Polit Behav
https://doi.org/10.1007/s11109-018-9443-y

ORIGINAL PAPER

The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence

Thomas Wood[1] · **Ethan Porter**[2]

Abstract Can citizens heed factual information, even when such information challenges their partisan and ideological attachments? The "backfire effect," described by Nyhan and Reifler (Polit Behav 32(2):303–330. https://doi.org/10.1007/s11109-010-9112-2, 2010), says no: rather than simply ignoring factual information, presenting respondents with facts can *compound* their ignorance. In their study, conservatives presented with factual information about the absence of Weapons of Mass Destruction in Iraq became *more* convinced that such weapons had been found. The present paper presents results from five experiments in which we enrolled more than 10,100 subjects and tested 52 issues of potential backfire. Across all experiments, we found no corrections capable of triggering backfire, despite testing precisely the kinds of polarized issues where backfire should be expected. Evidence of factual backfire is far more tenuous than prior research suggests. By and large, citizens heed factual information, even when such information challenges their ideological commitments.

All figures and tables in this paper can be replicated with the syntax available at the *Political Behavior* dataverse: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/AGRX5U

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11109-018-9443-y) contains supplementary material, which is available to authorized users.

✉ Thomas Wood
wood.1080@osu.edu

Ethan Porter
evporter@gwu.edu

[1] The Ohio State University, Derby Hall 154 N Oval Mall, Columbus, OH 43212, USA

[2] School of Media and Public Affairs, The George Washington University, 805 21st Street NW, Washington, DC 20052, USA

Published online: 16 January 2018

Keywords Backfire effect · Factual correction · Misinformation · Factual information

Introduction

Can citizensaffirm factual information about politics, even when the facts cut against their ideological and partisan beliefs? Or are they destined to view the facts through the "perceptual screen" (Campbell et al. 1960) that partisanship and ideology impress upon them? Nyhan and Jason (2010) offer strong evidence for the latter position: Subjects presented with facts correcting misperceptions relevant to their ideology responded by doubling down on their misperceptions. Particularly among conservatives, attempts to correct misperceptions activated a "backfire effect" against empirical facts, with subjects more strongly expressing a non-factual belief. Subsequent research has corroborated Nyhan and Reifler's initial finding. Confirming that the Affordable Care Act did not introduce "death panels" for terminal patients entrenched this conviction (Berinsky 2017; Nyhan et al. 2013). Indicating that President Obama is a Christian can make Republican respondents more strongly believe that he is a Muslim (Nyhan and Reifler 2017). Republicans were systematically less accurate in their estimates of unemployment when surveyed immediately after the release of a jobs report during the Obama presidency (Schaffner and Roche 2017). Equally significant is the backfire effect's prominence in popular political accounts. Political elites habitually cite this research to explain the public's failure to heed facts consistent with the elite's policy proposal of the moment.[1]

Another line of research, however, suggests that citizens are indeed capable of learning (Gerber and Green 1999; Howell and West 2009; Fishkin and Robert 2005). Sometimes with encouragements, and sometimes with small monetary incentives, citizens can absorb and retain complex new political information (Barnes et al. 2016; Fishkin 2005; Kuklinski and Quirk 2000); and factual receptivity can take place despite partisan differences (Bullock et al. 2013; Prior 2007; Prior et al. 2015).

The term "backfire effect" has an ambiguous quality, especially in popular accounts, where it is often used to describe any incidence of motivated reasoning. In this paper, we use "backfire effect" consistent with Nyhan and Jason 2010—when the average respondent is made less accurate on a factual question when exposed to a false claim and its correction, compared to those who only see the false claim. Examples of backfire by this standard are Nyhan and Jason 2010, Berinsky (2017), Nyhan et al. 2013, and Schaffner and Roche (2017). Also by this standard, we do not regard as backfire those studies which one: find only a change in political preferences (Redlawsk 2002; Kaplan Gimbel and Harris 2016; Thorson 2015;

[1] Google News search for the "backfire", "backlash", or "boomerang" effect and the names of Nyhan or Reifler returns over 300 unique articles. The 2010 backfire paper has also enjoyed remarkable academic attention. Among all papers printed in *Political Behavior* in the last 10 years, "When Corrections Fail" has been cited four times as much as the next most cited paper.

Trevors et al. 2016; Swire et al. 2017), two: studies which only find a corrections have a maladaptive effect on intended behavior (Nyhan et al. 2014), three: studies which find increased factual polarization *among corrected respondents* (Gollust et al. 2009), four: studies which only report a fluency effect where false claims are more easily remembered, since no corrections are randomly assigned in such designs (Sanna et al. 2002; Skurnik et al. 2005; Berinsky 2017), or five: studies where a framing effect compounds factual inaccuracy (Kahan et al. 2017).

Building on Nyhan and Reifler, the present paper maps the boundaries of the backfire effect. What issues provoke resistance to factual information? Which ideological groups, and members of which parties, are most likely to evince backfire? And which issues provoke which ideological groups to backfire? Nyhan and Reifler observed backfire on two high salience issues (whether WMD were found in Iraq and whether tax cuts ultimately increase tax receipts) and found no backfire on one less salient, more technical issue (the specific types of stem cell research prohibited by President Bush.) Testing only three issues makes it difficult to determine if backfire is caused by ideological group differences in factual receptivity, or if their results instead reflect the salience and ideological importance issues being corrected. That is, might liberals also prove factually maladaptive when presented with correction to the right issue?

Since Nyhan and Reifler found that backfire effect was heterogeneous between corrected issues, we staged five separate studies comprised of 52 commonly misconstrued policy areas. Table 1 summarizes the issues corrected. The issues were purposefully chosen to tap the most important ideological symbols along the political spectrum. For liberals, this meant correcting President Obama's statement about the role of drug sentencing in growing the incarcerated population, and contradicting Hillary Clinton's claim about the incidence of gun violence. For conservatives, we contradicted Donald Trump's claim about undocumented immigrants' criminal proclivities, and provided corrections about the incidence of abortion and teen pregnancy. If *any* correction should prompt a respondent to counter-argue unwelcome facts, and inadvertently entrench their pre-correction attitudes, these politically important, readily understood, recurrently debated issues stand out as likely candidates. In Carmines and Stimson's (1980) memorable turn of phrase, such issues are deemed "easy," because their recurrence in American politics has trained individuals on how to respond.

Across five studies in which we enrolled more than 10,100 subjects, we observed no backfire effects. When presented with facts that correct political leaders, subjects along the ideological spectrum are capable of heeding the correction and bringing their beliefs in alignment with the facts. This occurs even when the corrections directly conflict with subjects' ideological commitments.

To be sure, as we discuss below, some of the misperceptions were more salient than others. Yet despite the substantive differences between issues, across issues the result was uniform: subjects did not backfire. When presented with factual information, people can heed the facts, *even when* doing so forces them to separate from their political attachments. Of course, ideology and partisanship shaped the *extent* of our subjects' factual receptivity. The "perceptual screen" that partisanship instills (Campbell et al. 1960) is real. Yet evidence of differential learning along

Table 1 Issues, speakers, and corrections, by study

Issue	Speaker	Correction
Study 1: Corrections to 8 ideologically polarized issues—"Study 1" section		
(Full text of statements, corrections, and items can be found in Table 3 beginning on p. iii in the appendix)		
Spiraling gun violence	H.Clinton	Gun homicides declined by 50% since 1994
Drugs drive prison growth	Obama	Only 20% of prisoners incarcerated for Drug crimes
Solar has more jobs than oil	Clinton	Oil industry employs four times more people than solar power
Hedge fund mangers pay less tax than workers	H.Clinton	Average hedge fund manager pays twenty times as much tax
Discrimination sole cause of gender wage gap	Obama	Discrimination accounts for only a fraction of the wage difference
Obama cuts defense	Rubio	Obama has increased defense spending relative to Bush
Mexican immigrants disproportionately criminal	Trump	Undocumented offend at lower rate than general population
Rising violence against police	Cruz	Number of police officers killed currently at 130 year low
US taxes highest in world	Trump	US taxes second lowest among all OECD members
Study 2: Corrections to 8 issues where liberals and conservatives have misspoken—"Study 2" section		
(Full text of statements, corrections, and items can be found in Table 5 on p. xi)		
Obama passed TARP	Palin Obama	TARP was passed by George W Bush
Obama accommodates undocumented immigrants	Cruz Gutierrez	Obama deported undocumented at twice the rate of GW Bush
Spiraling teenage pregnancy rate	Carson Lee	Since 1991, black teen pregnancy fallen by 66%, 50% among whites
China holds most US debt	Romney Obama	China holds about 12% of US debt
Whites soon a racial minority	Graham Langoria	Whites a majority until 2045, majority of voters until 2070
Spiraling Chicago homicides	LaPierre Obama	Chicago homicides at 36 year low
Spiraling abortion rate	Ryan Obama	Abortion rate at 40 year low
Obama curtails drones' use	Graham Obama	Obama has ordered ten times as many strikes as his predecessor
Study 3: Concealing the corrections in a longer article—"Study 3" section		
(Articles and corrections can be found in Sect. A.5 on p. xiii Survey items can be found in Table 7 on p. xvii)		
Spiraling Chicago homicides	Obama LaPierre	Chicago homicides at 36 year low
Mexican immigrants disproportionately criminal	Trump	Undocumented offend at lower rate than general population

