

The challenge of addressing vivid lies spread by America's enemies in the cyber era.

Lee Ross, Department of Psychology, Stanford University

What can a social psychologist contribute to an examination of the problem of cyber-based attacks on US society? Recent targets include not only our elections but also the mutual trust, sense of shared values and purpose, and tolerance for diversity in political views, priorities, and concerns that are foundational to our democracy. In my view, that contribution involves awareness and appreciation of theory and research that helps us understand when and why people believe information that is factually false or misleading, and when and why they continue to believe, or at least be influenced by such information after it has been thoroughly discredited.

A second question is what, if anything is new and different about the problem of misinformation and “fake news” in the cyber age. Individuals, groups, and governments with malicious objectives have long concocted lies, employed propaganda, and found credulous targets to believe and spread such falsehoods? False reports and rumors about particular racial, ethnic, or religious groups have all too often prompted, and been used to justify, discrimination, exploitation, violence, and even genocide. Discrediting the Protocols of the Elders of Zion, or proving the innocence of Alfred Dreyfus, did little to end anti-Semitism. Countless exonerations of African-Americans lynched, imprisoned, or even executed for crimes they did not commit, and documentation of the lingering social and economic burdens created by segregation in much of our nation too often prompt rationalization or indifference rather than expression of guilt, shame and willingness to redress. Even today, lurid reports of rape, murder, and drug smuggling by illegal aliens, stoke anti-immigrant and anti-refugee sentiments. Moreover, reports of actual crime statistics—i.e., relative crime-rates for native born citizens, legal immigrants, and illegal immigrants, and percentages of illegal drugs entering our country across legal vs illegal points of entry— do not seem to faze the supporters of the US president who does the stoking.

I briefly address the “what if anything is different” question at various points in this essay, with full acknowledgment that I must rely upon cyber-savvy colleagues to enlighten me further about technical issues. At the outset of my essay, I can just state the obvious. It has never been easier, and cheaper, for a foreign power or other well-resourced organization not only to transmit misinformation to a vast audience but also to tailor messages to specific groups and even specific individuals. Furthermore, with today's social media, it never has been easier for recipients of such misinformation to pass it on to like-minded peers, in many cases embellishing that information and/or presenting it as factual rather than a claim of dubious provenance.

The bulk of my essay will deal with research on the extent to which, and ways in which, mixed, inaccurate, or debunked information can influence beliefs, and behavior in service of such beliefs. One issue is the willingness of people to believe what they read or hear, even if its credibility is dubious, particularly when what they read or hear is something they wish to believe or have a vested interest in believing, to be true. A related issue is the role that motivation versus sheer “cognitive laziness” —i.e., non-critical, non-deliberative, thinking—plays in such credulity. A second topic is the difficulty of getting people to adequately update and adjust their views

when information that they had initially taken to be veridical is challenged and discredited—again, especially in the case of views that serve their interests and reinforce important aspects of their identity. Some of the research and conceptual analysis I will discuss is my own, including studies conducted with students and colleagues over my fifty years at Stanford. However, I also touch on earlier research traditions, in particular work on dissonance theory and attitude change, and more contemporary work on prospect theory and loss aversion, both of which I know mainly as an admirer, or as an occasional commenter, critic, and synthesizer. I also make note of some more recent work done by investigators who employed *mTurk* and *Prolific* to explore research participants' assessments of the accuracy of false news headlines and their willingness to spread such posts to others.

At the outset of my discussion I want to acknowledge two limitations. The first is that the relevant laboratory studies and surveys typically dealt with the responses of participants who did not have particularly strong motives to create, believe, or spread misinformation. The second is that most of the research dealt with the immediate responses and with psychological processes displayed by individual participants rather than collective or group processes promoted by organized, well-resourced, sustained campaigns of deceit.

Selective Exposure and Confirmation Biases

Biases in exposure and memory: As the dissonance theory researchers led by Leon Festinger noted long ago, and most of us assume even if we are unaware of the relevant research, people are selective in the information they expose themselves to. They also show systematic biases in how critically they examine such information, in how much they remember, and in how they respond to subsequent efforts to discredit that information.

The early dissonance research focused heavily on consumer decisions (for example the purchase of a new car) and evaluations of chosen versus non-chosen alternatives. Some later studies examined choices about exposing oneself to political arguments that were consonant or dissonant with one's existing beliefs. The results of these studies, surprisingly, were mixed. Research participants did not consistently avoid discordant information. But once we turn our attention from laboratory studies with undergraduate research participants (who generally show a combination of low ideological commitment and high sensitivity to self-presentation concerns) and examine responses of a broader sample of people to contemporary news media, the findings linking political ideology to media choices are clear and unsurprising. A 2014 study by Mitchell et al reported that 88% of conservatives say they trust Fox News and 47% cite it as their main source of information about government and politics. Overall, they express greater trust than distrust in only of the 36 news sources they rated and two thirds of them say that most of their close friends share their views (and we can presume their media preferences). Liberals, by contrast, express more trust than distrust of 28 of the 36 media sources, and rely on a greater range of outlets (e.g., NPR, PBS, BBC as well as CNN and CNBC) than conservatives. Interestingly, liberals are more likely than are conservatives to block or defriend someone on a social network, and indeed to end a personal friendship because of politics.

