A reasoned approach to dealing with fake news

M. Anne Britt, Jean-François Rouet, Dylan Blaum and Keith Millis

To appear in: Policy Insights from the Behavioral and Brain Sciences


Abstract

We now have almost no filters on information that we can access, and this requires a much more vigilant, knowledgeable reader. Learning false information from the web can have dire consequences for personal, social and professional decision-making. Given how our memory works and the biases in selecting and interpreting information, it is more important than ever that we control our own cognitive and affective processing. For example, simple repetition of information can lead to increased confidence in its perceived truth, initial incorrect information remains available and can continue to have an effect despite learning the corrected information, and we are more likely to accept information that is consistent with our beliefs. Information evaluation requires readers to set and monitor for goals of accuracy, coherence and completeness, to employ strategies to achieve these goals, and value the time and effort-consuming this systematic evaluation. We present several recommendations to support a reasoned approach to fake news and manipulation.

Tweet: Supporting a reasoned approach to addressing fake news.
A reasoned approach

**Highlights:**

- Our interpretation of information is guided by our goals and prior beliefs.
- Recalling information can change our memory and beliefs (false memory effect) while information we find out was incorrect remains in memory and can continue to affect us (continued influence effect).
- Information repetition increases belief in its truth (Illusory truth effect) and we are more likely to accept belief consistent information (belief consistency) without seeking out disconfirming information (confirmation bias).
- We need to teach the public these biases and how to counteract them by setting goals to evaluate information for accuracy, relevance, and sufficiency.
- We need to increasing guidelines to protect the public and support for rigorous interdisciplinary research.

**Declaration of Conflicting Interests**

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The authors received no financial support for the research, authorship, and/or publication of this article.
A reasoned approach to dealing with fake news

On Dec 4, 2016, Edgar Maddison Welch fired an AR-15 assault rifle inside the Comet Ping Pong pizza restaurant in Washington D.C., while on a self-appointed mission to save children being held captive in the basement as part of sex-trafficking ring. According to stories circulating on the web at that time, Hillary Clinton was the head of the satanic sex ring using the restaurant. Welch was arrested, and no one was killed or injured, but how did he come to believe these conspiracy stories enough to decide to "self-investigate" with an AR-15 rifle? In a Dec 7, 2016 interview with the New York Times, Welch explained that there were many different sources on the internet leaving him with the “impression something nefarious was happening” (Goldman, 2016). He also routinely listened to Alex Jones, an Internet entertainer who promotes conspiracy theories. Welch said that his intel was not 100% but his goal was to “self-investigate” the situation.

In fact, the so-called “Pizza Gate” conspiracy spread quickly on the internet immediately before the November 2016 U.S. election. On October 30, a Facebook account using a fake identity posted that the NYPD suggested that emails found two days earlier on a laptop owned by a Clinton aide point to a Hillary Clinton pedophilia ring. This was quickly reposted thousands of times. On November 4th, Alex Jones (of the former InfoWars Web site) pronounced on one his programs: “When I think about all the children Hillary Clinton has personally murdered and chopped up and raped, I have zero fear standing up against her,” [...] “Yeah, you heard me right. Hillary Clinton has personally murdered children. I just can’t hold back the truth anymore” (Fisher, Cox, & Hermann, 2016). Within a few days, the news snowballed as #pizzagate Twitter topic received millions of retweets, many from foreign accounts (e.g., in the Czech Republic) and bots, and became “trending” news on Facebook (based on frequency algorithms alone). Also
helping to spread the story were “citizen journalists”. According to a Washington Post interview (Fisher et al., 2016) with one: “MacWilliams calls herself a journalist, but she does not try to be ‘100 percent accurate,’ either. She believes the beauty of the Internet is that people can crowdsource the truth. Eventually, what is real will emerge, she said.” The article concluded “Pizzagate — [...] is possible only because science has produced the most powerful tools ever invented to find and disseminate information.”

Americans acknowledge both the positive and negatives aspects of the Internet (Pew, 2018). Clearly the internet is helpful in gaining access to news, connecting people, and learning new skills. Negatives include fake news and misinformation, and privacy issues. The pizzagate incident points to several important challenges. In this new world, there are few if any gatekeepers for “truth”— The spread of information can be driven by many forces other than usefulness and reliability; for instance, the need to attract viewers to drive advertising revenue, fame, the sense of being part of something larger than ourselves. And in the end, it can have serious consequences beyond the intentions of the original source. Now more than ever, members of our society need to become aware of their own cognitive biases and how to avoid being exploited because of them.

