

# The Debunking Handbook

By John Cook and Stephan Lewandowsky

## Debunking the first myth about debunking

It's self-evident that democratic societies ought to base their decisions on accurate information. However, on many issues, particularly when vested interests are involved, misinformation can become entrenched in segments of the population (Jacques & Dunlap, 2008, Oreskes & Conway, 2010). Those engaged in the countering of misinformation can identify with the mythological Sisyphus, condemned for eternity to push a giant stone up a hill only to watch it repeatedly roll back down.

The most common misconception about misinformation is the notion that removing its influence is as simple as cramming more information down people's throats. This approach comes from the "information deficit model" of science communication, which assumes that public misperceptions are due to a lack of knowledge and the solution is more information. However, people don't process information as simply as a hard drive downloading data.

When dealing with misinformation, you're meddling with complex cognitive processes. Communicators wishing to successfully impart knowledge need to familiarise themselves with how people process information, how they modify their existing knowledge and how worldviews affect their ability to think rationally. It's not just *what* people think that matters, but *how* they think.

First, let's be clear about what we mean by misinformation: We use the label 'misinformation' to refer to any information that people have acquired that turns out to be incorrect. We are concerned with the cognitive processes that govern how people process corrections to information they have already acquired—if you find out that something you believe is wrong, how do you update your knowledge and memory?

Once people receive misinformation, it's exceptionally difficult to remove its influence. This was demonstrated in a 1994 experiment where people were exposed to misinformation in the context of a story about a fictitious warehouse fire, then given a correction clarifying that the misinformation was incorrect (Johnson & Seifert, 1994). Despite remembering and accepting the correction, people still showed a lingering effect, referring to the misinformation when answering questions about the story.

Is it possible to completely eliminate the influence of misinformation? The evidence indicates that no matter how vigorously and repeatedly we counter the misinformation, for example by repeating the correction over and over again, the influence remains detectable (Ecker, Lewandowsky, Swire, & Chang, 2011). The old saying got it right - mud sticks.

There is an additional complication. Not only is misinformation difficult to remove, debunking a myth can actually *strengthen* it in people's minds. Several different "backfire effects" have been observed:

- **Familiarity Backfire Effect:** When you repeat a myth within a debunking (e.g., on a flyer distributed by the [Centre for Disease Control on flu vaccinations](#): "Myth: The side effects are worse than the flu. Fact: the most likely side effect is a sore arm"), it makes people more familiar with the myth. This can lead to people, especially older adults, remembering the myth as true, especially after some delay (Skurnik, Yoon, Park, & Schwarz, 2005, Weaver, Garcia, Schwarz, & Miller, 2007).
- **Overkill Backfire Effect:** When a debunking includes too many arguments against the myth, this makes the debunking difficult to process and can lead to the myth being strengthened, simply

because its simplicity makes it more cognitively attractive by comparison to the correction (Schwarz, Sanna, Skurnik, & Yoon, 2007). People generally prefer simpler explanations over more complicated ones.

- **Worldview Backfire Effect:** If a debunking threatens one's world view or identity—for example, when strong supporters of President Bush are told that there were no 'Weapons of Mass Destruction' in Iraq—this can motivate people to double down on their beliefs, believing the myth stronger than ever (Nyhan & Reifler, 2010).

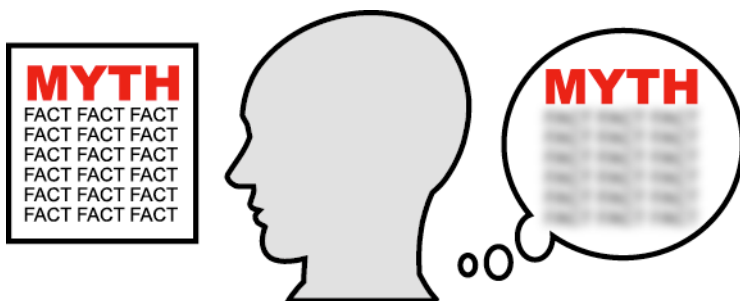
The last thing we want to do when debunking misinformation is blunder in and make matters worse. So this handbook has a specific focus - providing practical tips for effectively debunking misinformation and avoiding the various backfire effects. To achieve this, some understanding of the relevant cognitive processes is required and we explain some of the interesting psychological research in this area. We finish with some specific case studies as examples of effective rebuttals of common myths.

