

The Exception, Not the Rule?

The Rarely Polarizing Effect of Challenging Information*

Andrew Guess[†]
New York University

Alexander Coppock[‡]
Yale University

August 24, 2016

Abstract

Several prominent theoretical perspectives suggest that when individuals are exposed to counter-attitudinal evidence or arguments, their preexisting opinions and beliefs are reinforced, resulting in a phenomenon known as “backlash,” “backfire,” or “boomerang.” We investigate the prevalence of this effect. Should we expect that all attempts to persuade those who disagree will backfire? First, we formalize the concept of backlash and specify how it can be measured. We then present results from three survey experiments—two on Mechanical Turk and one on a nationally representative sample—in which we find no evidence of backlash, even under theoretically favorable conditions. While a casual reading of the literature on partisan information processing would lead one to conclude that backlash is rampant, we suspect that it is much rarer than commonly supposed. Researchers should continue to design well-powered randomized studies in order to better understand the specific conditions under which backlash is most likely to occur.

Prepared for the 2016 American Political Science Association Annual Meeting, Philadelphia

*The authors would like to thank David Kirby as well as the students in Columbia’s Political Psychology graduate seminar for their insight and support as this project developed. We thank Time-sharing Experiments for the Social Sciences for supporting Study 3. Deanna Kuhn and Joseph Lao generously shared copies of original experimental materials with us for Study 1. We benefited greatly from discussions at the New York Area Political Psychology Meeting in November 2014 and the Dartmouth Experiments Conference in July 2016. Study 1 and Study 2 were reviewed and approved by the Institutional Review Board of Columbia University (IRB-AAAN5213), and Study 3 was approved by both NYU (IRB-FY2016-639) and Columbia (IRB-AAAQ7729) IRB. We pre-registered the designs and intended analyses of Study 1 and Study 3 with EGAP: Experiments in Governance and Politics before primary data collection began. See: <http://e-gap.org/design-registration/registered-designs/>.

[†]Postdoctoral Fellow, Social Media and Political Participation (SMaPP) Lab, New York University. Email: guess@nyu.edu

[‡]Assistant Professor, Department of Political Science, Yale University. Email: alex.coppock@yale.edu

If people of opposing views can each find support for those views in the same body of evidence, it is small wonder that social science research ... will frequently fuel rather than calm the fires of debate.

—Lord, Ross, and Lepper (1979)

1 Introduction

For several decades, research on public opinion and information processing has presented a challenge for believers in evidence-based decision-making. Its prognosis for the body politic is dire: Instead of prompting a reconsideration of long-held views, counter-attitudinal evidence may actually strengthen preexisting beliefs, resulting in polarization. This prediction of *backlash* is associated with numerous theoretical perspectives and has led to an emerging consensus among scholars that attempts to persuade voters, challenge opponents, or correct factual misperceptions can often lead to the opposite of the intended effect.

According to this consensus, people work—consciously or not—to protect their worldviews via a series of complementary belief-preserving mechanisms (Kunda 1990). Examples include the *prior attitude effect*, the tendency to perceive evidence and arguments that support one’s views as stronger and more persuasive than those that challenge them; *disconfirmation bias*, in which people exert effort to counter-argue vigorously against evidence that is not congruent with their beliefs; and various forms of selective exposure and selective attention to congenial information, sometimes referred to as *confirmation bias* (Taber and Lodge 2006). The cumulative effect of these mechanisms is polarization: People exposed to the same information may respond by strengthening their preexisting views. For some subgroups, this polarization will manifest as backlash.

The canonical explication and demonstration of these mechanisms is Lord, Ross and Lepper (1979), in which both pro- and anti-death-penalty college students were exposed to mixed scientific evidence on the effectiveness of capital punishment on crime deterrence. To their surprise, the authors found that the subjects did not moderate their views; rather, those who initially supported the punishment became more pro-capital punishment on average by the end of the study, and those who opposed it became less so. This study helped inspire a

research agenda spanning psychology (Miller, McHoskey, Bane and Dowd 1993; Kuhn and Lao 1996), political science (Redlawsk 2002; Taber and Lodge 2006; Lau and Redlawsk 2006; Taber, Cann and Kucsova 2009; Nyhan and Reifler 2010), and other fields such as public health (Strickland, Taber and Lodge 2011; Nyhan, Reifler, Richey and Freed 2014). Backlash is one of the phenomena frequently invoked in this literature. How prevalent is it?

2 Expectations of Backlash

To our knowledge, the first appearance of backlash effects in the literature occurs in Lazarsfeld, Berelson and Gaudet (1944), who observed “several boomerangs upon people who resented what they read or heard and moved in the opposite direction from that intended” (p. 154). The authors note this in passing while explaining the reasons why personal contact may be more effective—and less likely to produce such “boomerangs”—than media messages.

Since then, a number of distinct theories have accommodated the possibility of backlash during the opinion formation process. In John Zaller’s **Receive-Accept-Sample model** (1992), opinions are generally stable. Depending both on individuals’ level of political awareness and the partisan mix of elite communications, people will exhibit varying levels of receptivity to new information. Those least likely to be swayed one way or another are people with low awareness (who are not likely to be exposed to political messages at all) and high awareness (who possess sufficient knowledge and sophistication about political issues to successfully avoid or resist contrary messages). Only those in between have the opportunity to both receive *and* accept new political messages, the balance of which will depend on the level of elite consensus. When confronted with information about an issue, people then sample from the “top of the head” considerations they have accumulated over time.

It is clear how such a model is consistent with individual-level stability of attitudes: new information is assimilated only under certain conditions, and even then is brought to conscious awareness only when required by a survey response or interpersonal context. And when new information *can* be absorbed, the predicted result is attitude change in the direction of the overall balance of arguments made by elite political actors. The model is generally compatible

with growing extremity in attitudes over time, particularly among high-awareness individuals who only seek out information that reinforces their existing considerations. But at the same time, it is also consistent with limited predictions of backlash: In a survey context, confrontation with challenging information may cause highly informed individuals to bring counterarguments to mind, creating a mix of considerations more hostile to that perspective (e.g., Lord, Ross and Lepper 1979; Kuklinski, Quirk, Jerit, Schweider and Rich 2000).

In contrast to Zaller’s memory-based approach, Lodge and Taber’s **John Q. Public model** (2006; 2013) explains political evaluations as the result of motivated reasoning driven largely by affective, unconscious processes. People make snap, emotion-laden judgments of political stimuli on the basis of affective tallies stored in memory. These tallies are in turn determined by primes and other subliminal cues accompanying the issues, candidates, or groups under consideration. Conscious processing of political information is thus little more than rationalization of the associated attitudes that elude our awareness. It is clear that from this perspective backlash effects should be possible: Evidence that challenges one’s political views could trigger “hot cognitions” about related topics, which in turn would motivate a search (in memory or elsewhere) for confirmatory information (Redlawsk 2002).

More recently, Kahan (2012) has applied the **theory of cultural cognition** to public perceptions of risk on issues such as climate change. This perspective suggests, for example, that endorsing factual positions at odds with scientific consensus can be “expressively rational” in the sense that it reinforces one’s membership in a cultural or ideological group. Such identity-protective cognition can be both conscious and unconscious, and it could lead to predictions of backlash via mechanisms similar to John Q. Public.

A final perspective derives from the **Bayesian ideal of opinion change**. This model provides a simple, mathematically coherent mechanism, via Bayes’ rule, for updating one’s prior beliefs in light of new evidence. The predictions of the model are subtle, leading to occasional disagreements about the expected pattern of evidence under various conditions. For example, whether “unbiased” Bayesian learning implies convergence or parallel updating in evaluations of political figures has been the subject of continuing debate (Gerber and Green 1999; Bartels 2002; Bullock 2009). Bayesian rationality has often been taken to rule out polarization, but

even this is possible in the presence of idiosyncratic likelihood functions, which determine the subjective probability of observing a piece of evidence given a particular state of the world. In other words, the Bayesian model is compatible with a wide range of empirical patterns, even including backlash (Benoît and Dubra 2014).

Thus far, we have outlined four theoretical perspectives that predict, or at least allow for, the possibility of backlash effects. How often do such effects occur? Focusing on randomized experiments, we searched the literature for evidence of backlash effects in response to informational treatments. Within the context of correcting factual misperceptions, there are several such studies. Nyhan and Reifler (2010) discovered evidence of “backfire” effects to corrections of misinformation embedded in mock news articles about WMD in Iraq and funding for stem cell research (see Wood and Porter 2016 for a replication and extension). In each case, the ideological group most threatened by the corrective information (conservatives in the former case, liberals in the latter) moved in the direction of greater agreement with the misinformed view as a result of the correction. In another study, Nyhan et al. (2014) showed that providing a correction about vaccine misperceptions can decrease vaccine-averse subjects’ reported intention to vaccinate; this finding was replicated in Nyhan and Reifler (2015). Finally, Zhou (2016) identifies so-called “boomerang” effects in framing experiments on Republicans’ responses to climate change messaging.

Alongside these findings are studies that either do not find convincing evidence of backlash or highlight alternative explanations. Redlawsk, Civettini and Emmerson (2010) examine a hypothesized “affective tipping point” or specific dose of counter-attitudinal information at which backlash stops and incorporation of the evidence begins. While the authors do not emphasize this point, the effects of small doses are too small to be distinguishable from zero.¹ The effects of large doses are positive and significant. A more straightforward case is Americans’ response to advances in gay rights: Bishin, Hayes, Incantalupo and Smith (2015) conclude from both experimental and over-time survey data that there is “no evidence of backlash by the public as a whole or among relevant constituent groups.” A recent study of political rumors (Berinsky 2015) found that backlash can be prevented through the use of partisan source credibility cues.