**Overview of cognitive biases**

One common source of bias is the human memory system. Research shows that memory does not work like a video recording of our life. Although we may briefly have a *verbatim* representation of the exact words of a text, these details are quickly lost without attention-demanding retention strategies, such as rehearsal (Sachs, 1967). We use our verbatim representation to create an interpretation of the author’s message by making connections to our prior knowledge and beliefs (Kintsch, 1998). In addition to representing the content, we can also
A reasoned approach

represent the source of the communication (e.g. Alex Jones of InfoWars) and associate content with a source and even identify relationships among sources (e.g., Alex Jones and @DavidGoldbergNY agree; while Reddit owners disagree) (Perfetti, Rouet, & Britt, 1999). In this way, we can represent otherwise inconsistent information across texts or webpages.

The nature and extent of this processing will depend on why a person is reading. People read information on the Web for a variety of reasons including: relieving boredom, keeping up-to-date on something, looking up something in order to do some task. In general, we read for a purpose or goal (Britt, Rouet & Durik, 2018), and these goals can impact what we read (e.g., CNN, FOX, Facebook posts), how we read (e.g., read only headlines and leads, read deeply, skim, skip until locate a piece of information), how long we read, etc. We do not simply represent only what the text says. We interpret the text, guided by our goals and based on prior knowledge as well as our attitudes and beliefs.

One set of biases originate with the quick loss of verbatim memory and potential over-reliance on an interpreted representation of the communication. For example, in one study, college students who were asked to read short, two-sentence arguments (e.g., “The U.S. is right to intervene in other countries’ affairs because local events can catastrophically impact the entire world.”), were only about 75% accurate at recalling the claim of the argument immediately after judging their agreement with it (Britt, Kurby, Dandotkar & Wolfe, 2008). Furthermore, the errors they made most often changed the meaning of the claim, although it may seem minor to the untrained eye (e.g., “The U.S. should intervene in other countries’ affairs”).

The problem of relying on an interpreted representation is illustrated in a well-known series of studies showing false memories can be created based on the situation at retrieval (false memory effect). Two groups of people watched a video of a car accident and one group was
asked how fast the car was going when it “smashed” into the other car. The other was asked how fast it was going when it “bumped” into another car (Loftus, 1979). People falsely remembered the car going faster in the first case and they were very certain that they had an accurate memory. This false memory effect can be reduced (but not eliminated) when a “false memory warning” is given prior to encoding (Roediger, 1996). But, when the warning is given *after* encoding, people are still likely to “remember” the false information and use it as if were true (Roediger, 1996). The effect is amplified the more often we encounter and retrieve the information (Roediger, Jacoby, & McDermott, 1996).

People are also highly affected by the mere repetition of information. Researchers have found that statements presented several times across occasions, led to more confidence in the truth of those statements as compared to statements that were not repeated (*Illusory truth effect*) (Polage, 2012). A related effect, *Availability Cascade effect*, shows how this Illusory truth effect can snowball at a more global level. Kuran and Sunstein (1999) presented an interesting analysis of the public attention and attitude toward three events (waste dumps in the Love Canal, the use of Alar chemical, and the TWA 800 crash) to show how repeated statements of beliefs can make claims seem truer and can be used to manipulate public opinion. It would seem inevitable that in an information environment such as the internet, social media algorithms that identify and highlight “trending” stories and source filters that restrict alternative information would produce an Availability Cascade.

Finally, *our memory system does not handle new discrepant information by simply replacing old information*. That is, initially faulty or wrong information remains available and can continue to have an effect despite encoding new, correct information. This is called the *continued influence effect* (Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012). For
example, when people read a story about a warehouse fire that was initially reported to be caused by oil paint cans but later discounted, readers continued to think that the fire was caused by the paint cans (Johnson & Seifert, 1994). If you think of human memory like a collection of assertions networked together by associative links of varying strength, updating information is more akin to creating a new link that tags the information as wrong but doesn’t remove it. Nevertheless, such a tag might be limited in scope, applying to one specific assertion but not to all the other information in memory connected to that assertion. In this case, the updated information will not necessarily be stored with the later-found-errorneous information without systematic updating all of the information learned. This continued influence of misinformation effect is difficult to overcome (Lewandowsky et al., 2012) and can even lead to back-fire effects as we discuss below.