Table 1 continued

Issue	Speaker	Correction
Spiraling abortion rate	Ryan	Abortion rate at 40 year low
Solar has more jobs than oil	Clinton	Oil industry employs four times more people than solar power
US Health care twice as expensive as all other states	Sanders	Swiss health care only 40% less expensive per capita than the US
WMD were found in Iraq	Bush	No WMD were found in Iraq
Study 4: Measuring the effect of survey item complexity—"Study 4" section		
(Articles and corrections can be found in Sect. A.9 on p. xvi. Survey items can be found in Table 11 on p. xxiii)		
WMD were found in Iraq	Bush	No WMD were found in Iraq
Spiraling abortion rate	Ryan	Abortion rate at 40 year low
Real unemployment rate > 30%	Trump	Unemployment is currently 4.9% (February 2016)
Tax cuts will pay for themselves	Trump	Trump tax plan will increase deficit by almost $10 trillion
EPA: fracking pollutes water	Sanders	EPA finds no systematic relationship
Solar has more jobs than oil	Clinton	Oil industry employs four times more people than solar power
Study 5: Comparing Turk subjects to nationally representative panel respondents—"Study 5" section		
(Articles and corrections can be found in Sect. A.11 on p. xxiv. Survey items can be found in Table 12 on p. xxvii.)		
Spiraling abortion rate	Ryan Obama	Abortion rate at 40 year low
Rising violence against police	Cruz	Number of police officers killed currently at 130 year low
Real unemployment rate > 30%	Trump	Unemployment is currently 4.9% (February 2016)
EPA: fracking pollutes water	Sanders	EPA finds no systematic relationship
Spiraling gun violence	H.Clinton	Gun homicides declined by 50% since 1994
Tax cuts will pay for themselves	Trump	Trump tax plan will increase deficit by almost $10 trillion

partisan lines does not obviate the overall learning we observe (Gerber and Green 1999). The average subject exposed to the correction subsequently expressed attitudes more in line with the facts.

Which Corrections are Prone to Counter-Argument?

Nyhan and Jason (2010) posit that factual backfire is a consequence of a psychological property called "counterarguing." At least since Campbell et al. (1960), political scientists have known that individuals employ a "perceptual

screen" when shopping the marketplace of ideas. If people were dispassionate Bayesians, they would willingly acquiesce to corrections of their ideological affiliates and to their adversaries alike. When examined in both experimental and observational settings, the average subject's search for political information is inconsistent with the Bayesian expectation. Individuals have a powerful preference for media which comports with their ideological preference (Stroud 2008). Exposure to pieces of of ideologically welcome and unwelcome evidence leads to attitudinal polarization, since the welcome evidence is adopted, and the unwelcome is scrutinized (Lord et al. 1979). Accounting for this pattern of self-deluding information search, Taber and Lodge's (2006) specification posits the competing influences of *accuracy* goals (those factors which motivate respondents to have a factually accurate understanding) and *partisan* goals (factors which motivate the construction of a mental image that comports with an individual's political preferences). Since the average person does not pay a cost for being factually inaccurate in their political beliefs, but might bear an expressive cost in adopting an ideologically incongruous fact, the partisan goals (weighing against factual accuracy) are expected to predominate.

The canonical lab treatments used to induce counter-argument are usually policy arguments. For instance, Taber and Lodge (2006) provided Stony Brook undergraduates with policy claims about affirmative action and gun control. In this setting, counterargument seems predictable: sophisticated subjects were provided valence arguments on two of the most polarizing debates in American politics. To abandon one's prior policy preferences is costly, while concocting a reason to discount unwelcome arguments is less taxing. Taber and Lodge's finding that counter-argument and attitudinal polarization were most apparent among the especially sophisticated undergraduates comports with this theory.

How this explains factual receptivity, or the lack thereof, is far from obvious. The receipt of a policy argument has clear logical implications—if one accepts the argument, there is a powerful expectation that your policy preferences be adapted accordingly. A factual correction seems less consequential. After being apprised that no WMD were found in Iraq, or that the Bush tax cuts reduced federal tax receipts, one can accept these facts without any glaring policy implication—rather, one might conclude that the war or the tax cuts were still worth it, on balance.[2] The possibility of counterargument, however. requires that respondents perceive facts as having inescapable, unwelcome political implications. Merely adopting or ignoring it is not tenable—the fact must be assailed, and a countervailing policy position more strenuously adopted.

The literature suggests two psychological mechanisms for the counterargument of political facts. Upon being presented an unwelcome fact, a respondent might:

1. Infer that a side advancing a "weak" argument (i.e., an argument that fails to connect with a relevant, preexisting consideration) has unintentionally advertised the overall weakness of their factual case, so that a respondent would

[2] In this way, the apparent difficulty in making one's policy preferences fit with one's factual attitudes is redolent of Americans' struggle to have their policy preferences fit with each other—what Converse (1964) famously described as poor "constraint."

move further away from the position suggested by this argument, than if the side had simply made no argument at all, or

2. Be inspired to counteract unwelcome arguments, by bringing to mind countervailing considerations. While such a respondent may merely have wanted to nullify a hostile argument, they might prove so unexpectedly adept at calling to mind offsetting considerations that they end up *more* certain of their prior preference.

Prior work (Brock 1967; Chong and Druckman 2013) has suggested that both mechanisms are apparent in certain settings. Importantly, each possibility predicts different types of corrections as prone to induce backfire. The first would suggest that corrections only vaguely related to a statement would indicate the weakness in the case correcting a speaker. This vague correction might have the unintended consequence of impelling the respondent to disregard the logic of those providing a correction. The second possibility, on the other hand, would suggest that corrections which forcefully contradict the speaker are more prone to factual backfire, by prompting subjects to marshal their cognitive forces to repel an unwelcome factual intruder. "Does Counterargument Explain Our Pattern of Findings?" section tests these possibilities for all our corrections.

Competing with both theories of counterargument is the possibility that subjects detest intellectual effort just as much, or more, than they fear adopting a counter-ideological consideration. Counterargument is a particularly effortful way to escape unwelcome political facts' logical consequences. Americans have a well documented facility for avoiding cognitive effort when deducing political attitudes (Sniderman et al. 1993; Mondak 1993, 1994). Specifically, ideology can be used as a *group heuristic*. Such a heuristic ignores traditional ideological arguments about policies and trade-offs, and instead uses the ideological spectrum to arrange political figures and perceived groups into a continuum of allies and adversaries (Jost et al. 2008; Conover and Feldman 1981). This models predicts different response to corrections than the counterargument model. In place of counterarguing an unwelcome fact, corrections are either ignored, or phlegmatically accepted, since the respondent knows that allegiance to their ideological pole is at root a *group affiliation*, rather than an intellectual commitment contingent on logic and factual consistency. At its maximum, this heuristic would suggest a correction (whether proximate or distal) would at worst be ignored, but never cause backfire. Ironically, if respondents' factual receptivity is determined by affective ideology, they should be made comparatively willing to acquiesce to facts, knowing that acquiescing to a counter-ideological fact will have scarce consequence for a broader set of attitudes and preferences. That is, in a polarized age, its difficult to even countenance changing one's political preferences, reducing the stakes of unwelcome ideological facts. Accordingly, this possibility would predict that correction strength (or proximity) would be unrelated to average correction sizes, and backfire would never be observed, since respondents would not feel obligated to counter-argue facts.

The relationship between average correction size and correction proximity therefore allow us to discriminate between the two types of counterargument, and the possibility that ideology serves as a heuristic precisely to avoid the kind of

costly cognitive work counterargument presumes. We find that neither proximate nor distal corrections feature small or large correction effects, on average. Since we observe no backfire in any test, and correctional proximity does not co-vary with correction size, our results suggest affective ideology, rather than counterargument, typify our subjects' response to factual correction.

Experimental Overview

Our experiments used real instances of misstatements by political leaders from both sides of the aisle. Some subjects were randomly vended a correction, consisting of neutral data from governmental sources. All subjects were then asked whether they agreed with the original misstatement. Two of the studies presented the misstatements as if they were excerpts from a longer newspaper article; three others presented subjects with complete fictitious news articles.

Table 1 summarizes the design of all five studies. In the first study, we showed subjects eight instances of actual public comments from political figures, in which the speaker diverged from available empirical evidence. Four of the misstatements came from Democrats; four came from Republicans. For the second study, we identified issues about which politicians from both ends of the ideological spectrum had made misstatements that could be corrected by reference to neutral data. For example, political leaders from both parties have made erroneous claims about the abortion rate.

In the third study, subjects read fictitious newspaper articles containing real misstatements from politicians. For each article, some subjects were randomly assigned to read a version of the article in which a corrective paragraph was embedded. The paragraph recited data from a neutral source. Subjects in this study were also exposed to a replication of the original Nyhan and Reifler news article about WMD in Iraq. In our fourth study, we tested if complicated survey items can induce backfire. Subjects again read fictitious newspaper articles with randomized corrections, before being presented a survey item from one of three levels of complexity.

In our final study, we compared the factual receptivity of Mechanical Turk subjects (who comprised our first four studies) to subjects drawn from a nationally representative survey panel. Both sets of subjects received identical corrections, and the experiments were concurrently.

Across five studies and 52 issues, among our 10,100 respondents, we did not observe a single instance of factual backfire. This does not lead us to conclude that backfire is categorically impossible. Certain issues and certain questions—perhaps asked at moments when ideology or partisanship, or both, are particularly salient—might plausibly trigger factual backfire. However, despite conducting most of our experiments during the height of the presidential primary, on issues of keen political interest, we were unable to locate a single issue of backfire. Our results are summarized in Fig. 1. Among liberals, 85% of issues saw a significant factual response to correction, among moderates, 96% of issues, and among conservatives, 83% of issues. No backfire was observed for any issue, among any ideological

cohort. The backfire effect is far less prevalent than existing research would indicate. It also does not appear to be the exclusive provenance of one ideological pole. When presented with information that conflicts with their political commitments, citizens take a Joe Friday approach: they choose just the facts, ahead of their ideology.