¹See footnote on page 579.

Finally, an emerging literature on rational learning argues that apparent factual misperceptions are at least partially an artifact of expressive responding by partisans (Prior, Sood, Khanna et al. 2015; Bullock, Gerber, Hill, Huber et al. 2015).

Given the mixed findings in the literature, it is important to understand the prevalence of backlash effects. How common are they? Just as important, what are the conditions under which backlash is most likely to be observed? The theoretical accounts outlined above hypothesize a few factors in particular. For example, cultural cognition would predict boomerang effects in response to factual claims that threaten one’s cultural or ideological worldview. The John Q. Public and related motivated reasoning accounts suggest a greater likelihood of backlash among the most politically knowledgeable and aware individuals as well as the strongest partisans.

In this paper, we present results from three well-powered randomized experiments, each designed to identify the effect of exposure to information on the attitudes and beliefs of different subgroups. We chose three distinct issues intended to cover a range of possible backlash triggers. As we detail in the next section, we operationalize the concept of “backlash” as the appearance of negative treatment effects for some subgroups—in other words, attitude change in the direction contrary to that suggested by the information presented.

Across all three studies, we find no evidence of backlash among theoretically relevant subgroups. This is most remarkable in our third study, on gun control, which was conducted on a nationally representative sample and fielded in the aftermath of the largest mass shooting in American history. Overall, we find that subjects respond in a manner consistent with rational learning. At the same time, a number of the theoretical accounts highlighted here can accommodate backlash as well as its absence. Our results suggest that while backlash may occur under some conditions with some individuals, it is the exception, not the rule.

The paper proceeds as follows. In the next section, we formalize the concept of “backlash” and specify how to measure it in data generated from a randomized experiment. Next, we outline our overall research approach for three separate studies. We then present the experimental results from each study in turn. We conclude with a discussion of the implications of our findings: The effects of information do not appear to vary dramatically from person

to person, suggesting that marginal updating in the direction of evidence is possible for large swaths of the population.

3 Measuring Backlash

Suppose that each individual i is endowed with three potential outcomes $Y_i(neg)$, $Y_i(control)$, and $Y_i(pos)$, corresponding to the attitude he or she would express if exposed to negative information, no information, or positive information. We define two individual-level treatment effects $\tau_{i,neg}$ and $\tau_{i,pos}$. $\tau_{i,neg}$ is defined as the difference between the negative and control potential outcomes: $Y_i(neg) - Y_i(control)$. $\tau_{i,pos}$ is defined analogously: $Y_i(pos) - Y_i(control)$. Individual i updates his or her view in the direction of evidence if $\tau_{i,neg} \leq 0$ and $\tau_{i,pos} \geq 0$. Individual i “backlashes” if $\tau_{i,neg} > 0$ or $\tau_{i,pos} < 0$.

Our expectation is that for *most* individuals and *most* treatments, $\tau_{i,neg}$ will be negative and $\tau_{i,pos}$ will be positive. Our main concern is whether or not there are *any* individuals for whom these signs are reversed. Unfortunately, due to the Fundamental Problem of Causal Inference (Holland 1986), we can never observe $\tau_{i,neg}$ or $\tau_{i,pos}$ for any individual. We can, however, estimate average causal effects. The Average Negative Treatment Effect (ANTE) is defined as $E[\tau_{i,neg}]$, where $E[\cdot]$ denotes the expectation operator. The Average Positive Treatment Effect (APTE) is defined analogously.

In the empirical sections below, we will present three randomized experiments in which we obtain estimates of the ANTE and the APTE. What can we conclude from these estimates? If the ANTE is estimated to be negative and the APTE is estimated to be positive, we cannot draw strong conclusions about whether or not $\tau_{i,neg}$ and $\tau_{i,pos}$ were “correctly” signed for all individuals; that is, we cannot conclude that there is no backlash simply because on average, individual effects have the expected sign. If, however, the ANTE or the APTE are estimated have the “wrong” sign, we can indeed conclude that at least some number of subjects experienced backlash.

We will extend this logic to subgroups of subjects. The CANTE and the CAPTE are the conditional cousins of the ANTE and the APTE, i.e., they refer to the average causal

effects conditional on membership in a subgroup. In particular, the majority of the theories of backlash enumerated above predict that backlash is most likely among individuals whose baseline opinions are opposed to the evidence that they see. To be specific, $Y_i(\textit{baseline})$ is a pre-treatment characteristic of individuals. $Y_i(\textit{baseline})$ is likely to be correlated with (but distinct from) $Y_i(\textit{control})$, the post-treatment outcome that subjects express when assigned to the control condition. We define “proponents” as those for whom $Y_i(\textit{baseline})$ is high and “opponents” as those for whom it is low.² Backlash theories predict that $\tau_{i,neg}$ is likely to be *positive* among proponents and that $\tau_{i,pos}$ is likely to be *negative* among opponents. If so, we are more likely to find CANTE estimates to be positive among proponents and CAPTE estimates to be negative among opponents.

Even if we fail to find “incorrectly” signed average causal effects among these subgroups, we will not be able to rule out incorrectly signed individual causal effects. We are left with something of an inferential dilemma: We are looking for evidence of backlash, but the failure to do so does not rule backlash out completely. Our empirical strategy is asymmetric: We can demonstrate that backlash occurs if we can estimate incorrectly signed average causal effects with sufficient precision, but we cannot conclusively demonstrate that it does not occur.

Another approach is to consider the variances of $Y_i(\textit{neg})$, $Y_i(\textit{control})$, and $Y_i(\textit{pos})$. If it is indeed true that $\tau_{i,neg}$ is negative for most, but positive for some, the variance of $Y_i(\textit{neg})$ will be higher than the variance of $Y_i(\textit{control})$. If effects are homogeneous across subjects, then the variance of the two sets of potential outcomes will be equal. We view an inspection of the variance of outcomes as partially informative about backlash: while backlash would be variance-increasing, so too could other patterns of treatment effects.

4 Research Approach

Our three studies share important design features, so we describe them together here for convenience. All three studies employ a within-and-between-subjects experimental design. First, respondents were invited to complete a pre-treatment (T1) survey in which we collect baseline

²What “high” and “low” mean in any specific context is a matter of judgment, and in one empirical application, we also estimate conditional effects among “moderates,” those whose values of $Y_i(\textit{baseline})$ are middling.

demographic information, importantly including measures of $Y_i(\textit{baseline})$. Second, respondents were invited back to a main survey (T2) in which treatments were allocated and post-treatment outcomes were collected.

We conducted these studies on two platforms, Amazon’s Mechanical Turk (MTurk) and a nationally representative sample administered by GfK. Obtaining nationally representative samples is expensive, and we gratefully acknowledge the assistance of the Time-sharing Experiments in the Social Sciences (TESS) organization in supporting this research. In recent years, social scientists have recognized the utility of MTurk as a tool for recruiting subjects (e.g., Buhrmester, Kwang and Gosling 2011). While opt-in samples collected via MTurk should not be considered representative of the U.S. population, they have been shown to outperform laboratory-based convenience samples in ensuring adequate variation across sociodemographic and political characteristics of interest (Berinsky, Huber and Lenz 2012). Typically, MTurk samples tend to skew younger, less wealthy, better educated, more male, whiter, and more liberal than the population as whole. We opted for this approach in Study 1 and Study 2 in order to boost the representativeness relative to student populations (Sears 1986; Clifford and Jerit 2014), and to better ensure generalizability of results via similarity of subjects, treatments, contexts, and outcome measures across domains (Coppock and Green 2015). The utility of MTurk samples for drawing inferences about the causal effects of information treatments depends on treatment effect heterogeneity. If the treatment effects for these subjects are substantially different from the effects for others, then MTurk is a poor guide to effects more generally. The evidence to date on this question suggests that estimates obtained from MTurk samples match national samples well (Mullinix, Leeper, Druckman and Freese 2015; Coppock 2016).

In all three studies, we estimate treatment effects of information separately for “proponents” and “opponents” as defined by pretreatment measures of our dependent variables. In Appendix A, we reproduce our analyses splitting our sample by ideology (liberals, moderates, conservatives) or partisanship (Democrats, independents, Republicans). Likely because these political categories are strongly correlated with proponent/opponent status, the results presented in the appendix are substantively quite similar to those presented in the main text.

5 Study 1: Capital Punishment

The treatments used in Study 1 are directly inspired by (Lord, Ross and Lepper 1979).³ In that study, subjects were presented sequentially with apparent scientific evidence both challenging and affirming the notion that the death penalty deters crime, what we would refer to as a “mixed evidence” condition. To this single condition, various combinations of pro-capital punishment, anti-capital punishment, and inconclusive evidence were added. We made minor updates to the text (changing the publication date of the fictitious research articles from 1977 to 2012, for example) and to the graphical and tabular display of the fabricated data using modern statistical software.

We recruited 1,659 MTurk subjects to take the T1 survey in which we gathered a series of pre-treatment covariates (age, race, gender, and political ideology) and two items concerning capital punishment: attitude toward the death penalty and belief in its deterrent effect. From the pool of 1,659, 933 subjects’ pre-survey responses indicated clear and consistent support or opposition to capital punishment. These subjects were invited to participate in the main survey. Among these, proponents were defined as subjects whose answers to the pre-treatment attitude and belief questions were between 5 and 7 on a 7-point scale. Opponents were defined as subjects whose answers to these questions were between 1 and 3. In total, 683 subjects participated in the main survey (287 proponents and 396 opponents).