In addition to these memory related problems, there are also several biases related to those assertions that we would categorize as beliefs or attitudes. In general, these biases affect what information we choose to attend to and how we process it, usually in the direction of maintaining the consistency of our attitudes and beliefs. Our beliefs can lead to biased search, selection and interpretation thereby reducing exposure to information that is not consistent with what we already believe.

Research has shown that we are more willing to accept consistent information and judge the quality of arguments more positively when they are consistent with our beliefs, as compared to information that opposes our beliefs (belief bias) (Lord, Ross, & Lepper, 1979). The other half of the belief bias effect is that we are also less likely to accept information that goes against our beliefs and judge the arguments as being of a lower logical quality (Edwards & Smith, 1996). We are more likely to scrutinize or attempt to find holes in the information that contradicts our
beliefs. In fact, in some cases, when we encounter disconfirming evidence, it may actually strengthen our evidence-opposing previous beliefs (*Back-fire effect*) or reject information because of the source it came from. All of these biases are exacerbated by our tendency to believe we are more unbiased reasoners than others (*Bias blind spot*).

We also have a strong tendency to seek out and use information that is in line with what we already believe to be true (*Confirmation bias*) (Nickerson, 1998); that is, we seek out information that confirms our beliefs rather than seeking information that could potentially disconfirm them. We also have a *myside bias* toward generating belief-consistent reasons (Baron, 1995) and ignore evidence counter to our own side when performing tasks such as writing an argumentative essay (Wolfe & Britt, 2008). Finally, beliefs even affect comprehension: readers spend longer rereading text that is contrary to their beliefs (Maier, Richter, & Britt, 2018) but have better memory for information that is consistent with them (Maier & Richter, 2013).

These biases are challenging to overcome because the internet is being designed to create “filter bubbles” which as Bill Gates notes, let “you go off with like-minded people, so you’re not mixing and sharing and understanding other points of view. It’s super important. It’s turned out to be more of a problem than I, or many others, would have expected” (Delaney, 2017).

**Challenges for the reader**

In addition to memory biases, people exhibit biases in how they reason about information. Our focus on reading is warranted by the fact that most of the information on the Internet is still conveyed through written texts (admittedly often in combination with speech and pictures), and that most adults have only a limited ability to understand and reflect critically on what they read (OECD, 2013). In this next section, we discuss how people read and evaluate the
A reasoned approach

support encounter for persuasive claims. The three characteristics for support include: assessing whether the support is accurate, coherent and complete (Blair & Johnson, 1987).

**Assessing the Accuracy of Information.**

People typically assume what others say is truthful and accurate unless there is reason for doubt (Grice, 1975). However, our cognitive biases work against this accuracy evaluation especially when we rely on our memory. What we think is true can be affected by what occurs in the retrieval situation because that can change our memory without undermining our confidence in the accuracy of that memory (false memory). When information should be updated because of an update or correction, the old information does not get erased or overwritten. Therefore, we may continue to believe things that have been revised or corrected and this is especially true for information that was highly elaborated before the update. Also, simple repetition (without additional support) can increase the believability of information (Illusory truth) and this can snowball (availability cascade). That repetition can even appear as corroboration even though the information is not repeated by independent sources. Our prior beliefs can also interfere with our evaluation of the accuracy evaluation. We are more willing to belief new information that is consistent with our views (belief bias).

Accurate evaluation requires readers to set and monitor accuracy as a goal, employ strategies to achieve that goal, and value the time and effort-consuming systematic evaluation. Often when reading on the Web, one has many competing goals and obtaining and verifying that the information is accurate is only one goal. Monitoring for accuracy may not have been a well-practiced cognitive activity in the past but it must become one today. Now more than ever, readers have to assess the accuracy of information which can range from clearly true or false to indeterminable. Indeed, the amount of falsehood and deception on the Web has given rise to a
plethora of fact-checking sites (e.g., https://www.snopes.com/, http://www.politifact.com/, and https://www.factcheck.org/) and even Google and Facebooks had begun efforts to label suspect stories.