### The Familiarity Backfire Effect

Myth debunkers face a catch-22 situation. To debunk a myth, you must mention the myth (otherwise, how will people know what you're talking about?). This makes people more familiar with the myth. When people become more familiar with information, they're more likely to accept it as true. Does this mean debunking a myth might actually reinforce it in people's minds?

To test for this backfire effect, people were shown a flyer that debunked common myths about flu vaccines (Skurnik, Yoon, Park, & Schwarz, 2005). Afterwards, they were asked to separate the myths from the facts. When asked immediately after reading the flyer, people successfully identified the myths. However, when queried 30 minutes after reading the flyer, people had more difficulty correctly identifying the myths. Some people actually scored worse after reading the flyer than they did beforehand. The flyer had reinforced the myths in their minds.

So the backfire effect is real. The driving force is the fact that familiarity increases the chances of accepting information as true. When people read the flyer, they immediately remember the details that debunk the myth, successfully identify the myths and "tag" them as false. However, as time passes, the memory of the details has faded and all people remember is the myth without the "tag" that identifies it as false. This effect is particularly strong in older adults because their memories are more vulnerable to forgetting of details.

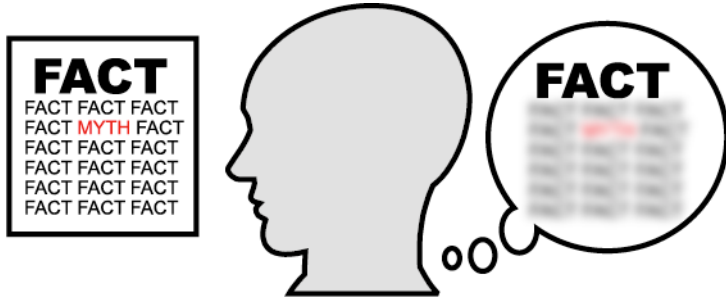


How does one avoid causing the Familiarity Backfire Effect? Ideally, avoid mentioning the myth altogether while correcting it.

***When seeking to counter misinformation, the best approach is to focus on the actual facts you wish to communicate.***

However, in many cases, not mentioning the myth is not a practical option. In this case, the emphasis of the debunking should be on the facts. The often seen technique of headlining your debunking with the

myth in big, bold letters is really the last thing you want to do. Instead, communicate your core fact in your headline. Your debunking should begin with emphasis on the facts, not the myth. Your goal is to increase people's familiarity with the facts.



### **The Overkill Backfire Effect**

You could file this one in the almost-too-obvious-to-mention box, but nevertheless, one principle that science communicators often fail to follow is making their content easy to process. This means easy to read, easy to understand and if possible, not too long. Information that is easy to process is likelier to be accepted as true (Schwarz, Sanna, Skurnik, & Yoon, 2007). For example, merely enhancing the color contrast of the printed font so it is easier to read can increase people's acceptance of the truth of a statement (Reber & Schwarz, 1999).

Common wisdom is that the more counter-arguments you provide, the more successful you'll be in debunking a myth. It turns out that the opposite can be true – when it comes to refuting misinformation, less can be more. Debunks that offered 3 arguments were successful in reducing the influence of misinformation while offering 12 arguments actually reinforced the myth (Schwarz, Sanna, Skurnik, & Yoon, 2007). Why? The Overkill Backfire Effect occurs because processing many arguments takes more effort than just considering a few.

### ***A simple myth is more cognitively attractive than an over-complicated correction.***

The solution is to keep your content lean, mean and easy to read. Encapsulating much information or complex concepts into short explanations is a difficult and time-consuming exercise. As Mark Twain famously wrote, "I didn't have time to write a short letter, so I wrote a long one instead". However, the effort is both worthwhile and essential. One of the most significant comments on science communication comes from Nobel Prize winning physicist Ernst Rutherford, who said "If you can't explain your physics to a barmaid it is probably not very good physics" (Jones 1998).

