective information being presented, a common and principally rational strategy as long as the counterarguments maintain ties to empirical reality. Attitude bolstering means to completely ignore the refutation and focus on supporting evidence.

In a study by Prasad and colleagues (2009), participants were presented with convincing evidence that was incongruent with both a specific misconception they held and

their general political worldview (specifically, Republican voters who believed in a link between Saddam Hussein and the terror attacks of 9/11 were presented with contrary evidence including a statement from President G. W. Bush himself). Faced with the evidence, only 2 % of participants explicitly acknowledged the inadequacy of their misconception and updated their beliefs. A higher proportion (14 %) denied holding the misconception in the first place, a surprising behavior that may have served to avoid "losing face." Most participants, however, displayed some form of motivated reasoning. The utilized strategies included the above-mentioned counter-arguing of facts and attitude bolstering. Prasad et al. also identified two other strategies. One was 'inferred justification,' which describes a form of flawed backward reasoning where people recursively infer the reasons that justify a present situation, ignoring evidence to the contrary. For example, some people argued that there must have been a link between Hussein and 9/11 because there must have been a reason for the U.S. invasion of Iraq. The other strategy identified by Prasad et al. was 'disputing rationality;' this is when people insist on their right to an opinion without factual reasoning, often combined with assertions of confidence that nothing could change their mind.

Munro (2010) additionally described a situation in which people discounted evidence by denying the in-principle amenability of a topic to scientific investigation ('scientific impotence discounting'). In that study, participants who believed that homosexuality was (or was not) a mental illness rejected evidence that homosexuals are not (or are) over-represented in psychological treatment facilities, and were more likely to agree with the suggestion that the topic could not be studied scientifically. In fact, the rejection of science even generalized to other topics, implying that when given evidence inconsistent with their beliefs, people became more skeptical of the scientific method in general.

Not surprisingly, motivated reasoning tendencies are less prevalent in open-minded people (Hart et al., 2009). Levitan and Visser (2008) asked their participants to rate the

heterogeneity of attitudes in their social network by having them list five members of their social network, and rate how much they agreed with each of those people's worldviews. The study demonstrated that people exposed to diverse attitudes through their social network are more open to change their attitudes in response to counter-attitudinal persuasive arguments, provided arguments are sufficiently strong (i.e., attitude diversity did not lead to attitude change in response to weak arguments).

Some researchers have argued that personality traits such as "social vigilantism"—the tendency to impose one's own beliefs on others—predict greater expressions of belief superiority and increased counter-arguing tendencies (Saucier & Webster, 2010). Other researchers have expressed the somewhat contrasting view that motivated reasoning tendencies are linked to low self-esteem (Wiersema, van Harreveld, & van der Pligt, 2012), negative mood (Jonas, Graupmann, & Frey, 2006), and threats of social exclusion (Greitemeyer, Fischer, & Kastenmüller, 2012).

The idea that people with low self-esteem (or people in a situation of reduced emotional or social stability) have a stronger need to defend their identity-defining attitudes is in line with self-affirmation theory. Self-affirmation theory claims that people use a variety of psychological adaptations, strategies, and distortions to protect the self from the threats of perceived failure or deficiencies (cf. Sherman & Cohen 2006). In the present case, this implies that people with low self-esteem and strong attitudes will find counter-attitudinal evidence particularly threatening to their identity and self-worth, and are hence more likely to engage in motivated reasoning. Motivated reasoning tendencies can therefore be softened by affirming people's self-worth, for example by focusing them on personal achievements before confronting them with attitude-incongruent corrective information (Cohen, Bastardi, Sherman, Hsu, McGoey, & Ross, 2007).