While the current US media can be a conduit of misinformation, its direct relevance to the topic of this essay lies in the fact that those responsible for television news play a potentially vital role in directing their viewers' attention to cyber-based misinformation, unsubstantiated claims, and malicious rumors. More importantly, media executives and commentators decide how much airtime to devote to the debunking of such misinformation, to the provision of factual rebuttals, to consideration of the identities of the sources of misinformation, and to discussion of their means and motives.

Studies dealing specifically with memory for, and/or critical scrutiny of, cyber-based misinformation are obviously needed and no doubt will be forthcoming. What I describe in this essay is work conducted by psychologists on the broader topic of memory and assimilation of information. Cognitive and social psychologists have shown that people are most likely to remember information that is vivid, concrete, and readily assimilated to familiar scripts or schemas. They also have explored both cognitive and motivational biases, showing that people are most likely to remember information that is congruent with their pre-existing beliefs and their motivation to feel justified in those beliefs, especially beliefs that are self-serving and/or that reflect well on their groups. (However I suspect that particularly insulting or denigrating claims about one's group are also disproportionately likely to be remembered rather than forgotten.)

The results of an early study on memory by Jones and Kohler (1959) anticipated later findings. The investigators reported that students tended to remember convincing arguments supporting their views on segregation better than unconvincing ones, but to remember unconvincing arguments refuting their views better than convincing ones. From this study, and later ones, it is a reasonable extrapolation to assume that people show better memory for convincing challenges and rebuttals to cyber-based misinformation than unconvincing ones only to the extent they find the original information dissonant with their existing views. When it comes to dubious information that strengthens views, logically persuasive challenges to veracity or cogent counter-arguments are unlikely to undo the damaging effects of such information.

Confirmation biases. The greatest body of research on the psychology of belief involves biases not in memory but in the interpretation of evidence and assessment (Nisbett & Ross, 1980, pp 168-192; Kunda, 1987, 1990; Ditto & Lopez 1992; Edwards & Smith, 1996; Nickerson, 1998; Tabor & Lodge, 2006). This research is most relevant to the impact of not of outright lies than that of information or evidence that such be subject to critical scrutiny. Philosophers throughout history have recognized and commented on the phenomenon of confirmation bias. Perhaps the most notable is Francis Bacon (1620) who (in the *Novum Organum*) wrote: *The human understanding when it has once adopted an opinion ... draws all things else to support and agree with it. And though there be a greater number and weight of instances to be found on the other side, yet these it either neglects or despises, or else by some distinction sets aside or rejects.*

A study that my colleagues and I conducted forty years ago (Lord, Ross, & Lepper, 1979) provided convincing evidence of this bias. The noteworthy feature of our study was the demonstration that the tendency for people to give information that accords with their belief more weight than discordant information, produces an effect on belief that cannot be justified on Bayesian grounds—ie that mixed evidence of equal probativeness on both sides of an issue is apt

to heighten rather than attenuate strength of belief. In our study, capital proponents and opponents read (fictional) descriptions of two studies. One study compared murder rates in U.S. states with and without the death penalty; the other compared murder rates in states before and after the introduction of that penalty. Participants first read a quick description of each study and then read a more detailed account of each procedure and comments about potential strengths and limitation of the relevant designs. Half were told that the findings with one design supported the deterrent effect and the findings with the other designed undermined it; half were informed of the opposite association between designs and outcomes. Order of presentation of the studies and findings was counter-balanced.

When participants were asked to rate the probativeness of the studies and to report any changes in their own belief, the results were very straightforward. Participant showed no overall tendency to rate either design as superior, or to rate the study they read first more positively than the study that followed. Instead, they rated the design of whichever study supported their views more positively than the ones that challenged their views. Moreover, after reading both studies, the overall result was that both capital punishment proponents and opponents (on average) reported a strengthening of their views. *See Figure 1.*

One again, the extrapolation from this and various follow-up studies on confirmation bias to the concerns of our task force is obvious. When cyber messages include facts, figures, and claims, whether merely selective or otherwise misleading, or created out of whole cloth, those messages are apt to be taken at face value when they serve to support existing political beliefs, allegiances, and personal motives, but subjected to rigorous scrutiny when they challenge those beliefs, allegiances, or motives. When misinformation targets individuals or groups on both sides of a political divide, the result will be heightened polarization and mistrust— even in the case of intelligent voters. In fact, Kahan et al. (2012) showed that when it comes to the vital issue of climate change the gap in the willingness of liberal Democrats versus conservative Republicans to endorse a statement attesting to the reality of global warming, and its link to human activity, increased in step-wise fashion as a function of the research participants' scores on a standard test of reasoning ability. *See Figure 2*

I should note that not all investigators cite motivational factors as the chief culprit in producing agreement rather than critical scrutiny rejection of misinformation. For example, Pennycock and Rand (2018) showed that, at least in the case of simple headlines reporting untrue statements (eg *Hillary Clinton has filed for divorce* or that *Mike Pence claims that gay conversion therapy saved his marriage*), greater reasoning ability was associated with greater skepticism about the veracity of the headlines. This effect, furthermore, held for participants with views at both ends of the political spectrum. These investigators also showed that simply instructing their research participants to process those headlines more mindfully and critically reduced ratings of likely accuracy. Intellectual laziness, rather than bias, they argued, can be the cause of belief in non-veridical information, particularly when such information is implausible.