The two main dependent variables measured subjects’ attitudes and beliefs about capital punishment. The *Attitude* question asked, “Which view of capital punishment best summarizes your own?” The response options ranged from 1 (I am very much against capital punishment) to 7 (I am very much in favor of capital punishment). The *Belief* question asked, “Does capital punishment reduce crime? Please select the view that best summarizes your own.” Responses ranged from 1 (I am very certain that capital punishment does not reduce crime) to 7 (I am very certain that capital punishment reduces crime).

³Authors who performed an earlier replication of the original design (Kuhn and Lao 1996) generously shared the original experimental materials, so we were able to use the identical wordings of the study summaries and descriptions.

5.1 Study 1: Procedure

Subjects were randomly assigned to one of six conditions,⁴ as shown in Table 1 below. All subjects were presented with two research reports on the relationship between crime rates and capital punishment that varied in their findings: *Pro* reports presented findings that capital punishment appears to decrease crime rates, *Con* reports showed that it appears to increase crime rates, and *Null* reports showed that no conclusive pattern could be discerned from the data. The reports used one of two methodologies:⁵ either cross-sectional (a comparison of 10 pairs of neighboring states, with and without capital punishment) or time-series (a comparison of the crime rates before and after the adoption of capital punishment in 14 states).

Subjects were exposed to both of their randomly assigned research reports—one time series and one cross-sectional within each treatment condition—according to the following procedure:

1. Subjects were first presented with a “Study Summary” page in which the report’s findings and methodology were briefly presented. Subjects then answered two questions about how their attitudes toward the death penalty and beliefs about its deterrent efficacy had changed as a result of reading the summary.
2. Subjects were then shown a series of three pages that provided further details on the methodology, results, and criticisms of the report. The research findings were presented in both tabular and graphical form.
3. After reading the report details and criticism, subjects answered a series of five questions (including a short essay) that probed their evaluations of the study’s quality and persuasiveness.
4. Subjects then answered the attitude and belief change questions a second time.

Subjects completed steps one through four for both the first and second research reports. After reading and responding to the first and second reports, subjects were asked two endline *Attitude* and *Belief* questions, identical to the pre-treatment questions.

The statistical models will operationalize the “information content” of the pair of reports seen by subjects in a linear fashion. The positive information content of two *Pro* reports is coded as 2, one *Pro* and one *Null* as 1, and so on. In order to allow the coefficient on information

⁴Following the original Lord, Ross and Lepper (1979) procedure, in treatment conditions 1, 3, and 6, the order of the reports’ methodology (time series or cross-sectional) was randomized, resulting in two orderings per condition. In treatment conditions 2, 4, and 5, both the order of the methodology and the order of the content were randomized, resulting in four orderings per condition. In total, subjects could be randomized into 18 possible presentations. This design was maintained in order to preserve comparability with the original study, but we average over the order and methodology margins to focus on the effects of information.

⁵The experimental materials are available in Appendix B.

content to vary depending on whether the information is pro- or counter-attitudinal, we will split the information content variable into positive information and negative information, as shown in Table 1.

Table 1 also presents the mean and standard deviation of both outcome variables by treatment group. These means provide a first indication that the treatments had average effects in the “correct” direction.

We can also use the standard deviations reported in the table to perform a test of whether the treatments polarized opinion. If the treatments were polarizing, the standard deviation in the successively more pro or more con treatment groups should be larger, but we do not observe this pattern. Formal statistical tests also reveal that the treatment groups do not differ with respect to the standard deviations of the outcomes. Relative to the *Null Null* condition, the differences-in-standard-deviations across groups are generally not statistically significant at the 5% level, according to a randomization inference test conducted under the sharp null hypothesis of no effect for any unit. The only exception is the difference-in-standard-deviations between the *Null Null* and *Pro Null* conditions for the *Support* dependent variable ($p = 0.04$). This inference does not survive common multiple comparisons corrections, including the Bonferonni, Holm, and Benjamini-Hochberg corrections. We conclude from these tests that the treatments do not polarize opinion in the sense of increasing its variance.

5.2 Study 1: Analytic Strategy

The relatively complicated design described above can support many alternative analytic strategies. Each report is associated with seven intermediate dependent variables in addition to the two endline dependent variables. Subjects could have been assigned to eighteen different combinations of research reports. Reducing this complexity requires averaging over some conditions and making choices about which dependent variables to present. We present here our preferred analysis. We will focus on the separate effects of positive and negative information on subjects’ T2 responses to the *Attitude* and *Belief* questions. Because the *Null Null* and *Pro Con* conditions are both scored 0 on both the positive information and negative information scales, we

Table 1: Study 1 (Capital Punishment) : Treatment Conditions

Condition	N	Positive Informa- tion	Negative Informa- tion	T2 Attitude		T2 Belief	
				Mean	SD	Mean	SD
Con Con	117	0	2	3.52 (0.19)	2.07 (0.07)	3.27 (0.15)	1.65 (0.09)
Con Null	116	0	1	3.15 (0.20)	2.12 (0.08)	3.11 (0.15)	1.58 (0.08)
Null Null	112	0	0	3.18 (0.20)	2.06 (0.09)	3.13 (0.15)	1.56 (0.08)
Pro Con	118	0	0	3.59 (0.21)	2.27 (0.07)	3.69 (0.16)	1.75 (0.09)
Pro Null	121	1	0	3.62 (0.21)	2.28 (0.07)	3.81 (0.16)	1.66 (0.09)
Pro Pro	102	2	0	3.69 (0.22)	2.22 (0.07)	4.13 (0.16)	1.61 (0.10)

Bootstrapped standard errors in parentheses.

include an intercept shift for the *Null Null* condition, as shown in Equation 1:

$$Y_i = \beta_0 + \beta_1(POS_i) + \beta_2(NEG_i) + \beta_3(Condition_i = \text{Null Null}) + \epsilon_i \quad (1)$$

We will estimate Equation 1 using OLS. $\hat{\beta}_1$ forms our estimate of the APTE and $\hat{\beta}_2$ our estimate of the ANTE. We will provide estimates for proponents and opponents separately, allowing for an easy comparison of $\hat{\beta}_1$ and $\hat{\beta}_2$ across subgroups.

5.3 Study 1: Results

Tables 2 and 3 present estimates of the effects of information on attitudes and beliefs about capital punishment. Focusing on the covariate-adjusted models, we estimate that a unit increase in positive information causes an average increase in support for capital punishment of 0.015 scale points among proponents and 0.068 scale points among opponents, neither of which is statistically significant. Negative information has a strong negative effect among proponents (-0.326, $p < 0.01$), and a weakly negative effect among opponents (-0.042, $p = 0.43$). These

estimates imply that moving from the *Con Con* condition to the *Pro Pro* condition would cause proponents to move $2 \cdot 0.015 + 2 \cdot 0.326 = 0.682$ scale points and opponents to move $2 \cdot 0.068 + 2 \cdot 0.042 = 0.220$ scale points. While the treatment effects do appear to differ by subject type ($p < 0.05$), we do not observe the “incorrectly” signed treatment effects that backlash would produce.

Table 2: Effects of Information on Support for Capital Punishment

	Dependent Variable: T2 Attitude Toward Capital Punishment			
	Among Proponents		Among Opponents	
Positive Information (0 to 2)	0.05 (0.10)	0.01 (0.09)	0.11 (0.09)	0.07 (0.06)
Negative Information (0 to 2)	-0.27*** (0.10)	-0.33*** (0.08)	0.01 (0.08)	-0.04 (0.05)
Condition: Null Null	-0.23 (0.20)	-0.19 (0.16)	-0.004 (0.14)	-0.07 (0.10)
Constant	5.86 (0.13)	0.44 (0.56)	1.76 (0.10)	0.66 (0.23)
Covariates	No	Yes	No	Yes
N	287	287	396	396
R ²	0.04	0.41	0.01	0.47

* $p < .1$; ** $p < .05$; *** $p < .01$

Robust standard errors are in parentheses.

The information content of the Null Null condition is coded 0.

Covariates include T1 Attitude, T1 Belief, age, gender, ideology, race, and education.

Turning to Table 3, we observe that the effects of the information treatments on belief in the deterrent efficacy of capital punishment are nearly identical for proponents and opponents. For both groups, moving from *Con Con* to *Pro Pro* results in an entire scale point’s worth of movement. Study 1 does not provide any direct evidence of backlash. For both proponents and opponents, treatment effects were always “correctly” signed.

6 Study 2: Minimum Wage

One possible objection to the results from Study 1 is that the conditions for backlash were not met. Specifically, the issue of capital punishment cuts across party lines and socioeconomic groups, so perhaps subjects are less inclined to “argue back” in the face of counter-attitudinal

Table 3: Effects of Information on Belief in Deterrent Efficacy

	Dependent Variable: T2 Belief in Deterrent Effect			
	Among Proponents		Among Opponents	
Positive Information (0 to 2)	0.15 (0.11)	0.12 (0.10)	0.35*** (0.10)	0.31*** (0.09)
Negative Information (0 to 2)	-0.35*** (0.11)	-0.36*** (0.10)	-0.16 (0.10)	-0.21** (0.09)
Condition: Null Null	-0.32 (0.21)	-0.28 (0.19)	-0.27* (0.16)	-0.30** (0.14)
Constant	5.08 (0.14)	1.69 (0.68)	2.47 (0.11)	1.44 (0.33)
Covariates	No	Yes	No	Yes
N	287	287	396	396
R ²	0.08	0.25	0.09	0.34

*p < .1; **p < .05; ***p < .01

Robust standard errors are in parentheses.