The motivated-reasoning research implies that pre-existing attitudes can create a major obstacle for attempts to correct misinformation, in particular when corrections are perceived to require attitudinal change. The literature just reviewed suggests that counter-attitudinal retractions are more likely to be accepted when people are reminded of attitudinal diversity in their social network, when they are in a good mood and self-affirmed, and not threatened by social exclusion.[1]

Correcting misinformation

We have already discussed that mere retractions (even when repeated) are fairly ineffective in reducing reliance on invalid or outdated information. Further, the above discussion of the familiarity backfire effect suggests that at least in older people, repeating misinformation during attempts to correct it carries the inherent risk of ironically strengthening the misconception (Skurnik et al., 2005, 2007). This finding presents a pragmatic challenge, because often a correction will require a myth to be repeated; for example, the statement 'the Great Wall of China is *not* visible from the moon' may inadvertently strengthen the link between 'the Great wall of China' and 'visible from the moon', yet it is unclear how the debunking could be performed in this instance without mentioning the misinformation. Finally, we have discussed how motivated reasoning can make it difficult to correct misconceptions when the correction runs counter to strongly held beliefs. In some instances, pre-existing beliefs can be so powerful that the presentation of corrective information can ironically fortify the misconception (viz. the 'worldview backfire effect'; Nyhan & Reifler, 2010).

[1]While it goes beyond the scope of this chapter, there are also strategies that aim more at changing behavioral tendencies rather than the attitudes and beliefs per se. These include the design of choice architectures (e.g., opt-in vs. opt-out approaches to superannuation or organ donation schemes) and the framing of information with respect to attitude-congruence (e.g., a person concerned about climate change might be persuaded to install solar panels on their house by arguments referring to their carbon footprint, but persons who are not concerned about climate change may find economical arguments more convincing). See Lewandowsky et al. (2012) for a brief review.

In opposition to those concerns involving backfire effects, there is a growing literature on the *deliberate* use of misinformation, and its refutation, as an educational tool. Students acquiring basic knowledge in any subject will typically harbor a variety of misconceptions, and techniques that can effectively reduce these are of interest to all teachers and science communicators. Directly addressing and refuting such misconceptions to introduce the valid information has been shown to be more effective than presenting the same valid information in a 'standard' teaching format (Kowalski & Taylor, 2009). Unlike plain retractions which simply state that some piece of information is not true, a refutational text is more detailed; although it explicitly presents the misinformation, it also provides a comprehensive explanation of why it is incorrect. For the 'Great Wall of China' example, a refutation might include supporting evidence such as the narrow width of the wall relative to the distance between the moon and the earth, and how that translates into a visual angle too small for the human eye to resolve. A refutational text might also try to explain why the misinformation was presented in the first place. For example, to understand why Andrew Wakefield suggested that autism was linked to the MMR vaccine, it helps to know that he received around half a million pounds in undisclosed payments from a lawyer preparing class action against the producers of the compound vaccine, and that there were plans to start a company to sell diagnostic tests (Deer, 2011). Arguably, such explanations will facilitate belief updating because they foster people's skepticism regarding the initial misinformation and its source (cf. Lewandowsky, Stritzke, Oberauer, & Morales, 2005, 2009).

A meta-analysis by Guzzetti, Snyder, Glass, and Gamas (1993) explored the efficacy of different interventions intended to reduce misconceptions in the classroom. Participants ranged from primary school students to university postgraduates, and a wide range of science misconceptions were incorporated. The authors found seven types of intervention being utilized within these studies, with refutational text and non-refutation texts (such as simple

narratives, or expository texts where new information is presented with no reference to the misconception), being the most common. Refutation texts were found to be the most effective strategy of all interventions, and significantly more effective in producing conceptual change than non-refutational text. Several studies included a post-test one or two months later, and refutational text was the only intervention to foster long-term conceptual change.

Efficacy of a refutation is dependent on a number of factors. Guzetti, Williams, Skeels, and Wu (1997) observed three high school physics classes for an 8-month period, accumulating qualitative data through interviews and questionnaires. They found that students ignored the refutation text and persisted with misinformation if (1) the refutation was poorly constructed and lacked clarity or (2) the students self-reported that they were only skimming the text. When students did not engage with the text, they mistakenly found evidence for their prior misconception within the refutation. It is also important that the misconception is based upon a lack of knowledge (which may be corrected by a valid explanation), rather than an individual's attitudes or belief system, as previously discussed.