Strategies for evaluating accuracy can target the content itself. Readers can actively question whether they find the information believable, whether it makes sense or is consistent with prior knowledge (Richter, Schroeder, & Wöhrmann, 2009). We have to monitor and protect against our belief bias because related beliefs are quickly available (Voss, Fincher-Kiefer, Wiley, & Silfies, 1993). This takes a deliberate effort of setting a goal to acquire accurate information. The content can also be evaluated for accuracy based on an assessment of the support provided. Some types of evidence are better (e.g., experiments, quotes, reasoning) than others (e.g., number of retweets). The evaluation of what makes “good” support, however, depends on one’s epistemic beliefs about what knowledge is (e.g., tentative and complex) and beliefs about how we get knowledge (e.g., gut feelings vs authority vs use of rules of inquiry) (Bråten, Britt, Strømsø, & Rouet, 2011) and other knowledge, skills and dispositions that lay readers often lack. Even undergraduate college students are not very skilled in detecting problems in descriptions of scientific studies (Kopp, Britt, Millis, & Graesser, 2012) and do not spontaneously set an explicit goal to obtain accurate information when reading search-result listings of global warming sources (Kopp, 2013). Instead they are “looking for support”, though not necessarily accurate support.

Readers can also evaluate features of the source of the information. Given that we cannot erase or overwrite memory, evaluating the source information would be most beneficial prior to encoding the content so that the reader can decide whether the source is reliable and knowledgeable enough to continue reading.
To illustrate, searching a news aggregator site for “Betsy DeVos” returned links to very different sites (e.g., “The Washington Post”, “Politico” and “Reason.com”) that vary in terms of their features of the author (e.g., author’s level of knowledge, motives or bias) and outlet (e.g., criteria for accuracy checking, checking tools employed). This is the easy case if one sets the goal to encode and evaluation the source (which is not often a reading goal). The reader also has to be able to understand that when a site uses a headline like “DeVos: Civil rights office will return to being a neutral agency”, it is stating that DeVos is saying “Civil rights office will return to being a neutral agency”. DeVos becomes the source of the statement and the reader’s knowledge of and evaluation of DeVos can be used to interpret it. In this case, the source precedes the content which helps. But in many cases the source of a statement comes after the content has been encoded, which is a problem because it is now in our memory before we know whether we trust the source. An even more challenging example is when all that is stated to entice a click is “Betsy DeVos isn’t ‘Enabling rape deniers’ by pushing for due process on college…” Without clicking it is not possible to know who is saying that Betsy DeVos is “enabling rape deniers”. This is challenging, since reading headlines without clicking to read the details is very common—leaving the reader with a memory of an assertion without being able to evaluate whether the source was knowledgeable or biased.

In the case of a news aggregator, some source information is actually presented. Many times, however, there is no information about key features of the author or outlet responsible to distribution. Even worse, the source information can actually be deceptive. For example, “ABCnews.com.co” or “Breaking-CNN.com” are cites which mimic the look, feel and logo of actual news sites but may present erroneous information and malware. When we talk with people
in real life, their “sourceness” is obvious. On the internet, the author as an entity can be obscured or even falsified (Britt, Rouet, & Braasch, 2013).

Finally, readers can look to corroborate information across independent sources. However, it can be challenging to find truly independent sources. For example, it may be that only 1 source came to the conclusion that 2-3 million illegal immigrants voted in the U.S. election, but one could have encountered it on Facebook, in discussion with friends and family, and on the television news. This could appear as corroboration, but it is not, because the information originally came from a single source.

**Assessing the relevance of support and search results.** Relevance of information can be evaluated according to goals for search (e.g., does this page, article or even paragraph give me the information that I am looking for?) or coherence (e.g., Does the support provided for their claim actual provide support? or Does the effect follow from the cause?).

For brevity sake, we will focus on claim-support relevance judgements. Of course, persuasion techniques, propaganda, and logical fallacies (e.g., ad hominem, emotional appeals) are very prevalent on the Web and can be persuasive, but they are beyond the scope of this article. Here we focus on an appeal to reason.