The information content of the Null Null condition is coded 0.

Covariates include T1 Attitude, T1 Belief, age, gender, ideology, race, and education.

evidence. Study 2 addresses this concern by focusing on the minimum wage, a highly contentious issue area that divides the political parties and other demographic groups. The design of Study 2 is similar in spirit to Study 1, but instead of presenting treatments as figures with accompanying text, the treatments in this study were short web videos in favor of and against raising the minimum wage.

6.1 Study 2: Procedure

A large number ($N = 2,979$) of survey respondents on MTurk were recruited to participate in a pre-treatment survey measuring demographic characteristics (age, gender, race/ethnicity, education, partisan affiliation, and ideological leaning) as well as baseline attitudes toward the minimum wage. From this large pool of survey respondents, 1,500 were invited to take part in the main survey testing the effect of videos on attitudes toward the minimum wage. Invitations to take part in the main survey were offered on a random basis, though more slots were offered to younger respondents and those with stronger views (pro or con) about the minimum wage. Of the 1,500 recruited to the main survey, 1,170 participated.

Subjects were exposed to two videos on the subject of the minimum wage. Two of the videos were in favor of minimum wage increases, one presented by John Green, a popular video blogger, and the other presented by Robert Reich, former U.S. Secretary of Labor and established left-leaning public intellectual. On the “con” side of the debate, one video was presented by an actor, and the other by economics professor Antony Davies. Within each side, one video featured a relatively young presenter and the other a relatively old presenter. Finally, two videos were included as placebos, addressing mundane requirements of state minimum wage laws. Links to all six videos are available in Appendix B, as well as screenshots that convey the production quality and mood of the videos.

Subjects were randomized into one of thirteen conditions: placebo, or one of the twelve possible orderings of the four persuasive videos. Subjects answered intermediate questions relating to how well-made and persuasive they found each video, then at the end of the survey, they answered two questions which serve as our main dependent variables. The *Favor* question asked, “The federal minimum wage is currently \$7.25 per hour. Do you favor or oppose raising the federal minimum wage?” The response options ranged from 1 (Very much opposed to raising the federal minimum wage) to 7 (Very much in favor of raising the federal minimum wage). The *Amount* question asked, “What do you think the federal minimum wage should be? Please enter an amount between \$0.00 and \$25.00 in the text box below.”

6.2 Study 2: Analytic Strategy

As in Study 1, we order the treatment conditions according to the amount of pro-minimum-wage video content. The information content of the *Con Con* conditions is scored -1, the *Pro Con* and *Placebo* conditions 0, and the *Pro Pro* conditions 1, as shown in Table 4. We will model responses as in Equation 1 and estimate separate regressions for opponents, moderates, and proponents. Opponents are defined as subjects whose pre-treatment *Favor* response was 4 or lower and whose *Amount* response was below the median response (\$10.00). Those with *Favor* responses of 4 or higher and *Amount* responses above the median are defined as proponents. All others are defined as moderates.

Table 4 presents the means and standard deviations by experimental group. Here again, we see an indication that the treatments had average effects in their intended directions. The means of the *Con Young* / *Con Old* condition are lower than the means of the mixed conditions, which are themselves lower than the means of the *Pro Young* / *Pro Old* conditions. These differences are all statistically significant. Turning to the differences-in-standard-deviations, formal tests under the sharp null of no effect lend some support to the hypothesis that the treatments lead to increases in the polarization of opinion—the differences between the placebo condition and the *Pro Old*, *Con Old* condition are statistically significant for both dependent variables at the $p < 0.01$ level. However, while increases in the standard deviations of outcomes would be a consequence of backlash, these increases could also result from some individuals having larger treatment effects than others, but with all effects still correctly signed.

Table 4: Study 2 (Minimum Wage): Treatment Conditions

Condition	N	Positive Informa- tion	Negative Informa- tion	Favor		Amount	
				Mean	SD	Mean	SD
Placebo	93	0	0	5.34 (0.18)	1.64 (0.12)	9.23 (0.30)	2.77 (0.35)
Con Young / Con Old	162	0	1	4.77 (0.15)	1.84 (0.08)	8.67 (0.21)	2.65 (0.25)
Pro Old / Con Old	165	0	0	4.99 (0.17)	2.09 (0.10)	9.99 (0.34)	4.28 (0.29)
Pro Old / Con Young	195	0	0	4.87 (0.14)	1.93 (0.07)	9.85 (0.23)	3.32 (0.24)
Pro Young / Con Old	169	0	0	5.01 (0.15)	1.90 (0.09)	9.30 (0.23)	3.03 (0.27)
Pro Young / Con Young	192	0	0	5.18 (0.12)	1.64 (0.09)	9.38 (0.21)	2.94 (0.32)
Pro Young / Pro Old	194	1	0	5.59 (0.12)	1.67 (0.11)	10.93 (0.29)	3.90 (0.29)

6.3 Study 2: Results

The results of Study 2 are presented in Tables 5 and 6. Again focusing on the covariate-adjusted estimates, the video treatments had powerful effects for all three subject types. Positive information had positive and statistically significant effects on subjects' preferred minimum wage amount; negative information had strongly negative effects.

A similar pattern of response is evident in Table 6. All coefficients are correctly signed. With the exception of negative information among opponents, all these coefficients are statistically significant.

Table 5: Effects of Information on Preferred Minimum Wage Amount

	Dependent Variable: T2 Amount					
	Among Opponents		Among Moderates		Among Proponents	
Pos. Info (0 to 1)	0.43 (0.51)	0.60* (0.33)	1.88*** (0.37)	1.51*** (0.30)	1.27*** (0.36)	1.37*** (0.29)
Neg. Info (0 to 1)	-0.61 (0.50)	-0.70** (0.31)	-0.83*** (0.18)	-0.93*** (0.19)	-1.93*** (0.31)	-1.70*** (0.30)
Condition: Placebo	-0.12 (0.53)	-0.60** (0.25)	-0.35 (0.27)	-0.46* (0.25)	-0.79* (0.42)	-0.49** (0.24)
Constant	6.89 (0.21)	2.49 (0.86)	9.37 (0.10)	5.64 (2.66)	11.99 (0.18)	1.31 (1.19)
Covariates	No	Yes	No	Yes	No	Yes
N	343	343	356	356	471	471
R ²	0.01	0.66	0.19	0.33	0.10	0.44

*p < .1; **p < .05; ***p < .01

Robust standard errors are in parentheses.

The information content of the Placebo condition is coded 0.

Covariates include T1 Amount, T1 Favor, age, gender, ideology, party ID, and education.

7 Study 3: Gun Control

It is useful to replicate a similar design on a large, nationally representative sample. For this issue in particular, the incidence of *complete* opposition to gun control in the United States is relatively low. Mechanical Turk samples may be inappropriate in this case because they tend to skew liberal, making the rarity of gun control opponents more of a problem in those

Table 6: Effects of Information on Favoring Minimum Wage Raise

	Dependent Variable: T2 Favor					
	Among Opponents		Among Moderates		Among Proponents	
Pos. Info (0 to 1)	0.86*** (0.25)	0.77*** (0.21)	0.56*** (0.16)	0.55*** (0.14)	0.29*** (0.09)	0.35*** (0.07)
Neg. Info (0 to 1)	-0.02 (0.25)	-0.13 (0.19)	-0.63*** (0.21)	-0.66*** (0.22)	-0.56*** (0.18)	-0.48*** (0.16)
Condition: Placebo	0.70** (0.27)	0.34 (0.25)	0.28 (0.24)	0.20 (0.21)	0.01 (0.19)	0.07 (0.17)
Constant	2.97 (0.11)	1.72 (0.53)	5.34 (0.08)	3.11 (1.22)	6.34 (0.05)	2.66 (0.53)
Covariates	No	Yes	No	Yes	No	Yes
N	343	343	356	356	471	471
R ²	0.05	0.47	0.07	0.18	0.06	0.33

*p < .1; **p < .05; ***p < .01

Robust standard errors are in parentheses.

The information content of the Placebo condition is coded 0.

Covariates include T1 Amount, T1 Favor, age, gender, ideology, party ID, and education.

populations.

More to the point, as noted in Section 3, the hypotheses we test are predictions about heterogeneous effects (here, the effect of information treatments on those who initially support or oppose gun control). Such subgroup analyses are especially sensitive to representativeness in a sample, and inferences require additional assumptions about the nature of treatment interactions (Gelman 2013; Mullinix et al. 2015).

Third, even if convenience samples contain a sufficient number of subjects from each subgroup of interest, it is possible that there are systematic differences between the members of a subgroup in Mechanical Turk samples and the corresponding members of that subgroup in other, more representative samples. In particular, some scholars have worried about selection bias in the case of conservatives on Mechanical Turk (Kahan 2013).

Finally, there are persistent and valid concerns about the tendency of MTurk workers to seek out the “right” answer or to exhibit other types of demand effects, especially given that members of that population tend to participate in many social science studies (Chandler, Mueller and Paolacci 2014). Given the types of information treatments that we propose to use—ones in

which the correct answers could be found with an Internet search—we want to ensure that our study is fielded to respondents who are less likely to succumb to these pressures.

7.1 Study 3: Procedure

We fielded Study 3 on a nationally representative sample ($N = 2,122$) administered by GfK from June 22-28, 2016, roughly 10 days after the mass shooting in Orlando, Florida and in the midst of a heated debate about terrorism and the regulation of firearms. In a preliminary wave of the survey, administered 3-10 days after the shooting, we determined whether subjects support (“proponent”) or oppose (“opponent”) stricter gun control laws. We also asked subjects four questions about their preferred gun control policies. We combine all four dependent variables into a composite index using factor analysis in order to improve power (Ansolabehere, Rodden and Snyder 2008).