Ensuring your content is easy to process means utilising every tool available to ensure your material is easy to process. Make sure you use simple language, short sentences, subheadings and paragraphs. Avoid dramatic language and derogatory comments that could put people off. Stick to the facts. End on a strong and simple message that people will remember and tweet to their friends, such as "97 out of 100 climate scientists believe that humans are causing global warning". or "Study proves that MMR vaccines are safe." Use graphics wherever possible to illustrate your points (more on this later).

Scientists have long followed the principles of the Information Deficit Model, which suggests that people hold erroneous views because they don't have all the information at hand. According to the Information Deficit Model, the solution is more information. However, too much information can backfire. This means adhering to the KISS principle long used by communicators: Keep It Simple, Stupid!

## The Worldview Backfire Effect

The third and arguably most potent backfire effect occurs when discussing emotionally charged topics that tie in with people's worldviews and sense of cultural identity. We all have a basic set of cultural values that we bring to bear in all aspects of daily life—for example, we all have an expectation of fairness in life, and we share a common belief that the world is a “just” place in which everyone more or less gets what they deserve. On other issues, however, we may differ: for example, some of us think that income inequality is a natural and indeed desirable consequence of the fact that some people choose to work particularly hard. Others reject that notion and think of equality as a more desirable state of society. When those basic cultural cognitions or worldviews are challenged, the consequences can be far-reaching.

There are a number of cognitive processes that cause people to unconsciously process information in a biased fashion. For those who are strongly fixed in their views, encountering counter-arguments can cause them to double down on their views with even greater strength.

One cognitive process that contributes to this effect is **Confirmation Bias**, where people selectively seek out information that bolsters their view. In one experiment, people were offered information on hot-button issues like gun control or affirmative action. Each parcel of information was labelled by source, clearly indicating whether the information would be pro or con (eg – the National Rifle Association or Citizens Against Handguns). Although instructed to be even-handed, people opted for sources that matched their pre-existing views. The result was that despite presenting people with a balanced set of facts, they reinforced their pre-existing views by gravitating towards the information they already agreed with. The polarisation was greatest among those with strongly held views (Taber & Lodge, 2006).

This pick-and-choose approach is made easier in today's society by the fragmentation of the media, allowing people to select like-minded sources such as cable TV channels (Kull, Ramsay, & Lewis, 2003). Similarly, the Internet allows people to congregate on blogs and online communities of similar ilk (Johnson, Bichard, & Zhang, 2009). This has been identified as a contributor to the polarization of views in society (Stroud, 2010).

However, what if one is presented only with arguments that run counter to one's worldview? In this case, the cognitive process that comes to the fore is **Disconfirmation Bias**, the flip side of confirmation bias. This is where people spend significantly more time and thought actively arguing against opposing arguments (Nyhan & Reifler 2010).

This was demonstrated in an experiment where Republicans who believed Saddam Hussain was linked to the 9/11 terrorist attacks were provided with evidence that there was no link between the two, including a direct quote from George Bush (Prasad, et al., 2009). Only 2% of participants changed their mind (although interestingly, 14% denied that they believed the link in the first place). The vast majority clung to the link between Iraq and 9/11, employing a range of arguments to brush aside the evidence. The most common response was attitude bolstering – bringing supporting facts to mind while ignoring any contrary facts. The process of bringing to the fore supporting facts resulted in the *strengthening* of people's erroneous belief.

If facts cannot dissuade a person from their pre-existing beliefs, and indeed can sometimes make things worse, how can we possibly reduce the effect of misinformation? There are two sources of hope. The first is to note that the Worldview Backfire Effect is strongest among those who are already fixed in their views. Among those who are not as firmly decided about hot-button issues, there is a greater chance that misinformation can be corrected. This suggests that messaging strategies should be directed towards the undecided majority rather than the unswayable minority.