A phenomenon that there is little disagreement about is the effect of prior or multiple exposure to messages. The relevant much-replicated finding, which some have termed the “illusory truth” phenomenon, is that repetition increases the perceived credibility of untrue

and/or implausible facts and statements. In a study particularly relevant to the concerns of this essay, Pennycock, Cannon, & Rand (2018) showed that although overall belief in an implausible “fake news” headline was only about 4% when it was presented for the first time, that percentage increased to 8% when the headline was being seen a second time. This effect, moreover, was apparent regardless of whether it was consonant or dissonant with the individual’s own political allegiance. Furthermore, inclusion of a warning about possible inaccuracy did not diminish this illusory truth effect.

Extrapolation of such results to the possible impact not of one or two pieces of misinformation but to a concentrated campaign that repeats false claims and embellishes them with memorable details is disquieting. Even if the percentage of fake news recipients who readily believe such claims is small, and that small cadre of believers (abetted and joined those with more sinister motives) disseminate a steady stream of untruths about an individual or group, the consequences can be large for society. Websites that claim to present the views or actions advocated by a particular candidate or of that candidate’s supporters, but are in fact contrived by individuals or groups that seek to discredit the candidate may pose a particular threat, especially as increasingly sophisticated technologies are employed in the deception.

Post-discrediting belief perseverance and its consequences

The failure of warnings to curb credulity on the part of participants in the Pennycock, Cannon and Rand study (and many others) is disturbing. However a phenomenon that is even more relevant to the concerns of this essay is the difficulty of after-the-fact undoing of the damage done by misinformation. My own involvement with this problem began with a specific professional concern faced by social psychologists. Early in my career, it was very common to deceive research participants in order to evoke some sentiment, to produce some cognitive or social conflict, to prompt some response, or to have those participants believe that their behavioral choices would have particular consequences for themselves, for a fellow research participant, for their group, or for some segment of society. At the conclusion of the study, the experimenter would reveal the deception (and ideally explain the reason for it) and ease any resentment or other negative effects that had been produced by the deception and the acknowledgment of it. Many of us feared that such post-experimental debriefing failed to do the trick, that it not only gave rise to lingering distrust of researchers but in some cases failed to fully undo whatever misperception had been fostered.

In an attempt to “bottle” this particular phenomenon, Mark Lepper and I, with our graduate student Michael Hubbard (Ross, Lepper, & Hubbard, 1975) conducted a study in which we had participants try to distinguish real from fake suicide notes. (The stimulus materials were in fact real suicide notes and fakes.) Feedback was provided, one note at a time, leading some participants to believe they had failed miserably (only 10 of 25 correct) and others to believe that they had succeeded brilliantly (24 out of 25 correct). Our participants then received an extensive debriefing, which included seeing the precise schedule of *correct* and *incorrect* they had heard, and were assured they would have heard regardless of which notes they had identified as authentic. The results were striking, and for those concerned about the ethics of deception studies, somewhat alarming. (See table below) About half the difference between the success and

failure condition participants in terms of their self-rated abilities at the suicide discrimination and related tasks remained even after that thorough debriefing. A similar perseverance effect was apparent for individuals who watched the detection performance of a peer from behind a one-way mirror and heard the initial performance score and relevant debriefing information.

Post-debriefing estimates of performance and ability

Measure	Actors' estimates	Actors' estimates	Observers' Estimates	Observers' Estimates
	Success Condition	Failure Condition	Success Condition	Failure Condition
Est. re: initial performance	18.3 correct	12.8 correct	19.0 correct	12.4 correct
Est. re: next performance	18.3 correct	14.3 correct	19.1 correct	14.5 correct
Rated ability at task	5.0	3.83	5.33	4.0

In a sense, this finding represents an example of anchoring and (inadequate) adjustment as illustrated in research by Tversky and Kahneman (1974). We hypothesized that upon succeeding or failing, research participants generated *reasons* why they personally might have been good at the task in question. (For example, they could account for their success by note their interest in human psychology and their conviction that they have been particularly empathetic listeners when their friends shared their troubles. Alternatively, they could attribute their failure to their generally optimistic personality and relatively comfortable and stress-free upbringing). Lepper and I went on to show that explaining a statistical association before learning it had not in fact been demonstrated produce continuing belief in that association (Ross, Lepper, Steinmetz & Strack, 1977; Anderson, Lepper, & Ross, 1980). We also made the *a fortiori* argument that if belief perseverance continues even after decisive debriefing, the lingering effects of less decisively discredited information must be far greater. What we did not do, alas, is show that such lingering effects of discredited information pertain to mainstream media reports or social media messages, or that that those effects are particularly strong when the information is congruent with the recipients' current beliefs, allegiances, and motivations.

In the cursory literature search I did in preparing this essay, I came across a large number of studies that documented the difficulty of correcting erroneous widespread public beliefs about issues like immigration and spending on various social programs. An excellent article by Nyhan & Riefler (1990), which included a thorough literature review on the success of correcting erroneous beliefs about public issues, featured the report of four studies in which undergraduates at a Midwestern Catholic college read mock newspaper articles (on-line) containing a statement from a political figure that reinforced a common misperception. In an initial study, the article pertained to the presence of WMDs in Iraq, in follow-up studies other issues included the tax cut and stem cell research. In one condition in each of these studies, the article contained corrective information immediately after the false or misleading statement; in the other condition, that corrective information was omitted. The finding was that such corrections were highly effective

when the challenge was congruent with the participant's political stance but largely ineffective when the original statement was congruent with his or her stance and the challenge was incongruent with that stance. Moreover, in the case of the participants with the strongest ideological views, such ideologically "unwelcome" corrections actually produced a "backfire effect" whereby beliefs in the original statement were strengthened rather than weakened.