The effectiveness of refutation texts is often explained within the 'conceptual change model' proposed by Posner, Stike, Hewson, and Gertzog (1982). This model suggests that there are four stages necessary for conceptual change to take place. The first step is *dissatisfaction* with one's own current understanding, which instigates cognitive conflict. The proposed replacement construct then needs to be *intelligible* (i.e., easily understood), *plausible*, and potentially *fruitful* (i.e., the student should be able to see the relevance of a correct understanding of matters and the potential applicability of the corrected knowledge to future problems).

Yet, considering the first step in Posner et al.'s (1982) model, to become dissatisfied with one's insufficient understanding, one would first need to notice an incongruity between one's current conception and the evidence. In fact, before one can notice such a discrepancy,

one's misconception, the presented evidence (and the associated correction) would have to be co-activated and aligned in working memory (cf. McCrudden & Kendeou, in press; van den Broek & Kendeou, 2008).

McCrudden (2012) looked at students' understanding of the biological term 'fitness,' which refers specifically to the number of an animal's offspring that survive into adulthood, a meaning quite different from the common interpretation in terms of physical fitness. Based on an analysis of think-aloud protocols and interviews, McCrudden argued that when reading a refutational text, co-activation of the misconception and the corrective evidence may occur automatically, but that the detection of a discrepancy between the two may require strategic monitoring processes (cf. Ecker et al., 2010). Using think-aloud protocols, McCrudden further demonstrated how merely noticing the discrepancy between misconception and evidence is not sufficient for a change in belief—the reader needs to strategically engage in resolving the discrepancy. Kendeou and van den Broek (2007) found that students with prior misconceptions engaged in such 'conceptual change strategies' only when reading refutation texts. These conceptual change or resolution strategies were reflected through participants vocalizing the juxtaposition between their misconception and the correct information, making text-based inferences, and paraphrasing (McCrudden, 2012; Kendeou & van den Broek, 2007).

The above discussion implies that one of the main reasons why refutational texts may be particularly effective in reducing misconceptions is that they set the stage for strategic 'conceptual change' processing. That is, refutational texts allow people to co-activate, align, and integrate their misconceptions with corrective evidence, which then facilitates the updating of their beliefs (cf. Ecker et al., 2012a; Johar & Roggeveen, 2007).

This view of corrections via refutation is consistent (although it does not require) the notion of memory reconsolidation, which has gained recent popularity (see Hardt, Einarsson,

& Nader, 2010, for a review, but also Ecker & Lewandowsky, 2012, for a critique). Reconsolidation theory proposes that when an item in memory is retrieved, it re-enters a transient, labile state, which may allow it to be modified, updated, or weakened, before is then re-stabilized (i.e., re-consolidated). When a refutation is read, the original misconception is necessarily recalled, and according to reconsolidation, this would render the misconception's representation labile and amenable to be updated and to accommodate the correct information.

Apart from the 'aligning' of misconception and correction, the effectiveness of using refutational materials in a classroom setting may have another 'operative ingredient:' It fosters critical thinking. Refutational texts encourage students to critically appraise what they know, to skeptically assess empirical evidence, and to endeavour to draw valid conclusions from the evidence (cf. Baker & Piburn, 1991; Berland & Reiser, 2008; Kuhn & Crowell, 2011; Manson, Gava, & Boldrin, 2008; ten Dam & Volman, 2004). Bedford (2010) suggested that the in-depth study of misinforming materials itself could improve students' understanding of scientific concepts, by motivating students to acquire the necessary knowledge and critical thinking skills to convincingly argue against misrepresentations. The explicit study of misinforming materials and the analysis of controversial issues can thus ameliorate students' critical thinking and reasoning skills. That is, reading refutational texts and discussing them in-class, building arguments based on evidence, will help students improve both their understanding of the subject matter and their argumentative reasoning skills. Alas, knowledge is currently often imparted as a set of unequivocal facts, and there is a lack of argument and debate in science classrooms (Osborne, 2010).