Study 1

In the first study, subjects were presented with eight genuine public comments from politicians. In each instance, the politician's comments were at odds with the facts. We selected misstatements about issues important to both ends of the ideological spectrum, and we evenly divided our speakers between the major parties. For each issue, a subject was randomly exposed, or not exposed, to a correction that cited neutral government data to rebut the speaker. For each issue, subjects were asked their agreed the speaker's position. For example, we presented subjects with remarks by Senator Ted Cruz on the killings of police officers. In October 2015, Senator Cruz said:

> The number of law enforcement officers killed as a result of violence has been on a precipitous upswing. If the police are intimidated, if they are scared, if they are not willing to do their jobs, we know the result. The result is the loss of life. The result is rising crime

Subjects randomly vended a correction were then told:

> In fact, according to public records, homicides of law enforcement officers have been declining for decades. Fewer police officers were killed in 2015 than any year since the 1890s.

All subjects were then asked to agree or disagree, on a five-point scale, with the following claim:

> The number of police officers killed in the line of duty is rising fast.

The full list of statements, correction, and survey items are available in "Experimental Overview" section. With a dateline at the top, the statements were designed to look like excerpts from news articles. All corrections provided came from governmental organizations and were cited as such. The order of issues was randomized for each subject. All speakers' partisan affiliations were prominently displayed.

More formally, each experiment is an adapted latin squares design (Cochran and Cox 1957):

Correction among liberals
Correction among moderates
Correction among conservatives
Absolute value of ideological effect
Undocumented disprop. criminal (Trump)
Discrimin. sole cause wage gap (Obama)
Surge in killings of police officers (Cruz)
US taxes highest in world (Trump)
Most prisoners serving drug sentences (Obama)
Workers pay more tax than financiers (Clinton)
Obama cuts defense spending (Rubio)
Gun violence at all time high (Clinton)
1
Obama accomodates undocumented (Cruz)
Spiraling abortion rate (Ryan)
Spiraling teen pregnancy rate (Carson)
Spiraling Chicago gun violence (LaPierre)
China holds majority of US debt (Romney)
Obama passed TARP (Obama)
Spiraling abortion rate (Obama)
Spiraling teen pregnancy rate (Lee)
Obama accomodates undocumented (Gutiérrez)
Obama restricts drones' use (Graham)
Whites imminently a minority (Graham)
Spiraling Chicago gun violence (Obama)
Whites imminently a minority (Longoria)
China holds majority of US debt (Obama)
Obama passed TARP (Obama)
Obama restricts drones' use (Obama)
2
Undocumented disprop. criminal (Trump)
Spiraling abortion rate (Ryan)
US healthcare is a privilege (Sanders)
Spiraling Chicago gun violence (LaPierre)
More jobs in solar than oil (Clinton)
Spiraling Chicago gun violence (Obama)
3
US taxes highest in world (Trump)
True UE rate 30%-40% (Trump)
Spiraling Abortion Rate (Ryan)
WMD found in Iraq (Bush)
EPA: fracking pollutes drinking water (Sanders)
More jobs in solar than oil (Clinton)
4
Surge in killings of police officers (Cruz)
US taxes highest in world (Trump)
True UE rate is 30%-40% (Trump)
Spiraling abortion rate (Ryan)
EPA: fracking pollutes drinking water (Sanders)
More jobs in solar than oil (Clinton)
Gun violence at all time high (Clinton)
Spiraling abortion rate (Obama)
5: Turk
Surge in killings of police officers (Cruz)
US taxes highest in world (Trump)
EPA: fracking pollutes drinking water (Sanders)
True UE rate 30%-40% (Trump)
Spiraling abortion rate (Ryan)
Gun violence at all time high (Clinton)
More jobs in solar than oil (Clinton)
Spiraling abortion rate (Obama)
5: Lucid
-2
-1
0
-1
0
-1
0
0
1
2
Difference along 5 point scale
Significance ○ p >= .05 ● p < .05

◀**Fig. 1** Summary correction effects by study and corrected issue. Each row summarizes a different experiment (the lower two rows report the two samples which comprise study 5). Among these columns, a positive significant estimate indicates backfire. The first three columns report correction effects by ideology. The fourth column reports the total ideological effect- specifically, the absolute value of the difference between the two ends of the ideological spectrum, averaging over corrected and uncorrected respondents. Within each row, items are sorted by ideological effects. To impart a coarse sense of both effects' size: 37% of the correction effects are larger, in absolute terms, than the average ideological effect.

$$\begin{matrix} T_1 & T_2 & \cdots & T_J \\ T_2 & T_3 & & \vdots \\ \vdots & & \ddots & \\ T_J & \cdots & & T_1 \end{matrix}$$

where *j* indexes the total number of treatments in some study, the rows indicate the order in which a respondent saw each treatment, and the columns indicate the possible permutations of treatments.

One complication comes in the form that every treatment was either corrected or uncorrected. That is, each respondent was exposed to some *c* count of corrections, where *c* is an element of the set {0, 1, 2...*j*}. Section A.14.2 on p. xxix describes the tests to ensure that these elements of correction exposure did not introduce confounding related to respondents' characteristics. Section A.14.3 on p. xxx describes results which show that this design did not introduce demand characteristics or otherwise affect responses.

Study 1: Results

Subjects (n = 3127) were recruited via Amazon's Mechanical Turk platform.[3] We estimated linear models of the following form

$$\text{Agreement}_i = \beta_0 + \beta_1\ (\text{ideology}) + \beta_2\ (\text{correction}) + \beta_3\ (\text{ideology} \times \text{correction}) \tag{1}$$

where *i* indexes issues. Agreement was measured on a five point Likert scale, with larger values indicating stronger agreement. Ideology is measured on a 7 point likert scale, with larger values indicating increased conservatism. Corrections are measured with a dummy variable.[4]

Figure 2 presents the predicted values and their 95% confidence intervals from each regression model, with ideology mapped to the x-axis and separate curves for uncorrected (light gray) and corrected (dark gray) ribbons. For each issue, the average corrected subject increased their agreement with the facts. No ideological group exposed to the correction moved in the opposite direction; that is, no group demonstrated backfire.

[3] To avoid the possibility of unintended panel conditioning, we excluded any Turk worker which had been participated in a prior study.

[4] The choice of the OLS model, and the specific measures for agreement, ideology, and correction, were chosen to be consistent with Nyhan and Jason (2010).

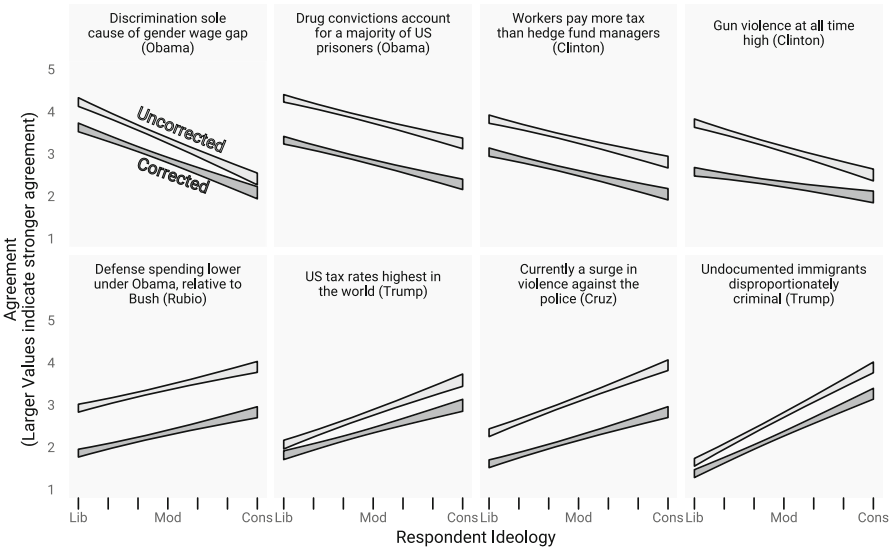


Fig. 2 Study 1 correction effects. Curves are the conditional predictions and their 95% confidence intervals drawn from the regression models described in Table 4 on p. iv. Issues are sorted by the overall relationship between ideology and agreement

The arrangement of issues within Fig. 2 is also informative. The facets are sorted by mean slope between agreement and ideology. Correction effects (depicted here as the vertical difference between ribbons) are as large as the *total* effect of ideology for certain issues. Despite what the backfire hypothesis would predict, neither issues of high nor low salience issues triggered factual backfire.

Study 2

For this study, we identified issues about which speakers from both sides of the aisle had departed from the available evidence. While the comments themselves were naturally distinct, they could both be corrected with reference to the same data. For each issue, subjects were randomly assigned to see a misstatement by either a Democratic or Republican politician. For example, in December 2015, Republican Senator Ted Cruz said the following on the subject of immigration:

> [As President], I will enforce the law. That means you stop the Obama administration's policy of releasing criminal illegal aliens. Do you know how many aliens Bill Clinton deported? 12 million. Do you know how many illegal aliens, George W. Bush deported? 10 million.

In July 2015, Democratic Representative Luis Gutierrez said the following:

> [President Obama] said he will flex his executive muscle, to be as big and as bold as he can be, to reduce deportations of undocumented immigrants...to keep families together. I saw our champion.