A limitation in my own past research on belief perseverance, and the research of most of my colleagues, was the failure to pursue significant behavioral consequences of unwarranted belief perseverance. In the introduction to his seminal *Theory of Cognitive Dissonance* volume, Festinger (1957) mentioned two dramatic examples whereby the behavior of people with strong, even self-defining, beliefs reflected dramatic resistance to disconfirmation. One example involved cases in which failed religious prophecies prompted true believers to increase rather than decreasing their efforts at proselytism. The other involved Japanese nationals living in the US who opted for repatriation to their homeland at the conclusion of World War II. According to a newspaper account of that day, those expatriates had thought that Japan would win the war, and despite what they read in newspapers and heard on the radio, continued to believe that the reports of their country's defeat and surrender were a hoax. Only after returning to Japan and witnessing firsthand their devastated and defeated homeland did they at last accept the veracity of the US media reports.

It is noteworthy in terms of the concerns of this essay that in both of these cases described by Festinger the resistance to unwelcome information occurred in contexts in which the resisters enjoyed a measure of social support. Indeed, in the case of the hapless individuals in the study by Festinger, Rieken, & Schachter (1956) who were at home rather than in the company of other believers when no rocket ship came to rescue them from the prophesized end of the world (and that the world was not destroyed) just drifted away from the group. They did not join in the effort to recruit new members to the group (which now proclaimed that it was only their faith that had prevented the originally prophesized cataclysm).

In short, efforts to refute false information, even the occurrence of events that directly contradict false beliefs, can prove disturbingly ineffective. Such refutations can even backfire if they prompt rationalizations and counter-arguments. I will comment more on this phenomenon in the next section of this essay, but first I think it is worth including another tidbit from the dissonance tradition. In his 1957 book, Festinger noted that an original impetus for his theorizing involved the report that immediately after an earthquake in India in 1934 the rumors disseminated by those in areas adjacent to the quakes were anxiety provoking rather than reassuring. Why would people chose to believe and spread rumors that went beyond what they know when those rumors made them feel more anxiety rather than less (as the reinforcement theorists of his day might have expected)? Festinger suggested that in the aftermath of the earthquake they accepted at face value and spread information that was congruent with what they were feeling—information that in fact justified those feelings. Again, the extrapolation should be clear. People whose feelings are initially stirred by a vivid story they heard and believed, and who continue to harbor those stirred feelings even after the story is discredited, will be inclined to seek out, believe, and spread information that is congruent with those feelings.

There is one other rumor transmission finding worth noting. In a 1955 experiment, Schachter and Burdick deliberately provoked rumor transmission at a private girl's school by having the principal conspicuously remove a girl from class, with no explanation but the statement to her, "Bring your hat and coat, you will be gone all day." The relevant finding was that the rumors initiated, and presumably passed on, by fellow students who liked that girl were predominately positive (e.g., "she is going to receive some honor"). By contrast those initiated by peers who did not like her were predominately negative (e.g., "she was caught cheating on a test"). The implication again is that false information congruent with pre-existing beliefs and sentiment is more likely to be accepted and even augmented, and to be transmitted and resistant to abandonment than information that is incongruent with existing beliefs and sentiments.

Collective and Counterfactual Rationalizations.

The original dissonance researchers did a fine job showing that their theory could generate non-obvious predictions, and their experiments were probably the cleverest in design and procedures in the history of my discipline. The researchers also were systematic and thorough when it came to specifying the conditions that increased the dissonance individual experienced when dealing with discrepancies between their behavior and their beliefs. The lion's share of studies focused on conditions under which individuals who had been induced to produce communications that were discrepant from their prior beliefs or otherwise to engage in behavior that seemingly required justification were most likely to reduce their dissonance by altering their beliefs. Their list of such moderating factors included perceived choice or freedom of action, commitment, effort, seriousness and irreversibility of consequences, and perceived responsibility for such consequences. (For reviews of this work, see Festinger, 1957, 1964; Brehm & Cohen, 1962; Aronson, 1969; Greenwald & Ronis, 1978; Cooper & Fazio, 1984; Cooper, 2007). It is likely that those same factors moderate responses both to fake news and to discrediting efforts. Thus, individuals who have spent time or money in service of their political beliefs, perhaps even paid a price in terms of friendships ruptured, and who have passed on misinformation in support of their views, are the people most likely to resist debunking efforts.

In retrospect, the experimental work on dissonance work was limited in several respects that are relevant to the concerns of this essay. First, despite the examples that Festinger described in his 1957 introduction to dissonance theory, when it came to experimentation the researchers focused heavily on self-reported beliefs rather than more consequential behaviors reflecting those beliefs. Second, they generally manipulated degree of dissonance evoked by particular experimental conditions rather than comparing the responses of individuals for whom particular pieces of information or particular actions would be consonant versus dissonant with their beliefs and values. Third, they focused almost exclusively on individual rather than social or collective processes—even though dissonance reduction and rationalization more generally have the gravest consequences when groups and their leaders encourage and even provide means for such rationalization.