To conclude, our analysis of misinformation correction strategies suggests that familiarity backfire effects may only be a concern when there is little explanation regarding the motivation behind the initial misinformation and the evidence (i.e., in simple "this is the

myth, this is the fact" situations). In situations in which more extensive explanations are possible, taking up and addressing the misinformation seems beneficial. It allows misconceptions and corrective evidence to be aligned, and seems to facilitate belief updating, in particular when a 'factual wrapper' can be established to encapsulate misinformation. This can be achieved by providing clear-cut evidence, and explaining the motivation behind the initial spreading of misinformation (i.e., a "this is the myth, this is why the myth was spread, this is the truth, this is the evidence" approach). Fostering critical thinking, skepticism, epistemological knowledge and an understanding of science from a young age thus seem crucial educational aims for contemporary information societies.

References

Anderegg, W. R. L., Prall, J. W., Harold, J., & Schneider, S. H. (2010). Expert credibility in climate change. *Proceedings of the National Academy of Sciences USA, 107*, 12107-12109.

Ansburg, P. I., & Heiss, C. J. (2012). Potential paradoxical effects of myth-busting as a nutrition education strategy for older adults. *American Journal of Health Education, 43*, 31-37.

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

Baker, D. R., & Pibum, M. (1991). Process skills acquisition, cognitive growth, and attitude change of ninth grade students in a scientific literacy course. *Journal of Research in Science Teaching, 28*, 423-436.

Bedford, D. (2010). Agnotology as a teaching tool: Learning climate science by studying misinformation. *Journal of Geography, 109,* 159-165.

Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General, 121*, 446-458.

Berger, J., Sorensen, A. T., & Rasmussen, S. J. (2010). Positive effects of negative publicity: When negative reviews increase sales. *Marketing Science, 29,* 815-827.

Berland, L., K., & Reiser, B., R. (2008). Making sense of argumentation and explanation. *Science Education, 93,* 26-55.

Bisby, J. A., Leitz, J. R., Morgan, C. J. A., & Curran, H. V. (2010). Decreases in recollective experience following acute alcohol: A dose-response study. *Psychopharmacology, 208*, 67-74.

Brannon, L. A., Tagler, M. J., & Eagly, A. H. (2007). The moderating role of attitude strength in selective exposure to information. *Journal of Experimental Social Psychology, 43*, 611-617.

Chater, N. (1999). The search for simplicity: A fundamental cognitive principle? *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 52*, 273-302.

Cohen, G. L., Bastardi, A., Sherman, D. K., Hsu, L., McGoey, M., & Ross, L. (2007). Bridging the partisan divide: Self-affirmation reduces ideological closed-mindedness and inflexibility in negotiation. *Journal of Personality and Social Psychology, 93*, 415-430.

De Beni, R., & Palladino, P. (2004). Decline in working memory updating through ageing: Intrusion error analyses. *Memory, 12* , 75-89.

Deer, B. (2011). How the vaccine crisis was meant to make money. *BMJ, 342*, art. c5258, 136-142.

Dehon, H. (2006). Variations in processing resources and resistance to false memories in younger and older adults. *Memory, 14*, 692-711.

DeStefano, F., & Thompson, W. W. (2004). MMR vaccine and autism: An update of the scientific evidence. *Expert Review of Vaccines, 3*, 19-22.

Diekelmann, S., Landolt, H., Lahl, O., Born, J., & Wagner, U. (2008). Sleep loss produces false memories. *PLoS ONE,* 3, e3512. doi:10.1371/journal.pone.0003512.

Doran, P. T., & Zimmerman, M. K. (2009). Examining the scientific consensus on climate change. *Eos, 90*, 21-22.

Echterhoff, G., Hirst, W., & Hussy, W. (2005). How eyewitnesses resist misinformation: Social postwarnings and the monitoring of memory characteristics. *Memory & Cognition, 33*, 770-782.