In the first instance, a Republican paints the president as weak on immigration enforcement; in the second, a Democrat paints the president as political ally of the immigrant community. Once again, on an issue-by-issue basis, some subjects randomly saw, or did not see, a correction. In this case, the correction read:

> In fact, according to the Department of Homeland Security, President Obama has deported illegal immigrants at twice the rate of his predecessor, President George W Bush.

Then, all subjects were asked to agree or disagree with the following, again on a five-point scale: "President Obama has been more tolerant of illegal immigration than previous presidents."

Table 5 on p. xi contains the full text of the speakers' remarks, the corrections appended, and the survey items.

Study 2: Results

Subjects (n = 2801) were recruited via Amazon's Mechanical Turk platform. Using the same model described in Eq. 1, we find that regardless of subjects' ideology, and regardless of whether the correction corrected a Democratic or Republican-aligned speaker, the average subject exposed to the correction brought their views closer in line with the facts. Figure 3 presents the predicted effect of ideology, the correction, for all 8 issues and both speaker ideologies. As with Fig. 2, darker ribbons below the lighter ribbons indicate movement toward the facts.

Of course, subjects with different political leanings responded differently to the correction provided to them. Unsurprisingly, respondents had residual reluctance to abandon co-ideologues and were generally eager to correct ideological opponents. Yet in no case did an ideological group respond to a co-ideologue being corrected by rejecting the correction.

Study 3

In the third study, our design extended that of Nyhan and Jason (2010), and concealed factual corrections within fictitious original news articles. The articles were designed to mimic an actual news article, with a visible dateline and headline. All subjects read eight articles, with each article containing an actual misstatement by politician. For each news article, some subjects were randomly assigned to view a version of the article that contained a factual correction provided by a neutral government source. For example, all subjects read a news article about Speaker Ryan's views on abortion, in which this quote by Ryan is displayed:

> Ryan's most cutting criticism, met with enthusiastic applause, was made of the President's changed policy on abortion : 'In the Clinton years, the stated goal was to make abortion 'safe, legal and rare.' Obama stands for an absolute, unqualified right to abortion—at any time, under any circumstances, and paid for by taxpayers.

Fig. 3 Study 2 correction effects. Curves are the conditional predictions and their 95% confidence intervals drawn from the regression models described in Table 6, which is on p. xii. Separate issues are included in columns, while speaker ideology is mapped to rows. Issues are sorted by the overall relationship between ideology and agreement ▶

Subjects who saw the corrected version of the article then saw the following:

> Statistics from the Center for Disease Control tell a different story. The number of abortions steadily declined during President Obama's first term, with fewer abortions in 2012 than any year since 1973.

All subjects were then asked to agree or disagree with the information provided by the correction. Section A.5 contains the full text of the articles and the survey items. Subjects in this study were also assigned to read the same news article about WMD in Iraq that Nyhan and Jason 2010 used; we describe those results in "WMD Backfire and Question Wording Effects" and "WMD Correction Results" sections.

Study 3: Results

Subjects (n = 977) were recruited via Amazon's Mechanical Turk platform. Correction effects were estimated with the same linear model depicted in Eq. 1. The predicted values for each model are depicted in Fig. 4. For every issue, the average subject who saw a newspaper article that contained a correction expressed less agreement with the factually inaccurate position, all along the ideological spectrum. On average, respondents were even willing to contradict co-ideologues, though these correction effects were smaller.[5] The effect of the correction observed in Study 3 was indeed smaller than the effects in the previous studies. This should not be surprising, as the correction itself was less conspicuous, since it was embedded in a longer body of text. In addition, conservatives were especially eager to reject liberal speakers, and conservatives were overall less responsive to corrections. Figure 4 demonstrates that the most conservative respondents were least responsive to corrections to conservatives and were most responsive to corrections of liberals.

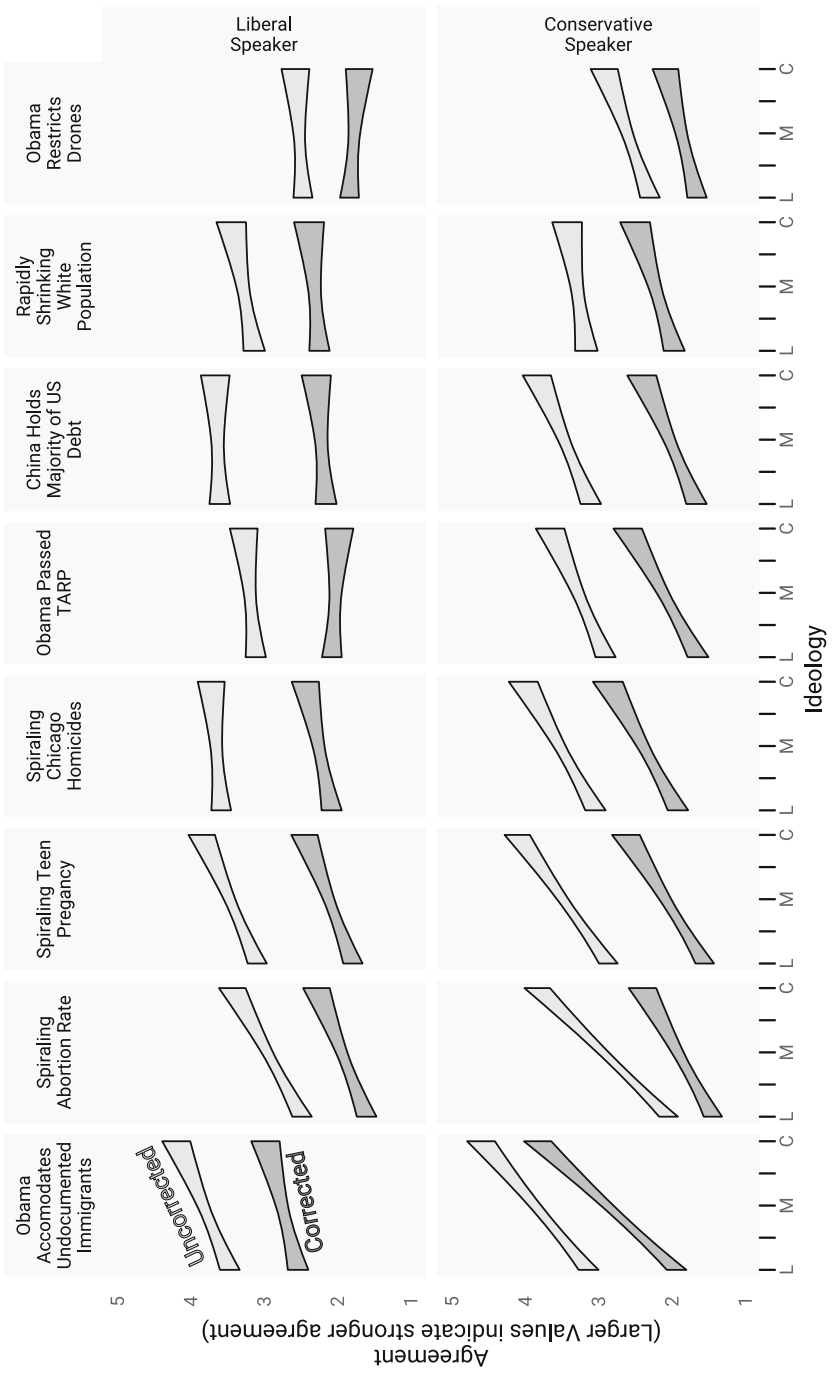
WMD Backfire and Question Wording Effects

All subjects in Study 3 were also presented the same fictitious article about WMD in Iraq used by Nyhan and Jason (2010), while some subjects also randomly shown Nyhan and Reifler's correction. However, we departed from Nyhan and Reifler in one crucial respect. To test whether backfire can originate due to question-wording, we randomly varied whether subjects were asked to agree with Nyhan and Reifler's original survey item, or a version we authored.

Nyhan and Reifler's original survey item read:

> Immediately before the U.S. invasion, Iraq had an active weapons of mass destruction program, the ability to produce these weapons, and large

[5] This relationship persisted if we compare respondents along the partisan scale. This result is described in Sect. A.14.1 on p. xxvi.


Liberal Speaker
Conservative Speaker
Obama Accomodates Undocumented Immigrants
Spiraling Abortion Rate
Spiraling Teen Pregancy
Spiraling Chicago Homicides
Obama Passed TARP
China Holds Majority of US Debt
Rapidly Shrinking White Population
Obama Restricts Drones
Uncorrected
Corrected
Agreement
(Larger Values indicate stronger agreement)
Ideology
L M C L M C L M C L M C L M C L M C L M C L M C
1 2 3 4 5

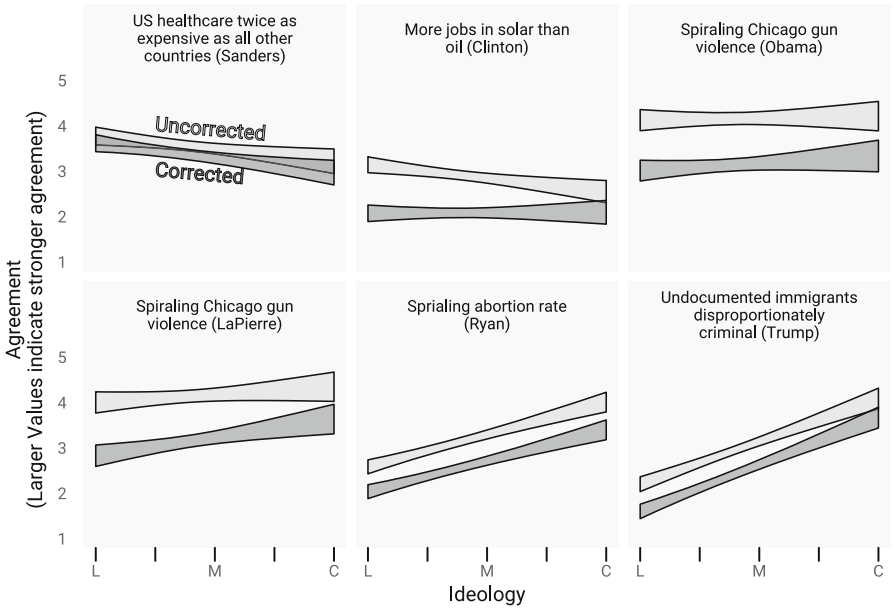


Fig. 4 Study 3 correction effects. Curves are the conditional predictions and their 95% confidence intervals drawn from the regression models described in Table 8 on p. xviii

> stockpiles of WMD, but Saddam Hussein was able to hide or destroy these weapons right before U.S. forces arrived.