In the experiment, subjects were randomly assigned to one of three conditions: no information (control), pro-gun-control information (positive), or anti-gun-control information (negative). The treatments we employed were modeled on those of Lord, Ross and Lepper (1979). Subjects were shown graphical evidence of a relationship between gun control policies and four outcome variables: gun homicides, gun suicides, gun accidental deaths, and gun assaults. The evidence was presented as if it were the central finding in “Kramer and Perry (2014)”, a fictitious academic article. (See Appendix B for questionnaire and stimuli.) We then collected our dependent variables and asked again about subjects’ “proponent” or “opponent” status.

Table 7: Study 3 Gun Control : Treatment Conditions

Condition	N	T2 Attitude		T2 Belief	
		Mean	SD	Mean	SD
Control	730	-0.03 (0.04)	0.04 (0.00)	0.69 (0.02)	0.02 (0.00)
Positive Information	702	0.06 (0.04)	0.04 (0.00)	0.72 (0.02)	0.02 (0.00)
Negative Information	690	0.00 (0.04)	0.04 (0.00)	0.62 (0.02)	0.02 (0.00)

Bootstrapped standard errors in parentheses.

7.2 Study 3: Analytic Strategy

We present results that we pre-registered in planned regression specifications. We use OLS with HC2 robust standard errors, separately by subject type. We employ survey weights (provided by GfK) for all models.

7.3 Study 3: Results

Table 8: Effects of Information on Gun Control Composite Scale

	Dependent Variable: Composite Scale			
	Among Opponents		Among Proponents	
Positive Information	0.05 (0.09)	0.05 (0.08)	0.06 (0.06)	0.04 (0.06)
Negative Information	0.04 (0.08)	−0.003 (0.08)	0.04 (0.06)	0.04 (0.06)
Constant	−0.91 (0.05)	−1.69 (0.28)	0.41 (0.05)	0.08 (0.24)
Covariates	No	Yes	No	Yes
N	718	718	1,359	1,359
R ²	0.001	0.14	0.001	0.14

*p < .1; **p < .05; ***p < .01

Robust standard errors are in parentheses.

Covariates include age, registration, education, hispanic ethnicity, gender, income, marital status, employment status, party ID, and ideology.

As shown in Table 8, the effects of the information treatments are small and, in most cases, with standard errors as large or larger than the estimated coefficients. Whether as a result of the issue itself, the nature of the sample, or the timing of the experiment, we do not find substantial effects of information on our composite measure of gun control policy preferences. Importantly, however, these negligible persuasive effects do not mask heterogeneity across subgroups: Pro-gun-control information has positive coefficients for both opponents and proponents of gun regulation, indicating a lack of evidence for backlash. Negative information similarly has positive coefficients for both subgroups in the unadjusted models (although for opponents with covariate adjustment, the coefficient is just below zero).

Turning to the proponent/opponent question, which asks whether respondents support

Table 9: Effects of Information on Gun Control Support

	Dependent Variable: Support Gun Control			
	Among Opponents		Among Proponents	
Positive Information	0.03 (0.04)	0.02 (0.04)	0.01 (0.02)	0.01 (0.02)
Negative Information	-0.02 (0.04)	-0.03 (0.04)	-0.08*** (0.02)	-0.07*** (0.02)
Constant	0.19 (0.03)	-0.03 (0.16)	0.93 (0.01)	0.77 (0.07)
Covariates	No	Yes	No	Yes
N	724	724	1,375	1,375
R ²	0.002	0.10	0.02	0.10

*p < .1; **p < .05; ***p < .01

Robust standard errors are in parentheses.

Covariates include age, registration, education, hispanic ethnicity, gender, income, marital status, employment status, party ID, and ideology.

“stricter gun control laws in the United States,” we similarly estimate small but positive coefficients for positive information across subgroups.⁶ Unlike with the composite index, however, we find negative effects of anti-gun-control information on both opponents and proponents. These estimates rise to conventional levels of significance for *proponents* of gun control, suggesting, contrary to predictions of backlash, that it is those in favor of gun control who are most receptive to evidence questioning its effectiveness. Together, these findings show a more robust persuasive effect on a generalized measure of support compared to specific, policy-related opinions.

Finally, in Appendix A, we replicate the finding of no backlash by examining the effect of the information treatments by party identification instead of proponent status. This is significant because of its commonly theorized role as a perceptual filter that could, in the RAS or JQP accounts, promote polarization or backlash.

Why didn’t subjects’ attitudes change as much as we expected them to, given the findings of significant and substantively meaningful effects in Study 1 and Study 2? The most obvious difference is the type of sample, which for this study was a representative online sample instead of Mechanical Turk. Some scholars have argued that subjects recruited in this way are eager

⁶We note that this particular analysis—testing the effect on this question—was not preregistered.

to provide expected responses to stimuli (Chandler, Mueller and Paolacci 2014). However, it is unclear why this would also not be the case in other online samples (which also recruit subjects in similar ways). Furthermore, an emerging body of evidence shows that treatment effects estimated from convenience samples generally replicate those found on nationally representative surveys (e.g., Coppock 2016).

Another possibility lies in the coincidental timing and nature of our study. Attitudes on gun control may already be more crystallized than those on other issues (e.g., Broockman and Kalla 2016). We then deployed this survey at the height of a national debate on the subject, meaning that our treatments competed for attentional resources within a saturated information environment. Viewed from one perspective, it is all the more surprising that we do not find evidence of the graphical treatments causing polarization in attitudes given the common expectation that people are highly susceptible to motivated reasoning in the context of an emotional and contentious political debate. On the other hand, however, it is also possible that views on gun control are sufficiently hardened that it is difficult (but not impossible) to move them one way or another.

8 Discussion

How common is backlash? Across three studies, we find no evidence of the phenomenon. Pro-capital-punishment evidence tends to make subjects more supportive of the death penalty and to strengthen their beliefs in its deterrent efficacy. Evidence that the death penalty increases crime does the opposite, while inconclusive evidence does not cause significant shifts in attitudes or beliefs. Arguments about the minimum wage likewise move respondents in the direction of evidence—toward supporting a higher or a lower dollar amount according to the slant of the evidence presented in the video treatments. Finally, evidence that gun control decreases gun violence makes people more supportive while evidence that it increases violence does the opposite.

The studies reported here were designed to encompass a variety of different types of information: scientific evidence described in tables and graphs in addition to more colloquial video

appeals. The issues covered vary along numerous dimensions: both “hot” (gun violence) and “cold” (minimum wage), easily mapped to the partisan divide or not, and of varying degrees of salience. The results do not depend on the particular issue or idiosyncratic features of the topics chosen. Additionally, in results not reported here, backlash effects do not materialize over time. In two studies in which we collected follow-up responses (Study 1 and Study 2), the initial findings persist at least 10 days after the initial experiment, although the magnitudes are somewhat attenuated.

We noted above that our research design is asymmetric. If we had found evidence of incorrectly signed effects (positive effects of negative information or vice versa), we could have concluded that backlash did indeed occur in our experiments. We did not find incorrectly signed effects—but we cannot conclude that backlash did not occur because there may have been some individuals (not exclusively defined by “opponent” or “proponent” status or partisan identity) that did in fact have an adverse reaction to the treatments. One intriguing possibility is that this very asymmetry contributes to the relatively widespread contention that presenting individuals with counter-attitudinal information is counterproductive: We can draw sharp inferences when backlash occurs but are left wondering when it does not.

These experiments show that when people are exposed to information, that on average, they update their views in the “correct” direction. However, one way in which these findings might not generalize to non-experimental contexts is if people selectively avoid counter-attitudinal information. Prior (2007) and Arceneaux and Johnson (2013) find that many individuals, if given the choice, simply consume entertainment rather than news information, thereby selecting out of both pro- and counter-attitudinal information in one stroke. On the other hand, (Bakshy, Messing and Adamic 2015) shows that while partisan Facebook users do appear to prefer pro-attitudinal news stories, they are exposed to and do consume a large amount of counter-attitudinal information. Future research should consider the conditions under which individuals could be induced to seek out larger or smaller doses of information with which they disagree.

A reasonable objection to these findings is that while subjects may not polarize when reading relatively sterile descriptions of academic studies, individuals may do so when actually arguing about some proposition with an opponent. Political disputes linger and do not easily resolve

when new information comes to light. We speculate that in contentious political environments, in which opposing sides routinely insult the other (or much worse), the introduction of evidence could induce a divergence in attitudes. Perhaps in such antagonistic contexts, individuals become distrustful of counter-attitudinal arguments. We leave the search for backlash effects in such contentious environments to future research.