Secondly, there are ways that messages can be presented without encountering the usual psychological resistance. For example, when worldview-threatening messages are coupled with so-called self-affirmation, people become more balanced in considering pro and con information (Cohen, Sherman, Bastardi, Hsu, & McGoey, 2007, Nyhan & Reifler, 2011).

Self-affirmation can be achieved by asking people to write a few sentences about a time when they felt really good about themselves because they acted on a value that was important to them. In comparison to people who received no affirmation, those who were self-affirmed became more receptive to messages that otherwise might have threatened their worldviews. Interestingly, the “self-affirmation effect” is strongest among those whose ideology was central to their sense of self-worth. Self-affirmation gives the facts a fighting chance. Another way in which information can be made more acceptable is by framing it in a manner that is less threatening to a person’s worldview. For example, Republicans are far more likely to accept an otherwise identical charge as a “carbon offset” than as a “tax”, whereas the wording has little effect on Democrats or Independents—because their values are not challenged by the word “tax” (Hardisy, Johnson, & Weber, 2010).

### **Elements of an effective debunking**

Assuming you successfully negotiate around the myriad of backfire effects, what is the most effective way to debunk a myth? The challenge is that once misinformation gets into a person’s mind, it’s very difficult to remove. This is even the case when people remember and *accept* a correction.

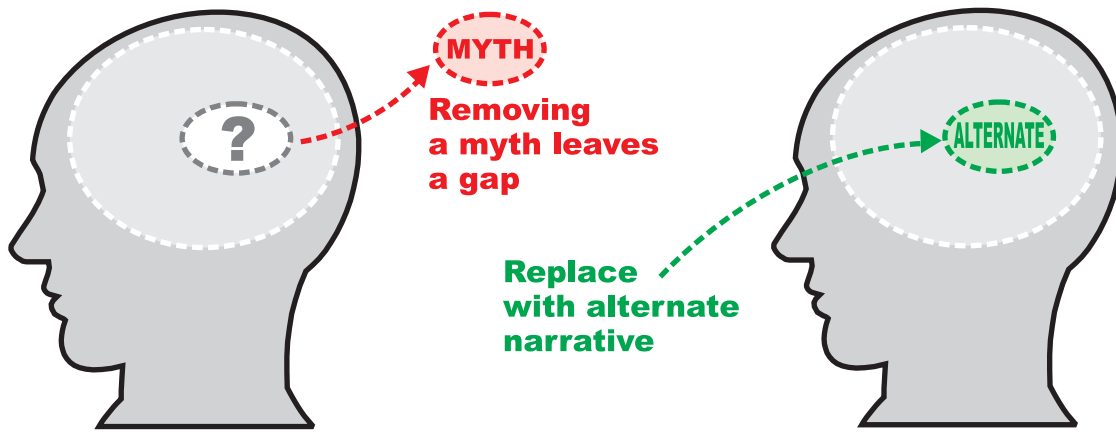
This was demonstrated in an experiment where people read a fictitious account of a warehouse fire (Seifert, 2002, Wilkes & Leatherbarrow 1988, Johnson & Seifert 1994). Mention was made of paint and gas cans along with explosions. Later in the story, it was clarified that the information about the paint and cans had been false and that those items were not present at the fire. Even when people remembered and accepted this correction, they still cited the paint or cans when asked questions about the fire. For example, when asked “why do you think there was so much smoke?”, people would routinely invoke the oil paint—that is, the very item they had just acknowledged was not present. When people hear misinformation, they build a mental model, with the myth providing an explanation. When you debunk the myth, you leave a gap in their mental model—and no way in which the presence of smoke could be easily explained. To deal with this problem, people prefer an incorrect model over an incomplete model, and in the absence of a better explanation, they opt for the wrong explanation (Ecker et al 2011).

It is perhaps not surprising, therefore, that the most effective way of discounting misinformation is by providing an **alternative explanation**. In the warehouse fire experiment, when an alternative explanation involving lighter fluid and stationery was provided, people were less likely to cite the paint and gas cans when queried about the fire.