Our alternative version read:

> Following the US invasion of Iraq in 2003, US forces did not find weapons of mass destruction.

The simpler item more directly taps respondents' factual understanding of the post invasion history, and does not offer parenthetical statements to measure attitudes about possible strategies Saddam Hussein might have employed to disperse or conceal WMD. The simpler version also more closely reflects the common interpretation of Nyhan and Reifler's finding. The average American voter is not expected to be familiar with the details of Iraq's prewar history. To adequately evaluate President Bush, and his undertaking to go to war to mitigate the threat posed by Iraqi WMD, citizens should be aware that no weapons were found. Our simpler item directly measures this factual understanding.[6]

WMD Correction Results

The WMD survey item to which subjects were exposed strongly conditioned their level of factual backfire. All respondents presented with Nyhan and Reifler's version

[6] Of course, the attitudinal consequence of this fact remains at a respondent's discretion, but functional democratic competence would seem to require that voters adopt a common set of basic political facts.

ignored the empirical correction, inferring nothing from the factual intervention (partially replicating the original finding.) Liberals presented with our simpler survey item adopted the factual correction. In Fig. 5, we plot the marginal effects of the correction, for both survey items. Respondent ideology appears on the x-axis. Effects above the red line indicate backfire—subjects rejecting the correction—while effects below the red line indicate uptake of the correction.

To understand why the two items yielded divergent effects, consider what distinguishes them. Both items relate to one overarching fact: despite the pledges of the Bush Administration, no WMD were found in Iraq. Our statements simply asks subjects to agree or disagree with this fact. In contrast, Nyhan and Reifler's item presents multiple ways to account for the failure to find WMD. A subject's understanding of WMD's absence in Iraq appears to have limited bearing on this item—instead, respondents may have interpreted the question as an invitation to appraise the war in general.

More broadly, the sheer number of facts in the Nyhan and Reifler version might have overwhelmed respondents, causing them to fall back with their ideological cohort, in which they had greater confidence. Imagine a conservative subject who understands that no WMD were found, but has no beliefs about the other facts that Nyhan and Reifler's item conveys. When presented with the WMD correction (and its implied criticism of her co-ideologue President) she might be pushed to adopt a *policy* position she believes is ideologically consistent.

Study 4

Might the complexity of survey items account for the incidence of factual backfire? That is, when shown a question which directly tests their factual understanding, respondents might privilege available facts. When shown a more convoluted item, respondents might weigh facts against their ideological priorities. This pattern would both account for previous findings and illuminate the role of factual interventions in conditioning attitudes.

To test this possibility, respondents were shown six mock newspaper articles, with our usual practice of randomizing article order and corrections.[7] Using Nyhan and Reifler's original WMD item as a model, we wrote comparably complex items for the other issues, along with moderate and simple versions. Each respondent saw one item per issue. To test the effect of item complexity, we used four issues by conservatives, and two by liberals, including a tax claim by Donald Trump that was comparable to the those which generated backfire in the original Nyhan and Reifer piece.

[7] Three articles were taken from study 3: the original Bush WMD article, the piece by Speaker Paul Ryan criticizing President Obama's policy toward abortion, and Secretary Hillary Clinton claim that twice as many Americans were employed in solar than in the oil industry. Three novel mock articles were also provided: Senator Sanders claiming that the EPA had found fracking was responsible for polluting water supplies, Donald Trump claiming that his tax cut plan would grow federal tax receipts, and Trump claiming that the true unemployment rate was actually higher than 30%. These mock articles can be read in Sect. A.9, which can be found in the appendix on p. xvi. The items can be read in Table 11 on p. xxiii.

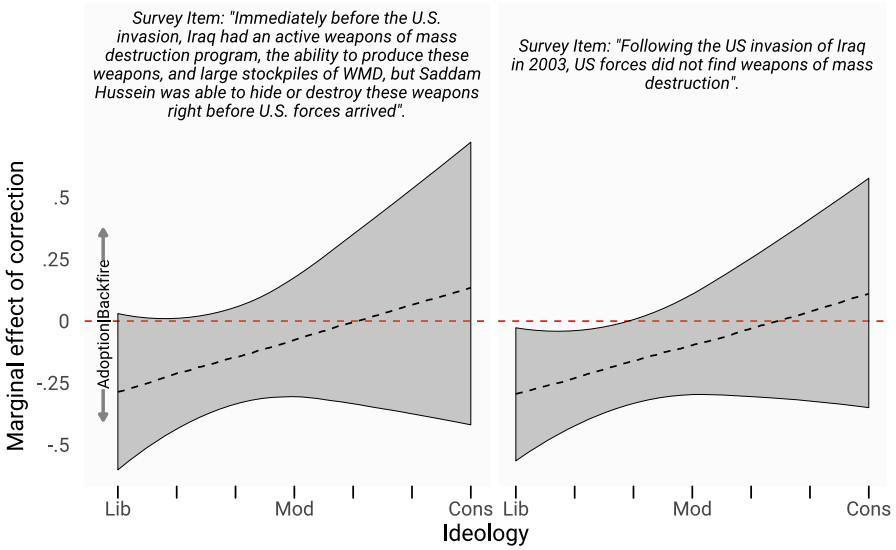


Fig. 5 Ideology and factual backfire to WMD discovery, by survey item wording. Longer survey item is from Nyhan and Jason 2010. Both groups were otherwise provided identical articles and corrections

Table 11 (found in Appendix, on p. xxiii) describes all 18 items. For instance, to measure agreement with Donald Trump's claim that the real unemployment rate was greater than 30%, the three items were:

Simple The true unemployment rate is greater than 30%.
Moderate After removing the effects of politicians interfering with the data, the true unemployment rate is greater than 30%.
Complex The unemployment rate has important political ramifications, and government statisticians are susceptible to threats and influence. After removing the effects of politicians interfering with the data, the true unemployment rate is greater than 30%.

Using Nyhan and Reifler's original WMD item as a model, a more complicated survey items offered preambulatory explanations for the fact which comprised the question's substance. Moderate items provide fewer explanations. All the items can be read in Table 11 on p. xxiii.

Study 4: Results

1333 respondents were recruited from Amazon's Mechanical Turk. Among the 18 combinations of survey item complexity and corrections, no backfire was observed. Table 9 on p. xix reports the coefficients for the linear models using the common specification (described in Eq. 1.) Of central importance to this study, however was the *overall* effect of survey item complexity on factual adoption, averaging over

items featuring liberal and conservative speakers. To test the overall effect, we estimate a multilevel model of the following type:

$$\begin{aligned} \textit{Response Level}: \quad & \text{agreement}_{i,j} = \beta_{0,j} + \beta_1\,(\text{correction}) \times \beta_2\,(\text{ideology}_i) \\ & \times \beta_3\,(\text{speaker ideology}_i) \times \beta_4\,(\text{survey item complexity}_k) \\ \textit{Issue Level}: \quad & \beta_{0,j} = U_{01} \end{aligned} \tag{2}$$

In other words, the model features a four-way interaction between the continuous indicators for complexity and respondent ideology, and dichotomous indicators for a correction and the perceived ideology of the speaker,[8] with the U_{01} being a random issue-specific intercept. The model estimates for this equation are provided in Table 10 on p. xx.

Figure 6 shows the marginal effects of each correction, grouping the ideology of speakers to the facet columns, and survey item complexity mapped to the row facets. Ribbons above the dashed horizontal indicate factual backfire. Among the six item × complexity combinations, none generated backfire. *Partially* consistent with example of WMD backfire from the original Nyhan and Reifler paper was the pattern of results observed among conservative respondents. The right hand facet in Fig. 6 demonstrates that conservatives are far less likely to demonstrate factual adherence when evaluated with a complex survey item. This suggests the initial backfire finding is partially a consequence of the original WMD question wording. Liberals do not demonstrate comparable variation in factual adherence in response to survey item complexity.

These items provide what should be an even easier test for backfire, by loading measures of factual attitudes with preambulatory clauses that should crowd out the correction's influence. Complicated items also seem prone to induce factual counter-arguing by listing separate ways to account for some outcome. This study also tested two items from Donald Trump, a candidate whose rejection of fact checkers and the judgment of the national political establishment is central to his appeal. Yet when Trump alleges a complicated political conspiracy to hide the true extent of unemployment, a correction which cites the very federal agency implicated by Trump (the Bureau of Labor Statistics) significantly improves our subjects' accuracy. The influence of facts even in these circumstance invites us to imagine the exceptional circumstances needed to engender factual backfire.