When we consider the effects of malicious disinformation delivered via social media, the goal of the disseminators goes beyond influencing individual recipients. They want to produce a rapid spreading of that disinformation. They know that the veracity of the information will be

challenged and that responsible media will eventually debunk it. However, they count on the fact that the emotions and motivations initially stirred and the vivid images evoked by the misinformation will exert a continuing influence even after the information itself is refuted. Indeed, an early study by attitude change researchers demonstrated the “sleeper effect” (Hovland, & Weiss, 1951; Cook & Flay, 1978; Kumkale & Albarracín, 2004) whereby the impact of initial clues regarding the trustworthiness of the source weakens over time, and the effects of low credibility messages approach those of high credibility messages.

Dissonance theory would suggest that this phenomenon may be due in part to the fact that people are motivated to hold onto and justify sentiments and beliefs that are congruent with pieces of misinformation when the validity of that information is challenged or even total debunked. This may take the form of an active search for additional information that is more credible, or at least less easily refutable—particular in the case where the individual has already engaged in actions congruent with that misinformation (passing it on to friends and Facebook contacts, defending its veracity and plausibility to doubters, etc.). This search includes reliance upon “counterfactuals” (see Effron, 2018; Roese & Olson, 2004). The form of these counterfactuals is often of the form that the individual or group who was accused of *X* *may not be guilty of X but they would do X if they had the chance or they may not be guilty of X but, given their character and motives, I bet that have done a lot of X-like things.*

The counterfactuals of this form I have heard include the ones that many Israelis use to rationalize the least defensible aspects of their state’s treatment of Palestinians in the occupied territories and its disproportionate military responses to crude and largely ineffective rocket attacks launched from Gaza. (These counterfactual include, *“If Hamas and Hezbollah could inflict casualties on us with the same ease and impunity we have inflicted casualties on them, the totals would be not in the hundreds but the hundreds of thousands”* and *“If the US were facing rocket attacks from Mexico, its retaliations would much less tempered than ours have been.”*).

One aspect of the disinformation campaign launched from abroad during the 2016 electoral process captured my attention because of its relevance on to my own research on inter-group enmity and mistrust. The campaign, most experts agree, was designed in part to produce a particular outcome—the election of Donald Trump. But it also served to foster greater political division. Media attempts to cover that cyber-based disinformation campaign inevitably produce and reinforce complaints of media bias, and American’s at opposite ends of the political spectrum inevitably are dissatisfied by how those at the other end of the spectrum characterize that coverage in general and respond to particular media outlets and commentators. A vivid illustration of partisan responses to media coverage was provided by a study Robert Vallone, Mark Lepper, and I reported more 35 years ago (Vallone, Lepper & Ross, 1984).

In that study we showed students five nights of actual major network coverage of a tragic event, the killing in the fall of 1982 of hundreds of Palestinian and Lebanese Shiite refugees in two refugee camps near Beirut Lebanon. There was no disagreement about the fact that the perpetrators were right-wing Christian militiamen. The controversial issue for the pro-Israeli and anti-Israeli partisans in our study was what responsibility, if any, Israeli officials had for the crime. The results of the study were dramatic. There was virtually no overlap in the partisans’

ratings of the coverage. Pro-Israeli viewers rated the content of the coverage and the motives of those presenting as anti-Israeli while the anti-Israeli viewers thought that it was pro-Israeli. Moreover, the more the viewers knew about the broader conflict, the more biased they rated the coverage. In reporting these findings, we noted that they were not the result of our having cleverly contrived stimulus materials to support our predictions. The producers of the coverage presumably wanted to be lauded for their fairness and objectivity. I hate to imagine what the consequences for intergroup trust and respect would have been if we had made each partisan group aware of the assessment offered by the other group.

The deeper theoretical point relates to what Gilovich and I refer to as the “objectivity illusion” (Gilovich & Ross, 2015; see also Ross & Ward, 1996; Pronin, Gilovich & Ross, 2004). People have the illusion that their perceptions and assessments are essentially objective and unbiased, and that, to the extent that others see the world differently, those others are displaying various cognitive or motivational biases, including those reflecting self-interest, absence of careful consideration, short-term thinking, susceptibility to propaganda, and concern with peer-group approval. When information of dubious validity is under consideration, and when opposing partisans learn of each other’s response to that information, and hear the tone of debunkers and debates about the validity of the broader claims related to that debunked information, there will similarly be a heightening of enmity, distrust, and disparagement. I am sure that such would be welcomed, and perhaps was sought, by the disseminators of false information during the 2016 election.

Counteracting Misinformation and Reducing its Consequences

There is by now a large literature on the degree to which differences in cognitive abilities, intellectual style, temperament, as well as political affiliations, class, and national culture influence practices of rationalization, dissonance reduction, and openness to change views in light of new information. Rather than reviewing that literature, which is tangential to the main concerns of my essay, let me merely suggest that that moderation and mediation by such factors is likely to be issue-specific, and heavily influenced by the relevance of the issue to personal and collective values. The dissonance theory literature, which now includes hundred if not thousands of studies, again pin-points several factors that heighten the stakes for recipients of misinformation and those who decide to pass on that information, including commitment, perceived freedom of choice, potential and actual consequences, and avenues for rationalization.