***So the most effective way to reduce the effect of misinformation is to provide an alternative explanation for the events covered by the misinformation.***

This is illustrated particularly clearly in fictional murder trials. Accusing an alternative suspect greatly reduced the number of guilty verdicts from participants who acted as jurors, compared to defences that merely explained why the defendant wasn’t guilty (Tenney, Cleary, & Spellman, 2009).

The best alternative must be plausible and explain all observed features of the event (Rapp & Kendeou 2007, Seifert 2002). When you debunk a myth, you create a gap in the person’s mind. To be effective, your debunking must fill that gap.



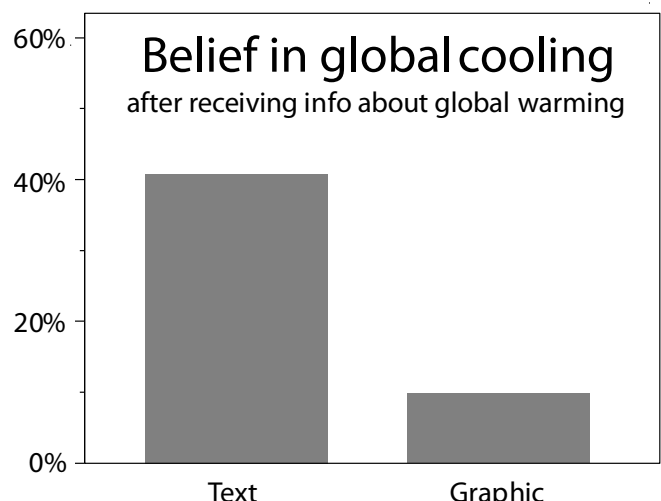
When a myth is exposed as wrong, one gap that may require filling is explaining *how* the myth is wrong in the first place. This can be achieved by exposing the rhetorical techniques used to misinform. A handy reference of rhetorical techniques common to many movements that deny a scientific consensus are found in [Denialism: what is it and how should scientists respond?](#) (Diethelm & McKee, 2009). The techniques include cherry picking, conspiracy theories and the raising of fake experts.

Another aspect of an alternative narrative might be to explain why the misinformer promoted the myth in the first place. For example, in experiments involving participants serving as jurors, people are more likely to dismiss tainted evidence if they are made suspicious of the motives underlying its introduction in the first place (e.g., “the prosecutor needs a conviction to get promoted”). Without inducing suspicion, instructions to disregard tainted evidence remain ineffective even though people claim to dismiss the evidence—that is, people earnestly believe they are not being influenced by tainted evidence, although their actual guilty verdicts show otherwise (e.g., Fein, McCloskey, & Tomlinson, 1997).

Suspicion about the source of misinformation has also been shown to reduce the influence of misinformation in “real life” (Lewandowsky, Stritzke, Oberauer, & Morales, 2005, 2009). In those studies, people who were skeptical of the motives underlying the Iraq War in 2003—i.e., those who rejected ‘Weapons of Mass Destruction’ as the true motive—were better able to differentiate between true and false information in the media. Skeptics rejected information that had been revealed to be incorrect, and they endorsed true information more accurately than people who were less skeptical.

Another key feature to effective rebuttal is an explicit warning prior to mentioning the myth. Experimentation with different rebuttal structures found the most effective combination included an alternative explanation and an explicit warning at the outset (Ecker, Lewandowsky, & Tang, 2011). In practice, it is not always possible to precede erroneous information with a warning; however, a generic warning can be created by pointing to the prior record of an information source—for example, by contrasting the accuracy of Fox News to that of National Public Radio in the U.S. (Kull, et al., 2003).

Graphics are also an important part of the debunker’s toolbox and have been shown to be significantly more effective in reducing misconceptions than text. When people read a refutation that conflicts with their pre-existing beliefs, they search for counter arguments and seize on ambiguities in order to construct an alternative interpretation. Graphics provide more clarity and less opportunity for misinterpretation. When self-identified Republicans were surveyed about their global warming beliefs, a significantly smaller



number thought global cooling was occurring when shown a graph of temperature trends compared to those who were given a textual description (Nyhan & Reifler, 2011). A survey also found that when shown data points representing surface temperature, people judged the warming trend correctly irrespective of their views towards global warming (Lewandowsky, 2011). If your content can be expressed visually, always opt for a graphic in your debunking.