Study 5

Our results indicate an unexpected level of factual responsiveness. Might the peculiar qualities of the Mechanical Turk population be responsible? Unlike the undergraduates who comprised the original Nyhan and Jason (2010) paper or the opt-in panel of knowledgeable Palin supporters who backfired when presented with

[8] For this study, President Obama, Secretary Clinton, and Senator Sanders are deemed liberal speakers, and President Trump and President Bush are deemed conservative speakers.

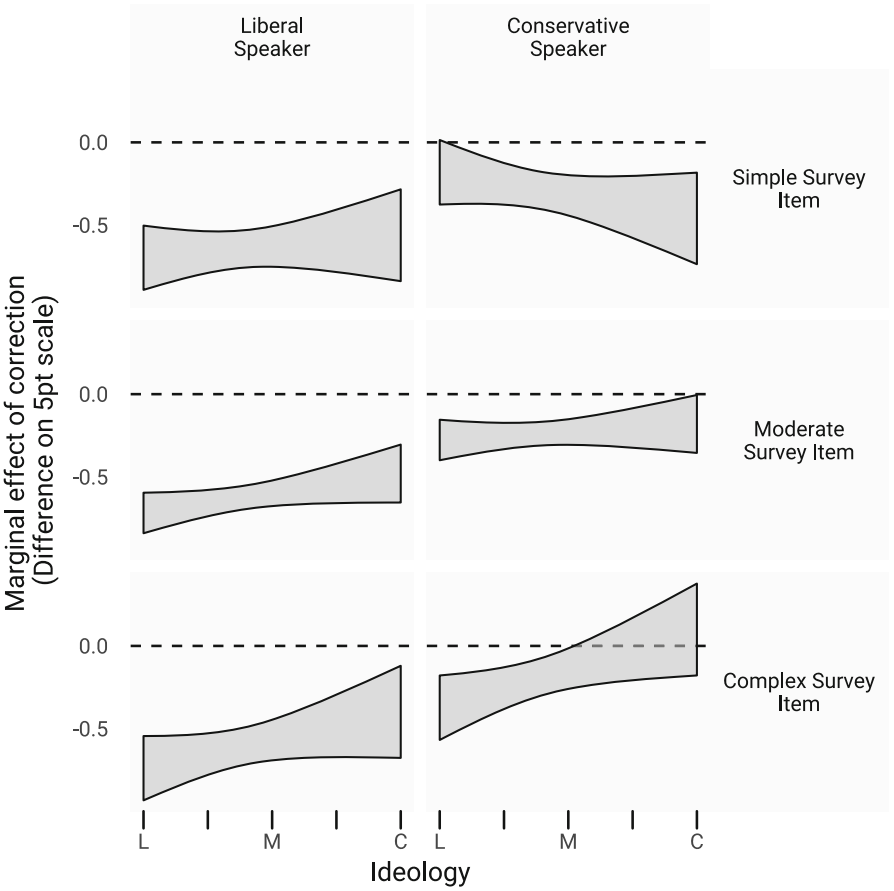


Fig. 6 Study 4's multilevel model tests the effect of survey item complexity on factual adherence. Ribbons show estimated marginal effects of the correction and their 95% confidence interval from the multilevel model in Table 10, on p. xx. Column facets show the separate effect of speaker ideology, and the x-axis shows the effect of respondent ideology

factual information about so-called "death panels" (Nyhan et al. 2013), MTurk workers might be more akin to those samples respondents who largely adopt factual information.[9] While the original backfire paper presents results from a small, convenience sample of undergraduates (who are generally be the least externally valid sample for the study of political attitudes; see Henrich et al. 2010), MTurk respondents might also be comparably biased on factual matters. For instance, given their habit of conducting multiple scientific surveys, these respondents might perceive factual responsiveness as a precondition for successfully completing a survey, and might therefore feign factual adherence due to financial motivations.

[9] For instance, the national representative panel who adopted the correction that the flu vaccines did not induce flu infections (Nyhan et al. 2015) or the national representative panel who accepted the correction that the MMR vaccines did not cause autism (Nyhan et al. 2014).

Similarly, given the well-publicized difficulty in sampling conservatives on the MTurk platform, liberal Turkers might masquerade as conservatives to participate in more studies, meaning these respondents do not reflect the typical conservative participant in a probability sample.

To address this possibility, we provide the same six statements and corrections to two samples drawn simultaneously from MTurk and Lucid, an online panel provider. Lucid draws from 30 separate survey recruitment firms, and through repeated interaction with survey respondents, allows experimental recruitment to match nationally representative survey margins.[10]

For this study, 995 respondents were drawn from the Lucid and 1024 respondents were drawn from Turk, over 6–8 June 2017. Table 2 reports the respective compositions of the Lucid and MTurk samples. Demographically, the nationally representative sample redresses those areas where Turk traditionally struggles—its sample is older, less educated, and more conservative.

Study 5: Results

Table 13 on p. xxviii reports the correction results by issue and sample using model 1. For all issues, no backfire was observed. Of central importance is the *difference* in factual adherence between Turk respondents and the nationally representative Lucid sample. To test this, we estimate a multilevel model described in Eq. 3, where i indexes respondents and j indexes issues, so that $\text{agreement}_{i,j}$ indicates the agreement of the ith respondent to the jth issue.

$$\begin{aligned} \textit{Response Level :} \quad & \text{agreement}_{i,j} = \beta_{0,\text{j}} + \beta_1\,(\text{correction}_\text{j}) \times \beta_2\,(\text{ideology}_\text{i}) \times \beta_3\,(\text{sample}_\text{i}) \\ \textit{Issue Level} \quad & \beta_{0,\text{j}} = \text{U}_{01} \\ & \beta_{2,\text{j}} = \text{U}_{21} \end{aligned} \tag{3}$$

That is, this model allows the intercept and the slope of the ideological response to vary at random between issues—to abstract away the variation in factual response that is apparent due to evaluations of an issue or a speaker. Of key importance is the

[10] Coppock and McClellan (2017) report an extensive test of the Lucid sample, comparing it to Turk, the Census Bureau's Current Population Survey, and the American National Election Study's (ANES) face to face and online samples. Treating the ANES face-to-face sample as the "gold standard", the Lucid sample is more psychologically similar to the ANES than the Turk sample on the Big-5 personality battery, and better matches the political knowledge and conservatism in the ANES. Coppock and McClellan also test the Lucid sample's ability to recover treatment effects in canonical social psychology experiments. Both Lucid and Turk samples recover the same framing effect observed the General Social Survey (a massive face-to-face survey instrument), improving the appetite for public spending when it is described as "assistance to the poor" or "caring for the poor" rather than "welfare." Both Lucid and Turk feature the same framing effect underpinning prospect theory [(the famous Tversky and Kahneman (1983) finding which shows risk tolerance is affected by framing possible outcomes as gains or losses.] Both Lucid and Turk recover indistinguishable experimental effects as observed in Hiscox (2006) in framing attitudes about free trade. Most importantly for this study, the one failed replication was on rumor corrections in the aforementioned Berinsky paper (2017), where Lucid respondents were unusually *resistant* to corrective information. This suggests that the Lucid sample is at least a comparably demanding sample in which to test factual adherence.

Table 2 Studies 5 sample composition, mechanical Turk and Lucid samples

	Age				Education				Employment				Gender	
	18–24	25–44	45–64	65+	HSD/less	Some coll.	BA degree	Adv. deg.	Employed	S.employ	Unemp.	Ret./oth.	Female	Male
Lucid	12	39	36	13	27	37	25	11	48	10	10	31	69	31
Turk	9	69	20	2	13	37	39	11	69	18	5	8	48	52

	Ideology			Income				Race				Region			
	Lib	Mod	Cons	< $45k	$45–75k	$75–120k	> $120k	White	Black	Hisp	Other	N. centl	N.east	South	West
Lucid	23	54	23	53	20	19	8	78	8	7	7	23	19	39	19
Turk	42	43	15	47	25	20	7	78	7	6	9	24	21	43	13

Each row reports the sample percentage within that variable

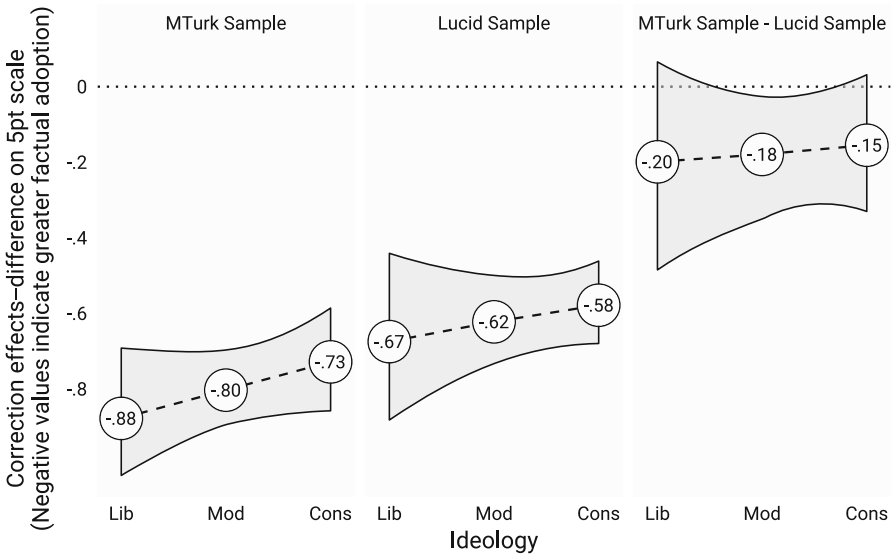


Fig. 7 Correction effects by sample. Ribbons indicate 95% confidence intervals, taken from Eq. 3. The third facet shows the 95% confidence interval on the difference in the respective samples' correction effects, indicating that moderate Turk respondents are marginally more factually adherent, but that those at the end of the ideological spectrum are not significantly different between Turk and the Lucid sample

interaction of the β_1 and β_3 terms, to estimate if MTurk or nationally representative respondents are more respondent to corrections. Figure 7 reports the difference between corrected and uncorrected respondents, by sample, in the first two facets. The third facet reports the *difference* in these two samples' correction effects. On average, Turk respondents are slightly more factually responsive, but these differences are insignificant at either end of the ideological spectrum (precisely where Nyhan and Reifler saw backfire in their 2010 paper.) Only among ideological moderates do we observe that opt-in respondents are meaningfully different than the nationally representative respondents.