Again, our cognitive system can work against the use of accurate reasoning. Relying on an interpreted representation makes it difficult to accurately evaluate whether the support is relevant for the specific claim made (we discuss this in more detail below). Indeed, readers routinely set goals to identify support but not necessarily “good” support and this may lead to accepting all support as “good” without much evaluation. As a result, a wider range of support may be seen as acceptable. Our beliefs can also interfere. We spend longer reading inconsistent information to scrutinize it and often judge those arguments as of lower structural quality.
Whereas, we read belief-consistent information more quickly and typically judge it to be of higher structural quality. It is difficult to judge the quality on an argument independent of one’s acceptance of its truth.

Even with arguments for which one doesn’t have strong beliefs, evaluation can be challenging. College students have been found to experience difficulty evaluating arguments that are structurally flawed (i.e., the reason failed to support the claim or there was no reason presented) as compared with those that are structurally acceptable (i.e., the reason supported the claim) (Larson, Britt, & Kurby, 2009). It is even more challenging to evaluate argument elements that are spread across time or an expanse of text. Doing so requires one to hold in working memory the precise claim while trying to detect the support, sometimes needing to go back and reread the precise claim. This type of monitoring takes effort. Readers have difficulty noticing discrepancies in even short texts (Otero & Kintsch, 1992; Steffens, 2016), although they are slightly better when the discrepant statements come from different sources (Braasch, Rouet, Vibert, & Britt, 2012). Thus, it can be quite challenging to monitor and evaluate the logical claim-reason connection for lay readers.

**Assessing the sufficiency of support and perspectives.** A final criterion for argument support is whether it is sufficient, i.e., complete to some degree and capable of justifying the claim. Professional writers and news reports typically consider more than a single perspective (Wolfe & Britt, 2008) and it is the hallmark of academic writing and the scientific method. For the rest of us, comparing perspectives can help us understand the broader situation and make an informed decision.

Biases are most prevalent for this criterion of support. Studies have shown that we are less likely to believe things that are inconsistent with our views (see belief bias above). We also
have poorer memory for belief-inconsistent information as well as that which is consistent.
College students tend to ignore other-side information when composing argumentative essays (Nussbaum & Kardash, 2005; Wolfe, Britt, & Butler, 2009) and they often do not make alternative-based objections when evaluating arguments without task supports (Shaw, 1996).
While readers often do not seek out alternative perspectives or disconfirming evidence, when it is made available, they do read some of it but generally do not include it in their arguments (Wolfe et al., 2009).

As with the other two evaluations, working against the myside bias and confirmation bias takes effort and requires that readers set specific goals to seek out, comprehend, and consider other-side information. In its simplest form, it means trying to find articles on the web that argue for the other side of the claim. It can also mean trying to find and account for all available evidence regardless of side of the claim it supports. It can also mean taking into account multiple sides or perspectives and therefore include or address the “alternative viewpoints” or even camps (e.g., fiscal conservative, liberals, social conservatives, and moderates). This is challenging when only one side is presented because it requires the search for alternative perspectives without knowing what those perspectives are or how to find out about them. It is also challenging with multiple texts (for instance a series of chat messages or Google page lists) because it requires reading more, comparing information, and possibly dealing with the emotions that come from reading perspectives outside one’s own. It is not something even college students take on readily (Kopp, 2013).

Policy Recommendations
Although we have described several ways that cognitive biases leave people susceptible to misinformation and exploitation, there are steps that policy makers can take to mitigate the problem. We present some thoughts below.

1. **Epistemic vigilance needs to be taught and not just in school.**

   In the world of Infowars, the public needs to acquire the intellectual skills needed to critically assess the accuracy, soundness and sufficiency of information, in other words: epistemic vigilance. This review points to the insufficiency of teaching simply rules (e.g., use “.gov” sites, look for an M.D.). Rules do not help because the principle behind the rule is lost and technology advances occur too quickly for heuristics to be useful. Instead, we need to teach the public how our memory and comprehension processes work and how our biases can lead us to fall prey to manipulators. Ideally, a citizen should be aware that conclusions from reading a post may be affected by the post’s wording, context, his or her prior beliefs, and the presence of similar posts. We also have to teach the public how to evaluate information for accuracy, relevance/soundness, and sufficiency. This includes understanding what constitutes reason-based argumentation and how that differs from persuasion and other forms of belief change. We also have to teach the importance of applying this knowledge especially when emotions run high and when the information relates to their beliefs.