Ecker, U. K. H., & Lewandowsky, S. (2012). Computational constraints in cognitive theories of forgetting. *Frontiers in Psychology, 3,* 400. doi: 10.3389/fpsyg.2012.00400

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

Ecker, U. K. H., Lewandowsky, S., Fenton, O., & Martin, K. (2012b). *Do people keep believing because they want to? The effects of racial attitudes on misinformation processing.* Unpublished manuscript.

Ecker, U. K. H., Lewandowsky, S., & Hogan, J. (2012a). *Repeating misinformation when retracting it—help or hindrance?* Unpublished manuscript.

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

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

Ecker, U. K. H., Zimmer, H. D., & Groh-Bordin, C. (2007). Color and context: An ERP study on intrinsic and extrinsic feature binding in episodic memory. *Memory & Cognition, 35*, 1483-1501.

Editors of the Lancet (2010). Retraction—Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children (Retraction of vol 351, pp. 637-641, 1998). *Lancet, 375*, 445.

Freed, G. L., Clark, S. J., Butchart, A. T., Singer, D. C., & Davis, M. M. (2011). Sources and perceived credibility of vaccine-safety information for parents. *Pediatrics, 127*, S107-S112.

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

Greitemeyer, T., Fischer, P., & Kastenmüller, A. (2012). The effects of social exclusion on confirmatory information processing. *European Journal of Social Psychology, 42*, 462-469.

Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting conceptual change in science: A comparative meta-analysis of instructional interventions from reading education and science education. *Reading Research Quarterly, 28*, 116-159.

Guzzetti, B. J., Williams, W. O., Skeels, S. A., & Wu, S. M. (1997). Influence of text structure on learning counterintuitive physics concepts. *Journal of Research in Science Teaching, 34*, 701-719.

Hardt, O., Einarsson, E. O., & Nader, K. (2010). A bridge over troubled water: Reconsolidation as a link between cognitive and neuroscientific memory research traditions. *Annual Review of Psychology, 61*, 141-167.

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin, 135*, 555-588.

Health Protection Agency (2011, 21 October). *Completed primary courses at two years of age: England and Wales, 1966-1977, England only after 1978 onwards.* Retrieved from http://www.hpa.org.uk/web/HPAweb&HPAwebStandard/HPAweb_C/1195733819251

Herron, J. E., & Rugg, M. D. (2003). Strategic influences on recollection in the exclusion task: Electrophysiological evidence. *Psychonomic Bulletin & Review, 10*, 703-710.

Heit, E., & Rotello, C. M. (2012). The pervasive effects of argument length on inductive reasoning. *Thinking and Reasoning, 18*, 244-277.

Henkel, L. A., & Mattson, M. E. (2011). Reading is believing: The truth effect and source credibility. *Consciousness and Cognition,20*, 1705-1721.

Hintzman, D. L. (2010). How does repetition affect memory? Evidence from judgments of recency. *Memory & Cognition, 38*, 102-115.

Jacks, J. Z., & Cameron, K. A. (2003). Strategies for resisting persuasion. *Basic and Applied Social Psychology, 25*, 145-161.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513-541.

Jennings, J.M., & Jacoby, L.L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology and Aging, 8*, 283-293.

Johar, G. V., & Roggeveen, A. L. (2007). Changing false beliefs from repeated Advertising: The role of claim-refutation alignment. *Journal of Consumer Psychology, 17*, 118-127.

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

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28.

Johnson, T. J., Bichard, S. L., & Zhang, W. (2009). Communication communities or ‘‘cyberghettos?’’: A path analysis model examining factors that explain selective exposure to blogs. *Journal of Computer-Mediated Communication, 15*, 60-82.

Jonas, E., Graupmann, V., & Frey, D. (2006). The influence of mood on the search for supporting versus conflicting information: Dissonance reduction as a means of mood regulation? *Personality and Social Psychology Bulletin, 32*, 3-15.

Kendeou, P., & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition, 35*, 1567-1577.