Does Counterargument Explain Our Pattern of Findings?

As we explained in "Which Corrections are Prone to Counter-Argument?" section, the two dominant theories of counterargument would predict either a positive or a negative relationship between correction accordance and correction size.[11] To measure this directly, we had 261 respondents[12] evaluate all our issue/correction pairs on the "connection" or "relatedness" of the correction and the statement—whether the a correction directly contradicted the factual basis of a misstatement, or

[11] In brief—a weak correction might inadvertently advertise the weakness of the corrective case, or a strong correction might have more obvious factual implications, and therefore inspire more forceful counterargument.

[12] These respondents were recruited on Mechanical Turk.

if it was only tangentially related to a misstatement's factual implication or insinuation. This allowed the respondent to assess the objective proximity of a misstatement and a correction, rather than using their ideology to assess ideologically welcome corrections as strong and ideologically opposed corrections as weak.[13] Immediately after they saw the statement and its correction, they were shown the following two statements:

> We're not asking if the **statement is wrong**, or **if the correction is right**.
> We're just asking how **closely related** they are.
> Is the factual correction above **unrelated**, **partially related**, or **closely related** to the political misstatement?

Each respondent was then shown a slider along a 100 point scale, with the bold labels in the sentence above at the 0, 50, and 100 positions. Figure 8 depicts the distribution of the sample means of the evaluations, with the corrections tested by Nyhan and Reifler depicted in red. This figure amply demonstrates the advantage of testing a large number of corrections—the Nyhan and Reifler study was only able to test three corrections which cluster in the middle of the accordance range, while we're able to study highly accordant corrections and those only vaguely related.[14]

Figure 8 demonstrates that accordance has no ostensible relationship with the scope of correction effects. Corrections deemed to accord *very* closely to the misstatements[15] and those deemed a distant correction[16] had about the same size correction. Since accordance and corrections effects are unrelated, it seems unlikely that counter-argument is an apt model for the average American's psychological response to factual correction. Rather, respondents behave as if they regard facts as banal—the cost of adopting facts is minimal, even when they're ideologically unwelcome.

[13] As a robustness check—there was no significant relationship between ideology and perceived accordance, for any of the tested pairs.

[14] It's instructive to consider those statement/correction pairs at either end of this spectrum. The statement by Senator Ted Cruz about the incidence of violence targeted at law enforcement, described above, was judged the most proximate correction. At the other end of this continuum is the 2012 claim by Congressman Paul Ryan that "Obama stands for an absolute, unqualified right to abortion—at any time, under any circumstances, and paid for by taxpayers" and the correction that "The number of abortions steadily declined during President Obama's first term, with fewer abortions in 2012 than any year since 1973." While Cruz makes a precise claim about the *change* in the incidence of killings of police officers, Ryan's statements merely suggested a spike in the incidence of abortion.

[15] An example of a *proximate* correction is Representative Gutiérrez's promise that President Obama would be the "champion...[of the] undocumented" paired with the evidence that Obama was a prodigious deporter of these residents. This correction/statement pair was scored 81.7 on a 100-pt scale of accordance.

[16] An example of a *distant* correction is Governor Romney's description of the United States using "a credit card ...issued by the Bank of China" and the correction that China holds about 15% of US debt. This correction/statement pair was scored 47.2.

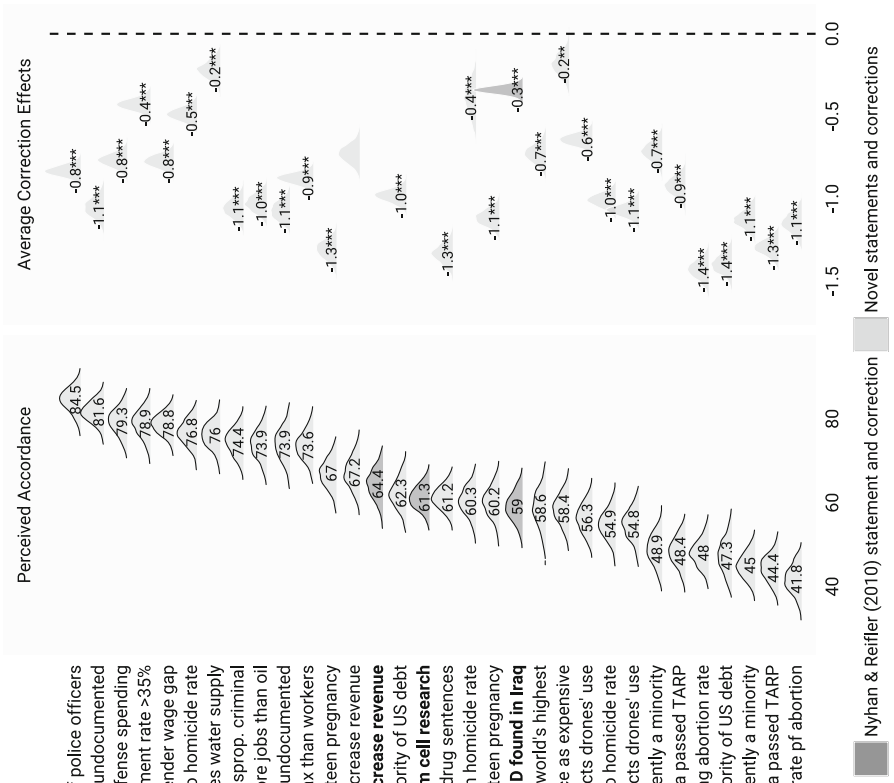


Fig. 8 Bootstrapped distribution of mean proximity between corrections and statements. Density plots in the left hand facet show the distribution of perceived accordance between correction and mis-statement pairs. Density plots in the right hand facet show the mean correction effect pair (averaging over the ideological cohorts). Asterisks indicate the significance of the correction. For reference, the statement/correction pairs tested by Nyhan and Jason (2010) are depicted in bold face. Compared to Nyhan and Reifer, we test both more and less accordant misstatment-correction pairs

Robustness Checks

To test if design effects reduces our estimates' validity, we performed three sets of robustness checks. First, to determine if distinct correction effects would be apparent if we conditioned on partisanship, we estimated every model separately using either ideology or partisanship as a continuous predictor interacted with the correction indicator (Wood and Oliver 2012). We then estimated the 95% confidence interval for the *difference* in fitted correction effects, given each predictor type. Neither partisanship nor ideology was able to elicit factual backfire, nor was there a substantive change in the pattern of effects. This analysis is discussed in Sect. A.14.1, and the results are presented in Fig. 9. Second, we checked the relationship between patterns of correction exposure and respondents' characteristics, to ensure that the randomization was successful. No observed characteristic predicted the number of corrections to which a respondent was exposed, or the maximum length of consecutive corrections, or the number of times a respondent observed three or more successive corrections. These results are presented in Table 14. Finally, we checked the relationship between patterns of correction exposure and our estimated correction effects, to test for the possibility of demand effects: specifically, that respondents would become aware that we were testing their factual receptivity after they had a seen a number of corrected and uncorrected statements. We find that corrections are about as large for the first issue presented to respondents as they are for the final issue. These results are presented in Fig. 10.

Discussion

We find that backfire is stubbornly difficult to induce, and is thus unlikely to be a characteristic of the public's relationship to factual information. Overwhelmingly, when presented with factual information that corrects politicians—even when the politician is an ally—the average subject accedes to the correction and distances himself from the inaccurate claim.

Our findings are consistent with one of the most well-documented aspects of mass public opinion: respondents shy away from cognitive effort, and will deploy shrewd strategies to avoid it (Lippmann 1922). The backfire hypothesis proposes that a subject, when furnished facts inconsistent with her ideological commitments, will resolve the challenge of these facts by concocting new considerations to offset the threatening information. Developing counter-arguments would be unusually effortful, as sophisticated respondents can simply filter out, rather than counterargue, unwelcome facts (Zaller 1992). If indeed subjects who backfire are counterarguing,[17] it is worth recalling that many of the samples in which backfire has been observed were gathered in university settings. Social psychologists have long known that students are unusually inclined toward cognitive effort (Cacioppo et al. 1983). Undergraduate subjects may therefore be more prone to displaying backfire,

[17] Contra our evidence in "Does Counterargument Explain Our Pattern of Findings?" section.

and these survey compositional factors might account for the previous findings of backfire.

Other findings that paint citizens as alarmingly ill-equipped for democracy have suffered from similar sample selection issues. For example, by one accounting, voters factor in local sports scores when deciding to reelect an incumbent (Healy et al. 2010). More recent research has challenged this finding (Fowler and Montagnes 2015). Research that claims to show widespread democratic incompetence may mistake the snapshot that any one study represents for the sum total of citizens' abilities. Our findings are not without their own limits. That voters acquiesce to facts presented to them does not mean that they have retained this information; rather, the facts we provided may quickly become inaccessible. Yet they do not reflexively reject that information—and plainly, they do not go to the effort of compounding inaccurate beliefs, as the backfire hypothesis would predict. Far from it: for about nine issues in ten, factual information significantly improves the average respondent's accuracy. At least for a brief moment, their perceptual screens dim, and the facts prevail.