Have researchers identified strategies that could be employed to forestall acceptance of false information sent via social media to individual who have not sought out such information or available on deceptively titled websites? I am sorry to say, there is not much research that I find encouraging. There are many studies showing that particular instructions that induce skepticism in research participants and more mindful and reflective consideration of the information they about to consider. Enhanced types of post-experimental debriefings have also been shown to reduce perseverance of the effects of false information. For example, in the previously described suicide note discrimination study by Ross, Lepper, & Hubbard (1975) such “process debriefing”, in contrast to standard debriefing that merely acknowledges the prior deception, largely eliminated the impact of deceptive success vs failure feedback. This was apparent in participants’

estimates of their earlier task performance, of their likely subsequent task performance and the ability at the relevant task, although even process debriefing failed to eliminate the effects of initial false information about the performance of a peer (see Table below).

Post-debriefing estimates of performance and ability

	<i>Actors' Estimates</i>		<i>Observers' Estimates</i>	
	Standard Debriefing	Process Debriefing	Standard Debriefing	Process Debriefing
Measure	Success-Failure	Success-Failure	Success-Failure	Success-Failure
Est of Initial Performance	5.5	1.2	7.0	5.0
Est. re: next performance	4.1	0.5	4.4	2.2
Rated ability at task	1.2	1.0	1.3	1.3

Another successful method involves having individuals actively “consider the opposite” (Lord, Lepper, & Preston, 1984)—that is, generate reasons why an existing belief and consider might be incorrect and to consider opposite possibilities—had a greater debiasing impact than explicit instructions to be fair and unbiased in considering evidence. Of course, it would be next to impossible to provide “process debriefing” or “consider the opposite” for every false message social media users receive. (If it were possible to identify such messages quickly enough to do so, it would be easier just to remove them as soon as they were transmitted).

Some of my colleagues have shown that general education about critical thinking can improve every day reasoning, and that explicit discussion of various cognitive and motivational biases can reduce susceptibility to misinformation and reduce bias. Certainly teaching Americans to resist lazy “system 1” or fast, shallow, non-deliberative, thinking (Kahneman, 2011; also Chaiken, 1980; Petty & Caccioppo, 1986; Petty, 1986) about important political issues would be a worthy goal for educators. While there is some evidence that education about statistical principle and logical fallacies can help people interpret statistical evidence more accurately (eg Nisbett, 2015; Nisbett, Jepson, Krantz & Kunda, 1985). I don’t know of any studies showing that lessons in reasoning or sensitivity to deception, can reduce the negative impact of cyber-based campaigns of misinformation of the sort that American were subjected to during the 2016 election and may face in future.

When we shift our attention from individual susceptibility to false information and individual resistance to debunking and focus on the dynamics of misinformation campaigns the challenge our social and political institutions face become more ominous. Those campaigns are apt to create groups and increase the influence of gatekeepers who pass on and augment misinformation, present dubious claims as simple fact, and respond to debunking with counterfactuals and claims that notwithstanding some inaccuracies “where there is so much smoke there must be fire.” Going forward, special attention must be paid to the development of

techniques to identify and resist coordinated campaigns of deception. An immediate task is one of prioritization. The first priority is obviously the protecting of the integrity and of the processes of voting and vote tabulation so that no external actors can gain access to and alter those processes. A second priority might be identification of false information sent to potential voters about dates and means of casting ballots—misinformation such that voting dates or places have been changed, that eligibility requirements have been changed, or even that one can cast one's ballot via some internet site. A third priority might simply be an early warning system that quickly identifies misinformation that is very widely and quickly disseminated—i.e., that is “going viral.”

It may be wishful thinking to imagine that public information and education campaigns can be designed to make the general public sophisticated and critical consumers of dubiously-sourced information. It may also be wishful thinking to imagine that we can somehow attach appropriate warning labels to such information, although it might be possible to enlist large numbers of individuals or even groups to be on the lookout for it, and report to some single, easily assessed website or phone-line. Indeed, in the best of all possible worlds one can envision the creation something akin to a 911 line available to all citizen who wish to check on the status of any communications relevant to electoral concerns and candidates. I suspect the technological challenges to creating such a public service can be met. It is the challenge of building public trust in such a service—and the willingness of the major political parties and media outlets to endorse the integrity and objectivity of that service that is apt to prove most difficult. Reestablishing trust in our most important institution—and in some case earning trust on the part of those whose distrust has proven warranted—is of course a goal that would pay dividends in meeting threats to our society that go beyond those posed by fake news and malevolent interference in our elections.

References Cited

- Aronson, E. (1969). The theory of cognitive dissonance: A current perspective. In L Berkowitz (Ed.) *Advances in experimental social psychology* (Vol. 4, pp. 1–34). New York: Academic Press.
- Anderson, CA, Lepper, M.R. & Ross, L. (1980) The perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality Psychology*, 39, 1037-1049.
- Brehm, J.W. & Cohen, A.R. *Explorations in cognitive dissonance*. 1962, New York: Wiley.
- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39(5) 752-766
- Cook, T. D.; Flay, B. R. (1978). The persistence of experimentally-induced attitude *change*. In L. Berkowitz (Ed.). *Advances in experimental social psychology* (Vol. 11, pp.1-57). Orlando Florida: Academic Press.
- Cooper, J. (2007). *Cognitive Dissonance: 50 Years of a classic theory*. London: Sage Publications
- Cooper, J. & Fazio, R.H (1984). A new look at dissonance theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 17, pp. 229-264). Orlando Florida: Academic Press.

- Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: use of differential decision criteria for preferred and non-preferred conclusions. *Journal of Personality and Social Psychology*, 63(4), 568–584.
- Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology*, 71(1), 5–24.
- Effron, D.A. (2018) It could have been true: How counterfactual thoughts reduce condemnation of falsehoods and increase political polarization *Personality and Social Psychology Bulletin*, 44(5), 729-745.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.
- Festinger, L. (1964). *Conflict, decision, and dissonance*. Stanford, CA, Stanford University Press.
- Festinger, L., Riecken, H.W., Schachter, S. *When Prophecy Fails* (1956). Minneapolis: University of Minnesota Press.
- Gilovich, T.D. & Ross, L (2015). *The wisest one in the room*. Free Press.
- Greenwald, A.G. & Ronis, D.L. (1978) Twenty years of cognitive dissonance: Case study of the evolution of a theory. *Psychological Review*, 85, 53-57.
- Kahan, D.M. Peters, E., Wittlin, M., Slovic, P Ouellette, L.L., Braman, D, & Mandel, G. (2012). *Nature Climate Change* volume 2, 732-735.
- Kahneman, D (2011). *Thinking fast and slow*. McMillan.
- Jones, E.E. & Kohler, R. (1959). The effects of plausibility on the learning of controversial statements. *Journal of Abnormal and Social Psychology*. 57, 315-320.
- Kumkale, G.T.; Albarracín, D. (2004). "The Sleeper Effect in Persuasion: A Meta-Analytic Review". *Psychological Bulletin*. 130 (1): 143–172
- Kunda, Z. (1987). *Motivated inference: Self-serving generation and evaluation of causal theories*. *Journal of Personality and Social Psychology*, 53, 636–647.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498.
- Lord, C. G., Lepper, M. R., & Preston, E. (1985). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, 47, 1231-1243.
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37, 2098–2109.
- Mitchell, A, Matsa, K.E. Gottfried, J & Kiley, J. (2015). *Political Polarization and Media Habits*. Pew research center, - internet.psych.wisc.edu
- Nickerson, R.S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175-220.
- Nisbett. R.E., (2015) *Mindware: Tools for smart thinking* City: Farrar, Straus, and Giroux
- Nisbett, R. E., Krantz, D. H., Jepson, D., & Kunda, Z. (1983). The use of statistical heuristics in everyday reasoning. *Psychological Review*, 90, 339-363

- Nisbett, R.E., & Ross, L (1980), *Human inference: Strategies and shortcomings of social judgment* Prentice-Hall, Century Series
- Nyhan, B. & Reifler, J. (2010) When corrections fail: The persistence of political misperceptions, *Political Behavior*, 32(2), 303-330.
- Pennycook, G & Rand, D.G. (2018) Lazy, not biased. Susceptibility to partisan fake news is better explained by a lack of reasoning than by motivated reasoning, *Cognition*, 10, 1016-1087.
- Pennycook, G., Cannon, T.D. (2018). Prior exposure increases perceived accuracy of fake news *Journal of Experimental Psychology: General*. 147, 1865-1880.
- Petty, R.E. (1986) *Communication and persuasion: central and peripheral routes to attitude change*. Springer-Verlag,
- Petty, R.E. & Cacioppo, J T (1986) *The elaboration likelihood model of persuasion. Advances in Experimental Social Psychology peripheral routes to attitude change* (vol 19) 196- xxx
- Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review*, 111, 781–799.
- Roese, N.J. & Olson, JM (Eds) (2014) *What might have been: The social psychology of counterfactual thinking*, Psychology Press.
- Ross, L., Lepper, M. & Hubbard, M. Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology*.1975, 880-892.
- Ross, L., & Lepper, M. (1980) . The perseverance of beliefs: Empirical and normative considerations.in R.A. Shweder & D.W. Fiske (Eds.). *New directions for methodology of behavioral science:: Fallible judgment in behavioral research*. Jossey-Bass.
- Ross, L., & Lepper, M., Strack, F. & Steinmetz, J.(1977). Social explanation and social expectation: The effects real and hypothetical explanations upon subjective likelihood. *Journal of Personality and Social Psychology*, 35, 817-829.
- Ross, L., & Ward, A. (1996) Naive realism in everyday life: Implications for social conflict and misunderstanding. In T. Brown, E. Reed, & E. Turiel (Eds.), *Values and knowledge* . Erlbaum, 1996.
- Schachter, S. & Burdick, H (1955) A field experiment on rumor transmission and distortion *Journal of Abnormal and Social Psychology*, 50. 363-671
- Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755–769.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Vallone, RP, Ross, L., & Lepper, MR. The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut Massacre. *Journal of Personality and Social Psychology*, 1985, 49, 577-585.

Addition Sources.