Kowalski, P., & Taylor, A. K. (2009). The effect of refuting misconceptions in the introductory psychology class. *Teaching of Psychology, 36*, 153-159.

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

Kuhn, D., & Crowell, A. (2011) Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science, 22,* 545-552.

Levitan, L. C., & Visser, P. S. (2008). The impact of the social context on resistance to persuasion: Effortful versus effortless responses to counter-attitudinal information. *Journal of Experimental Social Psychology, 44*, 640-649.

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

Lewandowsky, S., Stritzke, W. G. K., Oberauer, K., & Morales, M. (2005). Memory for fact, fiction, and misinformation: The Iraq War 2003. *Psychological Science, 16,* 190-195.

Lewandowsky, S., Stritzke, W. G. K., Oberauer, K., & Morales, M. (2009). Misinformation and the "War on Terror": When memory turns fiction into fact. In W. G. K. Stritzke, S. Lewandowsky, D. Denemark, J. Clare, & F. Morgan (Eds.), Terrorism and torture: An interdisciplinary perspective (pp. 179-203). Cambridge, UK: Cambridge University Press.

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition, 17,* 349-358.

Loftus, E. F., & Hoffman, H. G. (1989). Misinformation and memory: The creation of new memories. *Journal of Experimental Psychology: General, 118,* 100-104.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning & Verbal Behavior, 13*, 585-589.

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

Madsen, K. M., Anders, H., Vestergaard, M., Schendel, D., Wohlfart, J., Thorsen, P., Olsen, J., & Melbye, M. (2002). A population based study of measles, mumps and rubella vaccination and autism. *The New England Journal of Medicine, 347,* 1477-1482.

Manson, L., Gava, M., & Boldrin, A. (2008). On warm conceptual change: The interplay of text, epistemological beliefs and topic interest. *Journal of Educational Psychology, 100,* 291-309.

McCarthy, J. (9th March 2012). MMR Doctor Exonerated – Who's Guilty Now? *Generation Rescue.* Retrieved from http://www.generationrescue.org/latest-news/jenny-mccarthy/mmr-doctor-exonerated-who-s-guilty-now/

McCrudden, M. T., & Kendeou, P. (in press). Exploring the link between cognitive processes and learning from refutational text. *Journal of Research in Reading*. doi: 10.1111/j.1467-9817.2011.01527.x

McCrudden, M. T. (2012). Readers' use of online discrepancy resolution strategies. *Discourse Processes, 49*, 107-136.

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

Munro, G. D. (2010). The scientific impotence excuse: Discounting belief-threatening scientific abstracts. *Journal of Applied Social Psychology, 40*, 579-600.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175-220.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior, 32*, 303-330.

Oberauer, K., & Vockenberg, K. (2009). Updating of working memory: Lingering bindings. *Quarterly Journal of Experimental Psychology, 62*, 967-987.

Okado, Y., & Stark, C. (2003). Neural processing associated with true and false memory retrieval. *Cognitive, Affective and Behavioral Neuroscience, 3*, 323-334.

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science, 328*, 463-466.

Ost, J., Granhag, P.-A., Udell, J., & af Hjelmsäter, E. R. (2008). Familiarity breeds distortion: The effects of media exposure on false reports concerning media coverage of the terrorist attacks in London on 7 July 2005. *Memory, 16*, 76-85.

Ozubko, J. D., & Fugelsang, J. (2011). Remembering makes evidence compelling: Retrieval from memory can give rise to the illusion of truth. *Journal of Experimental Psychology: Learning Memory & Cognition, 37*, 270-276.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Towards a theory of conceptual change. *Science Education, 66*, 211-227.

Prasad, M., Perrin, A. J., Bezila, K., Hoffman, S. G., Kindleberger, K., Manturuk, K., . . . Powers, A. S. (2009). "There must be a reason": Osama, Saddam, and inferred justification. *Sociological Inquiry, 79*, 142–162.