## Anatomy of an effective debunking

Bringing all the different threads together, an effective debunking requires:

- **Core facts** – a refutation should emphasise the facts, not the myth. The presentation of facts should be restricted to key facts to avoid an Overkill Backfire Effect.
- **Explicit warning** – prior to any mention of the myth, text or visual cues should warn that the upcoming information is false.
- **Alternative explanation** – any gaps left by the debunking need to be filled. This may be achieved by providing an alternative causal explanation, how the myth is wrong or why the misinformers promoted the myth in the first place.
- **Graphs** – if the core facts can be displayed graphically, take advantage of this option.

### Sun and climate are going in opposite directions

Over the last 30 years of global warming, the sun has shown a slight cooling trend. Sun and climate are going in opposite directions. This has led a number of scientists independently concluding that the sun cannot be the cause of recent global warming.

• Core Facts

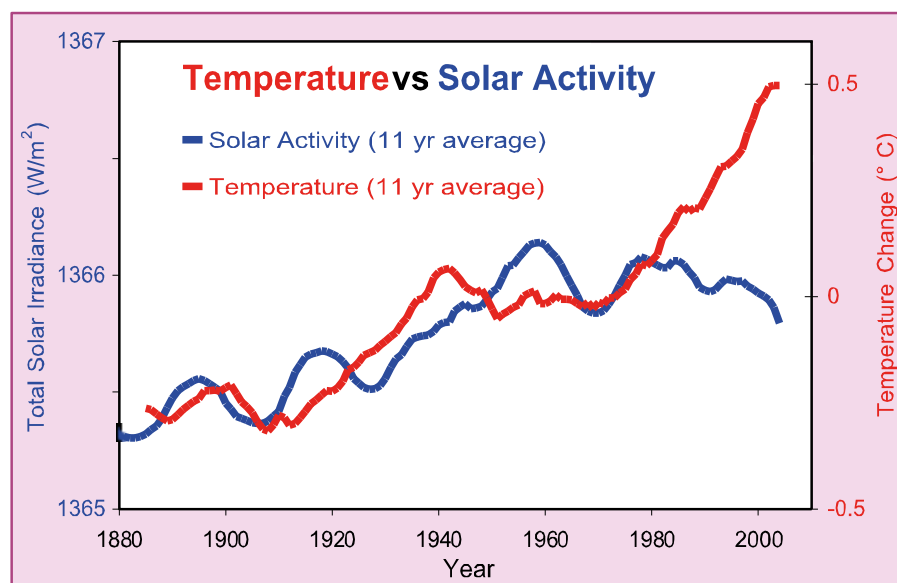
However, climate denier arguments seek to mislead the public and their claims about the sun are no exception. One of the most common and persistent climate myths is that the sun is the cause.

• Explicit Warning

• Myth

Climate deniers argue this by cherry picking the data - showing past periods when sun and climate move together but ignoring the last few decades when the two diverge.