Acknowledgements The authors would like to thank Leticia Bode, John Brehm, DJ Flynn, Jim Gimpel, Don Green, Will Howell, David Kirby, Michael Neblo, Brendan Nyhan, Gaurav Sood, and the participants at the Center for Stategic Initiatives workshop. Research support was generously furnished by the Cato Institute, and we owe a special debt of gratitude to Emily Ekins and David Kirby. All remaining errors are the responsibility of the authors.

References

Barnes, L., Feller, A., Haselswerdt, J., & Porter, E. (2016). Information and preferences over redistributive policy: A field experiment. Working Paper.

Berinsky, A. J. (2017). Rumors and health care reform: Experiments in political misinformation. *British Journal of Political Science, 47*(2), 241–262. https://doi.org/10.1017/S0007123415000186.

Brock, T. C. (1967). Communication discrepancy and intent to persuade as determinants of counterargument production. *Journal of Experimental Social Psychology, 3*(3), 296–309. https://doi.org/10.1016/0022-1031(67)90031-5.

Bullock, J. G., Gerber, A. S., Hill, S. J., & Huber, G. A. (2013). Partisan bias in factual beliefs about politics. Working Paper, March 2013, pp. 1–73. https://doi.org/10.1561/100.00014074.

Cacioppo, J. T., Petty, R. E., & Morris, K. J. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology*, *45*(4), 805–818. https://doi.org/10.1037/0022-3514.45.4.805.

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1960). *The American Voter*. New York: Wiley.

Carmines, E., & Stimson, J. (1980). The two faces of issue voting. *American Political Science Review, 74*(1), 78–91. https://doi.org/10.2307/1955648.

Chong, D., & Druckman, J. (2013). Counterframing effects. *The Journal of Politics, 75*(1), 1–16. Retrieved from http://www.journals.uchicago.edu/doi/abs/10.1017/S0022381612000837.

Cochran, W. G., & Cox, G. M. (1957). *Experimental designs*. New York: Wiley.

Conover, P. J., & Feldman, S. (1981). The origins and meaning of liberal/conservative self-identifications. *American Journal of Political Science, 25*(4), 617. https://doi.org/10.2307/2110756.

Converse, P. E. (2006). The nature of belief systems in mass publics (1964). *Critical Review: A Journal of Politics and Society*, *18*(1–3), 1–74. https://doi.org/10.1080/08913810608443650.

Coppock, A., & Mcclellan, O. A. (2017). Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. Retrieved from http://alexandercoppock.com/papers/CM_lucid.pdf.

Fishkin, J. S., & Luskin, R. C. (2005). Experimenting with a democratic ideal: Deliberative polling and public opinion. *Acta Politica*, *40*, 284–298. https://doi.org/10.1057/palgrave.ap.5500121.

Fowler, A., & Montagnes, B. P. (2015). College football, elections, and false-positive results in observational research. *Proceedings of the National Academy of Sciences*, *112*(45), 13800–13804.

Gerber, A., & Green, D. (1999). Misperceptions about perceptual bias. *Annual Review of Political Science*, *2*, 189–210. https://doi.org/10.1146/annurev.polisci.2.1.189.

Gollust, S. E., Lantz, P. M., & Ubel, P. A. (2009). The polarizing effect of News Media messages about the social determinants of health. *Public Health, 99,* 2160–2167. https://doi.org/10.2105/AJPH.2009.161414.

Healy, A., Malhotra, N., & Mo, C. (2010). Irrelevant events affect voters' evaluations of government performance. *Proceedings of the National Academy of Sciences, 107*(29), 12804–12809. Retrieved from http://www.pnas.org/content/107/29/12804.short.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The Weirdest people in the world?'. *The Behavioral and Brain Sciences*, *33*(2–3), 61–135. https://doi.org/10.1017/S0140525X0999152X.

Hiscox, M. J. (2006). Through a glass and darkly: Attitudes toward international trade and the curious effects of issue framing. *International Organization*. Cambridge University Press. http://doi.org/10.1017/S0020818306060255.

Howell, W. G., & West, M. R. (2009). Educating the public. *Education Next*, *9*(3), 40–47.

Jost, J. T., Nosek, B. A., & Gosling, S. D. (2008). Ideology: Its resurgence in social, personality, and political psychology. *Perspectives on Psychological Science, 3*(2), 126–136. https://doi.org/10.1111/j.1745-6916.2008.00070.x.

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened selfgovernment. *Behavioural Public Policy, 1*(01), 54–86. https://doi.org/10.1017/bpp.2016.2.

Kaplan, J. T., Gimbel, S. I., & Harris, S. (2016). Neural correlates of maintaining one's political beliefs in the face of counterevidence. *Scientific Reports, 6*(1), 39589. https://doi.org/10.1038/srep39589.

Kuklinski, J. H., & Quirk, P. J. (2000). Reconsidering the rational public: Cognition, heuristics, and mass opinion. In *Elements of reason: Cognition, choice and the bounds of rationality* (pp. 153–182).

Lippman, W. (1922). *Public opinion*. New York: Harcourt, Brace and Company.

Lord, C., Ross, L., & Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social*. Retrieved from http://psycnet.apa.org/journals/psp/37/11/2098/.

Mondak, J. J. (1993). Public opinion and heuristic processing of source cues. *Political Behavior, 15*(2), 167–192. Retrieved from http://www.jstor.org/stable/586448.

Mondak, J. J. (1994). Policy legitimacy and the Supreme Court: The sources and contexts of legitimation. *Political Research Quarterly, 47*(3), 675. https://doi.org/10.2307/448848.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2.

Nyhan, B., & Reifler, J. (2017). Answering on cue? How corrective information can produce social desirability bias when racial differences are salient. Retrieved from http://www.dartmouth.edu/~nyhan/obama-muslim.pdf.

Nyhan, B., Reifler, J., Edelman, C., Passo, W., Banks, A., Boston, E., et al. (2015). *Answering on cue?*. Hanover, NH: Dartmouth College.

Nyhan, B., Reifler, J., Richey, S., & Freed, G. L. (2014). Effective messages in vaccine promotion: A randomized trial. *Pediatrics*, *133*(4), e835–e842. https://doi.org/10.1542/peds.2013-2365.

Nyhan, B., Reifler, J., & Ubel, P. A. (2013). The hazards of correcting myths about health care reform. *Medical Care*, *51*(2), 127–132. https://doi.org/10.1097/MLR.0b013e318279486b.

Prior, M. (2007). Is partisan bias in perceptions of objective conditions real? The effect of an accuracy incentive on the stated beliefs of partisans. In *Annual conference of the Midwestern Political Science Association*.

Prior, M., Sood, G., & Khanna, K. (2015). The impact of accuracy incentives on partisan bias in reports of economic perceptions. *Quarterly Journal of Political Science*, *10*, 489–518. https://doi.org/10.1561/100.00014127.

Redlawsk, D. P. (2002). Hot cognition or cool consideration? Testing the effects of motivated reasoning on political decision making. *The Journal of Politics, 64*(4), 1021–1044. https://doi.org/10.1111/1468-2508.00161.

Sanna, L. J., Schwarz, N., & Stocker, S. (2002). When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *Journal of Experimental Psychology: Learning Memory and Cognition, 28*(3), 497–502. https://doi.org/10.1037//0278-7393.28.3.497

Schaffner, B. F., & Roche, C. (2017). Misinformation and motivated reasoning: Responses to economic news in a politicized environment. *Public Opinion Quarterly*. https://doi.org/10.1093/poq/nfw043.

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research, 31*(4), 713–724. https://doi.org/10.1086/426605.

Sniderman, P. M., Brody, R. A., & Tetlock, P. E. (1993). *Reasoning and choice: Explorations in political psychology*. Cambridge University Press.

Stroud, N. J. (2008). Media use and political predispositions: Revisiting the concept of selective exposure. *Political Behavior, 30,* 341–366. https://doi.org/10.2307/40213321.

Swire, B., Berinsky, A. J., Lewandowsky, S., & Ecker, U. K. H. (2017). Processing political misinformation: Comprehending the Trump phenomenon. *Royal Society Open Science, 4*(3), 160802. https://doi.org/10.1098/rsos.160802.

Taber, C. S., & Lodge, M. (2006). Motivated specticism in the evaluation of political beliefs. *Political Science, 50*(3), 755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x.

Thorson, E. (2015). Belief echoes: The persistent effects of corrected misinformation. *Political Communication*, *46*(9), 1–21. https://doi.org/10.1080/10584609.2015.1102187.

Trevors, G. J., Muis, K. R., Pekrun, R., Sinatra, G. M., & Winne, P. H. (2016). Identity and epistemic emotions during knowledge revision: A potential account for the backfire effect. *Discourse Processes*. https://doi.org/10.1080/0163853X.2015.1136507.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuititive reasoning: The conjuction fallacy in probability judgment. *Psychological Review, 90*(4), 293–315. http://psycnet.apa.org/journals/rev/90/4/293/.

Wood, T., & Oliver, E. (2012). Toward a more reliable implementation of ideology in measures of public opinion. *Public Opinion Quarterly*, *76*(4), 636–662. https://doi.org/10.1093/poq/nfs045.

Zaller, J. R. (1992). *The nature and origins of mass opinion*. New York: Cambridge University Press.