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

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

Rapp, D. N., & Kendeou, P. (2009). Noticing and revising discrepancies as texts unfold. *Discourse Processes, 46*, 1-24.

Reder, L. M., Angstadt, P., Cary, M., Erickson, M. A., Ayers, M. S. (2002). A reexamination of stimulus-frequency effects in recognition: Two mirrors for low- and high-frequency pseudowords. *Journal of Experimental Psychology: Learning Memory and Cognition, 28*, 138-152.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning Memory and Cognition, 26*, 294-320.

Redlawsk, D. P. (2002). Hot cognition or cool consideration? Testing the effects of motivated reasoning on political decision making. *Journal of Politics, 64*, 1021-1044.

Richter, T., Schroeder, S., & Wöhrmann, B. (2009). You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology, 96*, 538-558.

Roediger, H. L. III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 803-814.

Ruffman, T., Rustin, C., Garnham, W., & Parkin, A. J. (2001). Source monitoring and false memories in children: Relation to certainty and executive functioning. *Journal of Experimental Child Psychology, 80*, 95-111.

Saucier, D. A., & Webster, R. J. (2010). Social vigilantism: Measuring individual differences in belief superiority and resistance to persuasion. *Personality and Social Psychology Bulletin, 36*, 19-32.

Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology, 49*, 289-318.

Schwarz, N., Sanna, L. J., Skurnik, I., & Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. *Advances in Experimental Social Psychology, 39*, 127-161.

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. *Advances in Experimental Social Psychology, 38*, 183-242.

Singh, S., Kristensen, L., & Villaseñor, E. (2009). Overcoming skepticism towards cause related claims: The case of Norway. *International Marketing Review, 26,* 312-326.

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research, 31,* 713-724.

Skurnik, I., Yoon, C., & Schwarz, N. (2007). *Education about flu can reduce intentions to get a vaccination.* Unpublished manuscript.

Smith, A. & Yarwood, J. (2007). Tracking mothers' attitudes to MMR immunisation 1996-2006. *Vaccine, 25,* 3996-4002.

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging, 10,* 527-539.

ten Dam, G. & Volman, M. (2004) Critical thinking as a citizenship competence: teaching strategies. *Learning and Instruction, 14,* 359-379.

Troyer, A. K., & Craik, F. I. M. (2000). The effect of divided attention on memory for items and their context. *Canadian Journal of Experimental Psychology, 54,* 161-70.

van den Broek, P., & Kendeou, P. (2008). Cognitive processes in comprehension of science texts: the role of co-activation in confronting misconceptions. *Applied Cognitive Psychology, 22,* 335-351.

van Oostendorp, H., & Bonebakker, C. (1999). Difficulties in updating mental representations during reading news reports. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 319-339). Mahwah, NJ: Erlbaum.

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Schmidt, H. G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 796-800.

Wakefield, A., et al. (1998). Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet, 351,* 637-641. (retracted paper)

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,* 821–833.

Wiersema, D. V., van Harreveld, F., & van der Pligt, J. (2012). Shut yo ur eyes and think of something else: Self-esteem and avoidance when dealing with counter-attitudinal information. *Social Cognition, 30,* 323-334.

Wilkes, A. L., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 52A*, 165-183.

Wilson, E. A. H., & Park, D. C. (2008). A case for clarity in the writing of health statements. *Patient Education and Counseling, 72*, 330-335.

Wilson, E. A. H., & Wolf, M. S. (2009). Working memory and the design of health materials: A cognitive factors perspective. *Patient Education and Counseling, 74*, 318-322.

Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 934-945.

Zimmer, H. D., & Ecker, U. K. H. (2010). Remembering perceptual features unequally bound in object and episodic tokens: Neural mechanisms and their electrophysiological correlates. *Neuroscience and Biobehavioral Reviews, 34*, 1066-1079.

Author note: Preparation of this chapter was facilitated by Discovery Grants DP0770666 and DP110101266 from the Australian Research Council and by an Australian Postdoctoral Fellowship and an Australian Professorial Fellowship to the first and last author, respectively.