- American Psychological Association (2018). Why we're susceptible to fake news and how to defend against it. *Skeptical Inquirer*, 42 (6): 8–9.
- Allcott, H., & Gentzkow, M. (2017). *Social media and fake news in the 2016 election* (NBER Working Paper No. 23098). Retrieved from <http://www.nber.org/papers/w23089>
- Arkes, H. R., Boehm, L. E., & Xu, G. (1991). Determinants of judged validity. *Journal of Experimental Social Psychology*, 27, 576–605.
- Arkes, H. R., Hackett, C., & Boehm, L. (1989). The generality of the relation between familiarity and judged validity. *Journal of Behavioral Decision Making*, 2, 81–94.
- Bail, C. A., Argyle, L.P., Brown, T.W., Bumpus, J., & Voltovskiy. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Science*, 115 (37), 9216–9221.
- Beam, M. A., Hutchens, M. J., & Hmielowski, J. D. (2018). Facebook news and (de) polarization: reinforcing spirals in the 2016 US election. *Information, Communication & Society*, 21 (7), 940–958.
- Berger, J. & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research* 49, 192–205 (2012).
- Bronstein, M. V., & Cannon, T. D. (2017). Bias against disconfirmatory evidence in a large nonclinical sample: Associations with schizotypy and delusional beliefs. *Journal of Experimental Psychopathology*, 8 (1), 1–39.
- Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A. Caldarello G. Stanley H.E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceeding of the National Academy of Science* 113, 554–559
- Ecker, U. K. H., Hogan, J., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory & Cognition*, 6, 185–192.
- 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.
- Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, 144, 993–1002.
- Flanagin, A. J. (2017). Online social influence and the convergence of mass and interpersonal communication. *Human Communication Research*, 43(4), 450–463.
- Greenberg, J., & Jonas, E. (2003). Psychological motives and political orientation—the left, the right, and the rigid: comment on Jost et al. (2003). *Psychological Bulletin*, 129(3), 376–382.
- 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(6), 1420–1436.
- Kuklinski, J. H., Quirk, P. J., Jerit, J., Schweider, D., & Rich, R. F. (2000). Misinformation and the currency of democratic citizenship. *The Journal of Politics*, 62(3), 790–816.

- Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K.M., Menczer, F., . . . & Schudson, M. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096.
- Lewandowsky, S., Ecker, U., Seifert, C., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*, 106–131.
- Mitchell, A., Gottfried, J. Kiley, J. K. E & Matsa, K.E. (2014). Political polarization & media habits. *Pew Research Center*; www.journalism.org/2014/10/21/political-polarization-media-habit.
- Pennycook, G., & Rand, D. G. (2018b). *Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking*. Available at SSRN: <https://papers.ssrn.com/sol3>
- Polage, D. C. (2012). Making up history: False memories of fake news stories. *Europe's Journal of Psychology*, *8*, 245–250.
- Roese, N. J., & Epstude, K. (2017). The functional theory of counterfactual thinking: New evidence, new challenges, new insights. *Advances in Experimental Social Psychology*, *56*, 1-79.
- Schwartz, M. (1982). Repetition and rated truth value of statements. *American Journal of Psychology*, *95*, 393–407.
- Schweitzer, M. E., & Hsee, C. K. (2002). Stretching the truth: Elastic justification and motivated communication of uncertain information. *Journal of Risk and Uncertainty*, *25*, 185-201.
- Self, W. (2016). Forget fake news on Facebook – the real filter bubble is you. *New Statesman*, Nov.16.
- Silverman, C. (2011). The backfire effect: When your deepest convictions are challenged by contradictory evidence, your beliefs get stronger. *Columbia Journalism Review*
- Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research*, *31*, 713–724.
- Stangor, C. & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures, *Psychological Bulletin*, *111* (1): 42–61.
- Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning*, *13*(3), 225–247.
- Swire, B., Ecker, U. K. H., & Lewandowsky, S. (2017). The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 1948–1961.
- Vosjoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*, 1146-1151.
- Wood, T; Porter, E (2018). The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence *Political Behavior* (forthcoming).

FIGURES

Figure 1 Confirmation Bias in Evaluating Mixed Evidence

Changes in Views of Death Penalty Proponents and Opponents After Reading Two Studies with Opposite Results

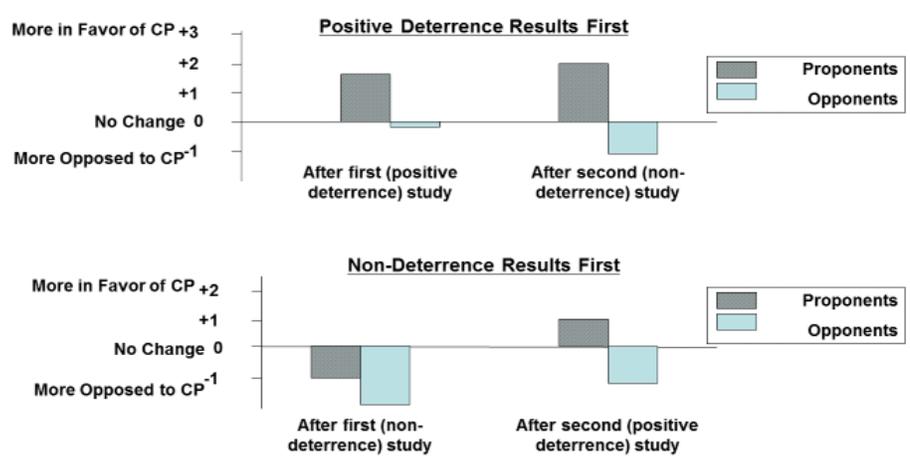


Figure 2: Association between agreement with a scientific claim and reasoning ability

There is "solid evidence" of recent global warming due "mostly" to "human activity such as burning fossil fuels."