References

- Ansolabehere, Stephen, Jonathan Rodden and James M. Snyder. 2008. "The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting." *American Political Science Review* 102(2):215–232.
- Arceneaux, Kevin and Martin Johnson. 2013. *Changing Minds Or Changing Channels?: Partisan News in an Age of Choice*. University of Chicago Press.
- Bakshy, Eytan, Solomon Messing and Lada A Adamic. 2015. "Exposure to ideologically diverse news and opinion on Facebook." *Science* 348(6239):1130–1132.
- Bartels, Larry. 2002. "Beyond the Running Tally: Partisan Bias in Political Perceptions." *Political Behavior* 24(2):117–150.
- Benoît, Jean-Pierre and Juan Dubra. 2014. "A Theory of Rational Attitude Polarization." *Available at SSRN 2529494*.
- Berinsky, Adam, Gregory Huber and Gabriel Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20(3).
- Berinsky, Adam J. 2015. "Rumors and health care reform: experiments in political misinformation." *British Journal of Political Science* pp. 1–22.
- Bishin, Benjamin G., Thomas J. Hayes, Matthew B. Incantalupo and Charles Anthony Smith. 2015. "Opinion Backlash and Public Attitudes: Are Political Advances in Gay Rights Counterproductive?" *American Journal of Political Science*.
- Broockman, David and Joshua Kalla. 2016. "Durably reducing transphobia: A field experiment on door-to-door canvassing." *Science* 352(6282):220–224.
- Buhrmester, Michael D., Tracy Kwang and Samuel D. Gosling. 2011. "Amazon's Mechanical Turk: A New Source of Inexpensive, yet High-Quality, Data?" *Perspectives on Psychological Science* 6(1):3–5.
- Bullock, John G. 2009. "Partisan Bias and the Bayesian Ideal in the Study of Public Opinion." *The Journal of Politics* 71(03):1109–1124.
- Bullock, John G, Alan S Gerber, Seth J Hill, Gregory A Huber et al. 2015. "Partisan Bias in Factual Beliefs about Politics." *Quarterly Journal of Political Science* 10(4):519–578.
- Chandler, Jesse, Pam Mueller and Gabriele Paolacci. 2014. "Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers." *Behavior Research Methods* 46(1):112–130.
URL: <http://dx.doi.org/10.3758/s13428-013-0365-7>
- Clifford, Scott and Jennifer Jerit. 2014. "Is There a Cost to Convenience? An Experimental Comparison of Data Quality in Laboratory and Online Studies." *Journal of Experimental Political Science*.
- Coppock, Alexander. 2016. Positive, Small, Homogeneous, and Durable: Political Persuasion in Response to Information PhD thesis Columbia University.

- Coppock, Alexander and Donald P. Green. 2015. "Assessing the Correspondence Between Experimental Results Obtained in the Lab and Field: A Review of Recent Social Science Research." *Political Science Research and Methods* 3(1):113–131.
- Gelman, Andrew. 2013. "Does it matter that a sample is unrepresentative? It depends on the size of the treatment interactions." <http://andrewgelman.com/2013/09/04/does-it-matter-that-a-sample-is-unrepresentative-it-depends-on-the-size-of-the-treatment/> Accessed: 2016-01-01.
- Gerber, Alan and Donald Green. 1999. "Misperceptions About Perceptual Bias." *Annual Review of Political Science* 2(1):189–210.
- Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396):945–960.
- Kahan, Dan. 2013. "Fooled twice, shame on who? Problems with Mechanical Turk study samples, part 2." <http://www.culturalcognition.net/blog/2013/7/10/fooled-twice-shame-on-who-problems-with-mechanical-turk-stud.html>. Accessed: 2016-01-01.
- Kahan, Dan M. 2012. "Ideology, motivated reasoning, and cognitive reflection: an experimental study." *Judgment and Decision Making* 8:407–24.
- Kuhn, Deanna and Joseph Lao. 1996. "Effects of Evidence on Attitudes: Is Polarization the Norm." *Psychological Science* 7(2):115–120.
- Kuklinski, James H., Paul J. Quirk, Jennifer Jerit, David Schweider and Robert F. Rich. 2000. "Misinformation and the Currency of Democratic Citizenship." *Journal of Politics* 62:790–816.
- Kunda, Ziva. 1990. "The Case for Motivated Reasoning." *Psychological Bulletin* 108:480–498.
- Lau, R.R. and D.P. Redlawsk. 2006. *How Voters Decide: Information Processing During Election Campaigns*. Cambridge University Press.
- Lazarsfeld, Paul F., Bernard Berelson and Hazel Gaudet. 1944. *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. New York: Columbia U. Press.
- Lodge, Milton and Charles S Taber. 2013. *The Rationalizing Voter*. Cambridge University Press.
- Lord, Charles S., L. Ross and M. Lepper. 1979. "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence." *Journal of Personality and Social Psychology* 37:2098–2109.
- Miller, Arthur G., John W. McHoskey, Cynthia M. Bane and Timothy G. Dowd. 1993. "The Attitude Polarization Phenomenon: Role of Response Measure, Attitude Extremity, and Behavioral Consequences of Reported Attitude Change." *Journal of Personality and Social Psychology* 64(4):561–574.

- Mullinix, Kevin J., Thomas J. Leeper, James N. Druckman and Jeremy Freese. 2015. "The Generalizability of Survey Experiments." *Journal of Experimental Political Science* 2:109–138.
- Nyhan, Brendan and Jason Reifler. 2010. "When Corrections Fail: The Persistence of Political Misperceptions." *Political Behavior* 32(2):303–330.
- Nyhan, Brendan and Jason Reifler. 2015. "Does Correcting Myths about the Flu Vaccine Work? An Experimental Evaluation of the Effects of Corrective Information." *Vaccine* 33(3):459–464.
URL: <http://linkinghub.elsevier.com/retrieve/pii/S0264410X14015424>
- Nyhan, Brendan, Jason Reifler, Sean Richey and Gary L. Freed. 2014. "Effective Messages in Vaccine Promotion: A Randomized Trial." *Pediatrics* .
- Prior, Markus. 2007. *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*. Cambridge University Press.
- Prior, Markus, Gaurav Sood, Kabir Khanna et al. 2015. "You cannot be serious: The impact of accuracy incentives on partisan bias in reports of economic perceptions." *Quarterly Journal of Political Science* 10(4):489–518.
- Redlawsk, David 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.
- Redlawsk, David P., Andrew J. W. Civettini and Karen M. Emmerson. 2010. "The Affective Tipping Point: Do Motivated Reasoners Ever "Get It"?" *Political Psychology* 31(4):563–593.
- Sears, David O. 1986. "College Sophomores in the Laboratory: Influences of a Narrow Data Base on Social Psychology's View of Human Nature." *Journal of Personality and Social Psychology* 51(3):515–530.
- Strickland, April A, Charles S Taber and Milton Lodge. 2011. "Motivated reasoning and public opinion." *Journal of health politics, policy and law* 36(6):935–944.
- Taber, Charles S., Damon Cann and Simona Kucsova. 2009. "The Motivated Processing of Political Arguments." *Political Behavior* 31(2):137–155.
- Taber, Charles S. and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50(3):755–769.
- Wood, Thomas and Ethan Porter. 2016. "The elusive backfire effect: mass attitudes' steadfast factual adherence." *Working Paper* .
- Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.
- Zhou, Jack. 2016. "Boomerangs versus Javelins: How Polarization Constrains Communication on Climate Change." *Environmental Politics* 25(5):788–811.
URL: <http://dx.doi.org/10.1080/09644016.2016.1166602>

Appendix A: Effects of Information by Ideology and PID

In the main text, we presented estimates by proponent or opponent status, defined by $Y_{baseline}$, measured pre-treatment. In this appendix, we present parallel tables in which we subset our samples by ideology or partisan identification. In Study 1, we did not collect information about subjects' partisan attachments, so we use subjects' self-reported identification as a liberal, moderate, or conservative to subdivide the sample. In Studies 2 and 3, we conduct our analyses separately for Democrats, independents, and Republicans. In the tables that follow, we show that in no case do we estimate positive effects of negative information or negative effects of positive information. That is, we do not observe backlash in any of the groups defined by ideology in Study 1 or partisanship in Studies 2 and 3.

Table 10: Study 1: Effects of Information on Support for Capital Punishment By Ideology

	Dependent Variable: T2 Attitude Toward Capital Punishment					
	Among Liberals		Among Moderates		Among Conservatives	
Positive Information (0 to 2)	0.14 (0.18)	0.01 (0.06)	-0.20 (0.29)	0.10 (0.12)	0.16 (0.27)	0.01 (0.11)
Negative Information (0 to 2)	-0.10 (0.16)	-0.07 (0.05)	0.06 (0.26)	-0.29** (0.12)	-0.13 (0.25)	-0.26** (0.12)
Condition: Null Null	-0.48* (0.29)	-0.15 (0.12)	-0.07 (0.53)	-0.08 (0.19)	0.66 (0.40)	-0.26 (0.21)
Constant	2.69 (0.21)	0.25 (0.23)	3.96 (0.37)	0.29 (0.46)	5.19 (0.29)	1.14 (0.50)
Covariates	No	Yes	No	Yes	No	Yes
N	374	374	163	163	121	121
R ²	0.02	0.85	0.01	0.86	0.03	0.80

*p < .1; **p < .05; ***p < .01

Robust standard errors are in parentheses.

The information content of the Null Null condition is coded 0.

Covariates include T1 Attitude, T1 Belief, age, gender, proponent, race, and education.

Table 11: Study 1: Effects of Information on Belief in Deterrent Efficacy By Ideology

	Dependent Variable: T2 Belief in Deterrent Effect					
	Among Liberals		Among Moderates		Among Conservatives	
Positive Information (0 to 2)	0.40*** (0.15)	0.30*** (0.09)	0.01 (0.20)	0.14 (0.15)	0.21 (0.18)	0.11 (0.14)
Negative Information (0 to 2)	-0.17 (0.14)	-0.18** (0.09)	-0.12 (0.17)	-0.35** (0.14)	-0.44** (0.22)	-0.50*** (0.15)
Condition: Null Null	-0.45* (0.23)	-0.25* (0.14)	-0.14 (0.39)	-0.18 (0.26)	-0.15 (0.36)	-0.73** (0.30)
Constant	2.95 (0.18)	1.40 (0.33)	3.86 (0.21)	1.45 (0.59)	4.86 (0.24)	1.74 (0.61)
Covariates	No	Yes	No	Yes	No	Yes
N	374	374	163	163	121	121
R ²	0.07	0.60	0.004	0.55	0.09	0.60

*p < .1; **p < .05; ***p < .01

Robust standard errors are in parentheses.