• Alternate Explanation



• Graph

## Bibliography

- Cohen, G. L., Sherman, D. K., Bastardi, A., Hsu, L., & McGoey, M. (2007). Bridging the Partisan Divide: Self-Affirmation Reduces Ideological Closed-Mindedness and Inflexibility in Negotiation. *Personality & Soc. Psych.*, 93 (3), 415-430.
- Diethelm, P., & McKee, M. (2009). Denialism: what is it and how should scientists respond? . *Eur J Public Health* , 19 (1), 2-4.
- Ecker, U. K., Lewandowsky, S., & Tang, D. T. (2011). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition* , 38, 1087-1100.
- Ecker, U. K., Lewandowsky, S., Swire, B., & Chang, D. (2011). Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction. *PSYCHONOMIC BULLETIN & REVIEW* , 18 (3), 570-578.
- Fein, S., McCloskey, A. L., & Tomlinson, T. M. (1997). Can the jury disregard that information? The use of suspicion to reduce the prejudicial effects of pretrial publicity and inadmissible testimony. *Personality & Social Psychology Bulletin*, 23, 1215-1226.
- Hamilton, L. (2011). Education, politics and opinions about climate change: Evidence for interaction effects. *Climatic Change* , 104, 231-242.
- Hardisty, D. J., Johnson, E. J., & Weber, E. U. (2010). A Dirty Word or a Dirty World? Attribute Framing, Political Affiliation, and Query Theory. *Psychological Science*, 21, 86-92.
- Jacques, P. J., & Dunlap, R. E. (2008). The organisation of denial: Conservative think tanks and environmental skepticism. *Environmental Politics* , 17, 349-385.
- Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When discredited information in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition* , 20 (6), 1420-1436.
- 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.
- Jones, J.P. (1998). Point of View: STAS and BehaviorScan--Yet Another View. *Journal of Advertising Research*, 38, 51-53.
- Jou, J., & Foreman, J. (2007). Transfer of learning in avoiding false memory: The roles of warning, immediate feedback, and incentive. *Quarterly Journal of Experimental Psychology* , 60, 877-896.
- Kahan, D. (2010). Fixing the communications failure. *Nature* , 463, 296-297.
- Kull, S. R., Ramsay, C., & Lewis, E. (2003). Misperceptions, the media, and the Iraq war. *Political Science Quarterly* , 118, 569-598.
- Lewandowsky, S. (2011). Popular Consensus: Climate Change Is Set to Continue. *Psychological Science*, 22, 460-463.
- 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. *Terrorism and torture: An interdisciplinary perspective* , 179-203.
- Munro, G. D. (2010). (2010). The Scientific Impotence Excuse: Discounting Belief-Threatening Scientific Abstracts. *Journal of Applied Social Psychology* , 40, 579-600.
- Nyhan, B., & Reifler, J. (2011). Opening the Political Mind? The effects of self-affirmation and graphical information on factual misperceptions .



- Nyhan, B., & Reifler, J. (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior*, 32 (2), 303-330.
- Oreskes, N., & Conway, E. M. (2010). *Merchants of doubt*. Bloomsbury Publishing.
- Prasad, M., Perrin, A. J., Bezila, K., Hoffman, S. G., Kindleberger, K., Manturuk, K., et al. (2009). "There Must Be a Reason": Osama, Saddam, and Inferred Justification. *Sociological Inquiry*, 79, 142-162.
- Rapp, D. N., & Kendeou, P. (2007). Revising what readers know: Updating text representations during narrative comprehension. *Memory & Cognition*, 35, 2019-2032.
- Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, 8, 338-342.
- Schwarz, N., Sanna, L., 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.
- Seifert, C. M. (2002). The continued influence of misinformation in memory: What makes a correction effective? *The Psychology of Learning and Motivation*, 41, 265-292.
- Skurnik, I., Yoon, C., Park, D., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research*, 31, 713-724.
- Song, S., & Schwarz, N. (2010). If it's easy to read, it's easy to do, pretty, good, and true: fluency effects on judgment, choice, and processing style. *The Psychologist*.
- Stroud, N. J. (2010). Polarization and Partisan Selective Exposure. *Journal of Communication*, 60, 556-576.
- Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50, 755-69.
- Tenney, E. R., Cleary, H. M., & Spellman, B. A. (2009). Unpacking the doubt in "Beyond a reasonable doubt:" Plausible alternative stories increase not guilty verdicts. *Basic and Applied Social Psychology*, 31, 1-8.
- Weaver, K., Garcia, S. M., Schwarz, N., & Miller, D. T. (2007). Inferring the popularity of an opinion from its familiarity: A repetitive voice sounds like a chorus. *Journal of Personality and Social Psychology*, 92, 821-833.
- Wolfe, C. R., & Britt, M. A. (2008). Locus of the myside bias in written argumentation. *Thinking & Reasoning*, 14, 1-27.