   In schools, we need interdisciplinary experts (e.g., journalists, information-science experts, teachers, cognitive scientists) to work together to develop curricula that can support the development of transferable knowledge. We also need professional development to support teachers in implementing effective instructional approaches to epistemic vigilance.

   But teaching epistemic vigilance should not be limited to school. Bad actors throughout the world recognize that social media is an effective way to influence the fabric of our
A reasoned approach

democratic society (e.g., elections, beliefs, harmful actions). And they are correct. The problem is too important to be limited to a few lectures or activities during school hours. It is even possible that Welch would have acted differently if epistemic vigilance was in the zeitgeist.

Learning cannot end at with primary or secondary schooling. Presumably, new technologies will emerge that will need new forms of epistemic vigilance. For example, DeepFakes are realistic videos which show someone (politician, celebrity) saying something that they did not say. Because they are hard to detect, the public has a right to be educated about them, and ways to detect them. A step towards this goal would be public service announcements on the potential dangers of having low epistemic vigilance, such as psychological (e.g., trolls), physical (e.g., swatting) and financial (e.g., phishing) harm.

2. We need a set of regulations or guidelines for using social media, based on ethics and awareness of cognitive biases.

Another recommendation is to focus on knowing how posts can lead to harmful actions taken by others. We think that the initial “pizza gate” Facebook post is akin to shouting “fire” in a movie theater in its outcome. Indeed, recent evidence suggests that people contribute more to spreading false news than robots (Vosoughi, Roy, & Aral, 2018). Whether it is a single malevolent person who gets people to falsely run for their lives, or an inaccurate viral post that gets Internet users roiled up, people are potentially harmed (e.g., Mele, 2016). We note that over the past 15 years, social media companies have gradually implemented rules of use, but we think they should be scrutinized in regard to ethics and factors that affect epistemic vigilance. For example, should a tweet be retweeted if there is clear evidence that the information is suspect? It is possible that pizzagate could have been prevented if more was done to prevent the number of inaccurate posts that resulted from the initial inaccurate one. It is true that Alex Jones was
ultimately kicked off of YouTube, Facebook and Twitter, but it took a while. Public policies and regulations may be the next condition for the sustained development of public communication online. Governments have set up agencies dedicated to understanding, tracking and educating citizens on potential harmful events (e.g., the Centers for Disease Control; National Institutes of Health). Regulations have also been created and enforced in order to limit the negative impact of major innovations such as the automobile. With the growth of social media, news agencies, and the availability of the Internet, we believe, comes the responsibility to identify causes of harmful events (e.g., “pizzagates”, bullying, school shootings) and the proliferation of inaccurate information via the Web. Of course, the U.S. guarantees free speech, but there are limits to free speech when it insinuates illegal activity, obscenity or invokes pandemonium. In some countries, laws prohibit public statements of racial hatred or insults toward law enforcement officers. The EU has recently begun to take steps for countering fake news (European Union, 2018). Just like the automobile created a need for traffic lights and speed limits, the Internet calls for some sort of guidelines as to what is permissible to post given the inadequacies that humans have in regard to the search for truth. Clearly, one’s motivation to find “the truth” is not enough because Welch was in pursuit of the truth, but likely fell prey to the illusory truth effect.

3. More funding is needed to understand and teach epistemic vigilance.

In 2009, the U.S. Department of Education’s Institute of Education Sciences launched the “Reading for Understanding Research Initiative” to fund projects to extend reading research beyond decoding and simple coherence-based inferencing (Douglas & Albro, 2014). We need similar leadership to drive rigorous interdisciplinary research to examine developmental trajectories, longitudinally, to study argument comprehension, evaluation and production and factors relating to epistemic vigilance. We need to identify suitable interventions for different
A reasoned approach

populations of readers in K-12 and adulthood. We also need developmental data to understand what is appropriate (e.g., perspective taking, coordinating cognition, attention and inhibition, strategic processing, epistemic beliefs) for a given population.
References


Delaney, K.J. (Feb 2017). Filter bubbles are a serious problem with news, says Bill Gates.