The information content of the Null Null condition is coded 0.

Covariates include T1 Attitude, T1 Belief, age, gender, proponent, race, and education.

Table 12: Study 2: Effects of Information on Preferred Minimum Wage Amount by Partianship

	Dependent Variable: T2 Amount					
	Among Democrats		Among Independents		Among Republicans	
Pos. Info (0 to 1)	1.63*** (0.41)	1.48*** (0.28)	1.58*** (0.43)	1.22*** (0.31)	0.07 (0.63)	0.80** (0.37)
Neg. Info (0 to 1)	-1.44*** (0.31)	-1.38*** (0.24)	-0.68* (0.37)	-1.04*** (0.30)	-0.79 (0.57)	-1.05*** (0.28)
Condition: Placebo	-1.11** (0.50)	-0.67** (0.26)	-0.14 (0.47)	-0.47** (0.21)	0.17 (0.41)	-0.29 (0.25)
Constant	10.86 (0.17)	3.19 (0.91)	9.45 (0.20)	2.77 (0.93)	8.10 (0.23)	1.41 (1.64)
Covariates	No	Yes	No	Yes	No	Yes
N	469	469	416	416	226	226
R ²	0.09	0.52	0.06	0.60	0.01	0.69

*p < .1; **p < .05; ***p < .01

Robust standard errors are in parentheses.

The information content of the Placebo condition is coded 0.

Covariates include T1 Amount, T1 Favor, age, gender, ideology, initial position, and education.

Table 13: Study 2: Effects of Information on Favoring Minimum Wage Raise by Partianship

	Dependent Variable: T2 Favor					
	Among Democrats		Among Independents		Among Republicans	
Pos. Info (0 to 1)	0.39*** (0.14)	0.41*** (0.11)	0.87*** (0.21)	0.75*** (0.14)	0.26 (0.36)	0.52* (0.28)
Neg. Info (0 to 1)	-0.45** (0.21)	-0.56*** (0.17)	-0.34 (0.26)	-0.54*** (0.19)	0.21 (0.38)	0.11 (0.27)
Condition: Placebo	-0.09 (0.26)	0.02 (0.17)	0.54* (0.28)	0.15 (0.18)	0.63 (0.47)	0.46 (0.38)
Constant	5.89 (0.07)	3.05 (0.46)	4.86 (0.12)	1.49 (0.45)	3.83 (0.15)	0.11 (1.88)
Covariates	No	Yes	No	Yes	No	Yes
N	469	469	416	416	226	226
R ²	0.03	0.44	0.05	0.67	0.01	0.52

*p < .1; **p < .05; ***p < .01

Robust standard errors are in parentheses.

The information content of the Placebo condition is coded 0.

Covariates include T1 Amount, T1 Favor, age, gender, ideology, initial position, and education.

Table 14: Study 3: Effects of Information on Gun Control Composite Scale by Partisanship

	Dependent Variable: Composite Scale					
	Among Democrats		Among Independents		Among Republicans	
Positive Information	0.23* (0.12)	0.12 (0.09)	-0.002 (0.10)	0.03 (0.08)	0.04 (0.09)	0.05 (0.08)
Negative Information	0.08 (0.10)	0.06 (0.08)	-0.003 (0.10)	0.07 (0.08)	0.02 (0.09)	0.04 (0.07)
Constant	-0.57 (0.07)	-0.52 (0.31)	-0.05 (0.07)	-0.78 (0.23)	0.42 (0.06)	-0.81 (0.22)
Covariates	No	Yes	No	Yes	No	Yes
N	569	569	848	848	660	660
R ²	0.01	0.43	0.0000	0.36	0.0004	0.34

*p < .1; **p < .05; ***p < .01

Robust standard errors are in parentheses.

Covariates include age, registration, education, hispanic ethnicity,

gender, income, marital status, employment status, initial position, and ideology.

Table 15: Study 3: Effects of Information on Gun Control Support by Partisanship

	Dependent Variable: Support Gun Control					
	Among Democrats		Among Independents		Among Republicans	
Positive Information	0.11*	0.05	−0.001	0.02	−0.02	−0.01
	(0.06)	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)
Negative Information	−0.06	−0.07*	−0.07*	−0.05	−0.06*	−0.04
	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
Constant	0.39	0.51	0.68	0.31	0.91	0.34
	(0.04)	(0.13)	(0.03)	(0.10)	(0.02)	(0.09)
Covariates	No	Yes	No	Yes	No	Yes
N	573	573	858	858	668	668
R ²	0.02	0.48	0.01	0.48	0.01	0.47

*p < .1; **p < .05; ***p < .01

Robust standard errors are in parentheses.

Covariates include age, registration, education, hispanic ethnicity, gender, income, marital status, employment status, initial position, and ideology.

Appendix B: Experimental Materials

Study 1

Questions

The *Attitude* question asked, “Which view of capital punishment best summarizes your own?” The response options ranged from 1. “I am very much against capital punishment.” to 7. “I am very much in favor of capital punishment.” The *Belief* question asked, “Does capital punishment reduce crime? Please select the view that best summarizes your own.” Responses ranged from 1. “I am very certain that capital punishment does not reduce crime.” to 7. “I am very certain that capital punishment reduces crime.”

Sample Treatment: Con (Cross Section)

Does Capital Punishment Prevent Crime?

One of the most controversial public issues in recent years has been the effectiveness of capital punishment (the death penalty) in preventing murders. Proponents of capital punishment have argued that the possibility of execution deters people who might otherwise commit murders, whereas opponents of capital punishment denied this and maintain that the death penalty may even produce murders by setting a violent model of behavior. A recent research effort attempted to shed light on this controversy.

The researchers (Palmer and Crandall, 2012) decided to look at the difference in murder rates in states that share a common border but differ in whether their laws permit capital punishment or not. Carefully limiting the states included to those which had capital punishment laws in effect or not in effect for at least five years, they compiled a list of all possible pairs and then selected ten pairs of neighboring states that were alike in the degree of urbanization (percentage of the population living in metropolitan areas), thus controlling for any relationship between the size of urban population and crime per capita. They also limited the capital punishment states to those which had actually used their death penalty statutes, thus controlling for the possibility that the mere existence of the death penalty may not carry the same weight unless

capital punishment is known to be a possibility. Using the murder rate (number of willful homicides per 100,000 population) in 2010 as their index, they assembled the table and graph shown on the next page. They reasoned that if capital punishment has a deterrent effect, the murder rates should be lower in the state with capital punishment laws.

The results, as shown in the table and graph below, were that in eight of the ten pairs of states selected for their study the murder rates were **higher** in the state with capital punishment laws than in the state without capital punishment laws. The researchers concluded that the existence of the death penalty does not work to deter murderers.

Critics of the study have complained that selection of a different set of ten neighboring states might have yielded a far different, perhaps even the opposite, result.

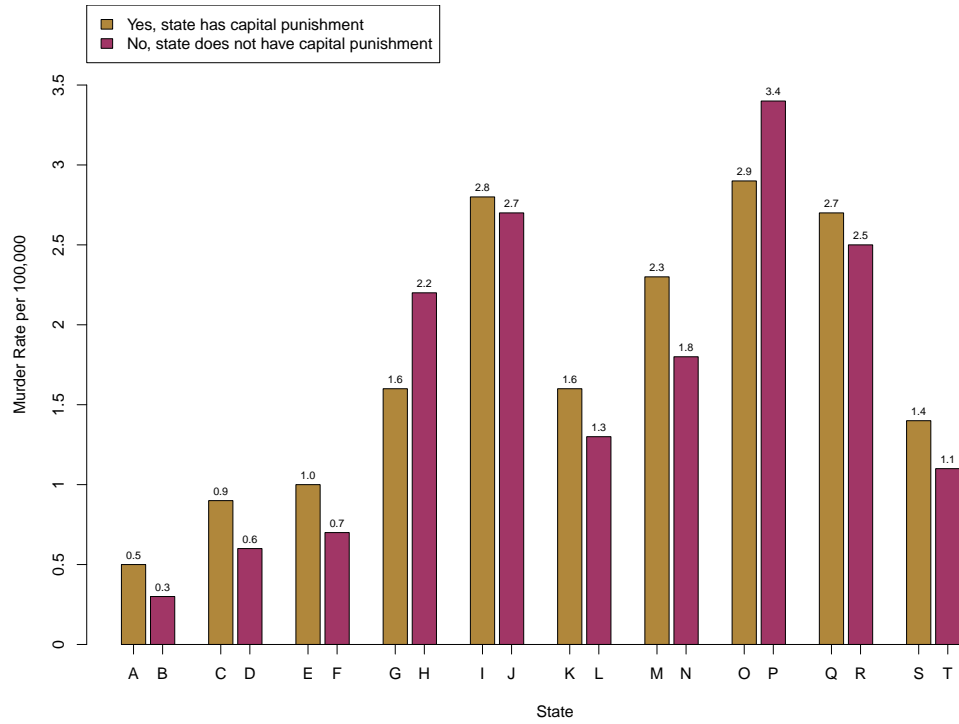
In replying to this criticism, Palmer and Crandall (2013) have recently reported a replication of their study, using a different set of ten states that share a common border but differ in whether their laws permit capital punishment or not. The results of this second study were essentially the same, murder rates being higher in the capital punishment state for seven of the ten comparisons.

Murder Rate in 2012 for Neighboring States With and Without Capital Punishment

Pair	State	Murder Rate	Capital Punishment	Pair	State	Murder Rate	Capital Punishment
1	A	0.5	Yes	6	K	1.6	Yes
	B	0.3	No		L	1.3	No
2	C	0.9	Yes	7	M	2.3	Yes
	D	0.6	No		N	1.8	No
3	E	1.0	Yes	8	O	2.9	Yes
	F	0.7	No		P	3.4	No
4	G	1.6	Yes	9	Q	2.7	Yes
	H	2.2	No		R	2.5	No
5	I	2.8	Yes	10	S	1.4	Yes
	J	2.7	No		T	1.1	No

Table reproduced with permission from Palmer and Crandall (2012)

Murder Rate in 2012 for Neighboring States with and without Capital Punishment
 Reproduced with permission from Palmer and Crandall (2012)



Study 2

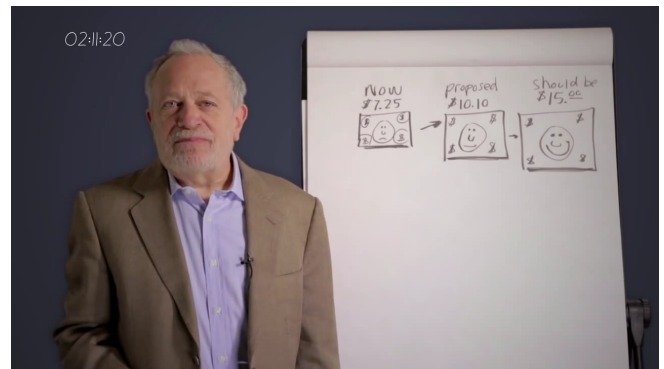
Questions

The *Favor* question asked, “The federal minimum wage is currently \$7.25 per hour. Do you favor or oppose raising the federal minimum wage?” The response options ranged from 1. “Very much opposed to raising the federal minimum wage” to 7. “Very much in favor of raising the federal minimum wage.” The *Amount* question asked, “What do you think the federal minimum wage should be? Please enter an amount between \$0.00 and \$25.00 in the text box below.”

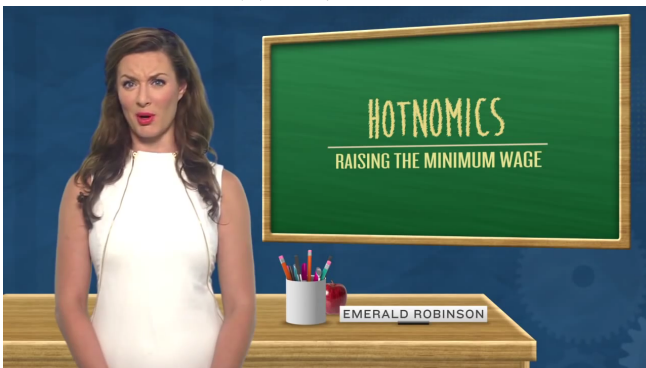
Treatments



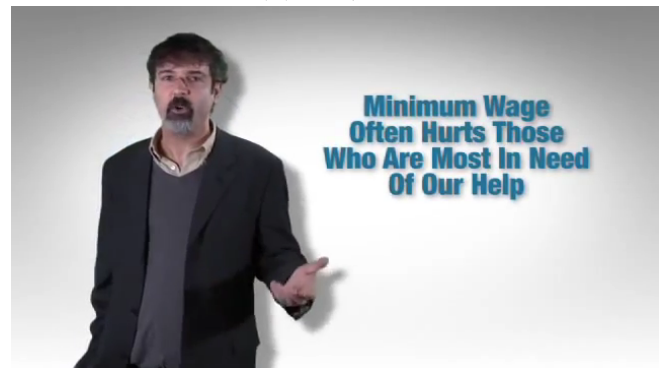
(a) Pro/Young



(b) Pro/Old



(c) Con/Young



(d) Con/Old

Figure 1: Study 2 Treatment Videos

Table 16: Study 2 Treatment Video URLs

Treatment Video	URL
Pro/Young	http://youtu.be/ZI9aDHLptMk
Pro/Old	http://youtu.be/G0qt153V3JI
Con/Young	http://youtu.be/hFG1Ka8AW6Q
Con/Old	http://youtu.be/Ct1Moeaa-W8

Study 3

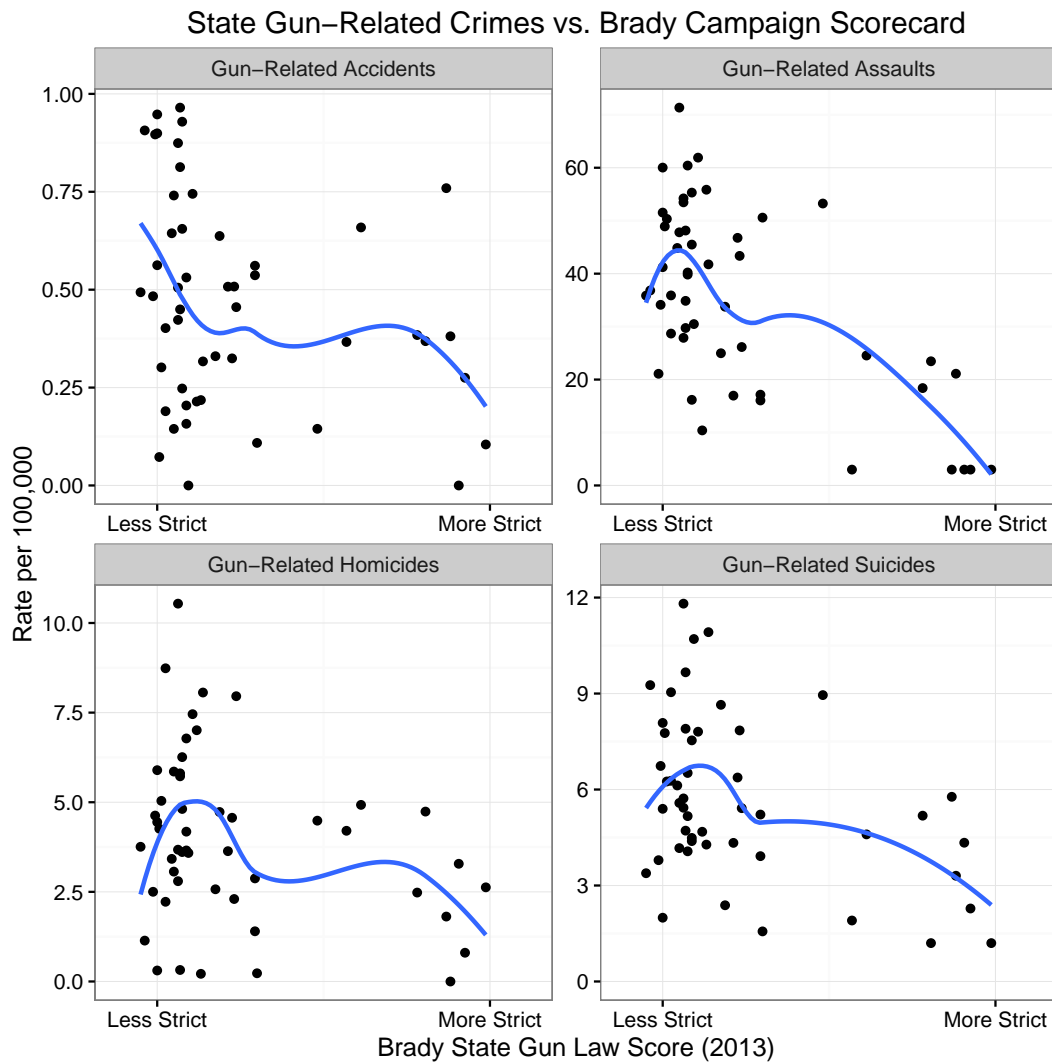
Questions

1. Do you support or oppose stricter gun control laws in the United States?
 - (a) I support stricter gun control laws
 - (b) I oppose stricter gun control laws
2. What do you think is more important: protecting the rights of Americans to own guns, or regulating gun ownership?
3. Do you support or oppose a nationwide ban on the sale of assault weapons?
4. Do you support or oppose a nationwide ban on the possession of handguns, except by the police and other authorized persons?
5. Suppose more Americans were allowed to carry concealed weapons if they passed a criminal background check and training course. If more Americans carried concealed weapons, would the United States be safer or less safe?

Pro Treatment

Kramer and Perry (2014) studied the relationship between gun laws and gun-related crimes in all 50 U.S. states. As a proxy for state-level gun regulations, they used the scorecard developed by the Brady Campaign to Prevent Gun Violence, a pro-gun-control group, which ranks states from 0 (negligible restrictions) to 100 (strong restrictions). They found that on average, states with stricter policies on gun ownership and possession tend to have lower levels of firearm-related accidents, assaults, homicides, and suicides.

The figure below displays their main findings:



Con Treatment

Kramer and Perry (2014) studied the relationship between gun laws and gun-related crimes in all 50 U.S. states. As a proxy for state-level gun regulations, they used the scorecard developed by the Brady Campaign to Prevent Gun Violence, a pro-gun-control group, which ranks states from 0 (negligible restrictions) to 100 (strong restrictions). They found that on average, states with stricter policies on gun ownership and possession tend to have higher levels of firearm-related accidents, assaults, homicides, and suicides.

The figure below displays their main findings:

State Gun-Related Crimes vs. Brady Campaign Scorecard

