Motivated Responding in Studies of Factual Learning*

February 11, 2017

Abstract

Large partisan gaps in reports of factual beliefs have fueled concerns about citizens' competence and ability to hold representatives accountable. In three separate studies, we reconsider the evidence for one prominent explanation of these gaps—motivated learning. We extend a recent study on motivated learning that asks respondents to deduce the conclusion supported by numerical data. We offer a random set of respondents a small financial incentive to accurately report what they have learned. We find that a portion of what is taken as motivated learning is instead motivated responding. That is, without incentives, some respondents give incorrect but congenial answers when they have correct but uncongenial information. Relatedly, respondents exhibit little bias in recalling the data. However, incentivizing people to faithfully report uncongenial facts increases bias in their judgments of credibility of what they have learned. In all, our findings suggest that motivated learning is less common than what the literature suggests, but also that there is a whack-a-mole nature to bias, with reduction in bias in one place being offset by increase at another place.

*We would like to thank Princeton Research in Experimental Social Science (PRESS) for financial support. We are also very grateful to Dan Kahan for generously sharing experimental design details; Doug Ahler, Martin Bisgaard, Katie McCabe, Peter Mohanty, and Markus Prior for insightful comments on a previous draft; participants of the PRESS workshop for suggestions about the experimental design; and finally, the editor of this journal and three reviewers for their critical feedback and guidance. The data and code necessary to replicate the results in this paper are available at https://dataverse.harvard.edu/dataverse/polbehavior.

Surveys of political knowledge frequently find large partisan differences in reported beliefs about policy-relevant facts. When Democrats control the White House, Democrats are likelier than Republicans to report that economic conditions are improving; the reverse holds true under a Republican administration (Bartels 2002). Similar patterns exist on answers to factual questions about healthcare (e.g., Nyhan 2010), foreign policy (e.g., Jacobson 2010), and social services (e.g., Jerit and Barabas 2012), among other issues. In general, when people are surveyed about facts, they are more likely to report having beliefs that are *congenial* to their existing beliefs and attachments than beliefs that are *uncongenial*. While we expect values and preferences to differ in a pluralistic society, we do not expect disagreement over established facts. Factual disagreement fuels concerns about the public's ability to form preferences that are in line with their values or interests (e.g., Hochschild 2001), hold representatives accountable (e.g., Bartels 2002), and productively deliberate with each other (Shapiro and Bloch-Elkon 2008; Muirhead 2013).

We consider two distinct explanations for differences in survey reports of factual beliefs: motivated learning and motivated responding.[1] According to the former, when people are presented with information, they are more likely to deduce that the information supports a congenial conclusion than an uncongenial conclusion (Jerit and Barabas 2012; Kahan 2013; Kahan et al. 2017). This bias in learning is thought to stem from the psychological drive to hold internally consistent beliefs that reflect positively on one's core attachments, such as political party, ideology, or cultural identity (Kunda 1990). Motivated learning is especially concerning from a normative standpoint. It suggests that even when people come across the same information, they can form very different beliefs about the conclusions supported by the information. (Note that learning about conclusions supported by information is different from believing that the conclusions are correct.) Such bias makes reducing gaps in factual beliefs yet harder—mere provision

[1]Partisan gaps in factual beliefs may also arise from selective exposure to information (see Stroud 2010, though see also Prior 2013), and motivated assessments of credibility of information (see, e.g., Lord et al. 1979).

of accurate information is even less likely to work, and may even backfire (see, e.g., Nyhan and Reifler 2010).

However, differences in survey reports of factual beliefs do not always reflect differences in what people believe. Instead, they may be artifacts of the survey response process. Respondents sometimes give congenial but inaccurate answers in response to factual questions even when they have accurate but uncongenial facts at hand (Bullock et al. 2015; Prior et al. 2015). Other times, respondents are ignorant, having no relevant cognitions, and they offer a congenial answer as their best guess (Luskin et al. 2013). In both cases, the survey response process inflates estimates of bias in factual beliefs. We extend this line of thinking to estimates of *learning*, the process that produces such beliefs. Like studies of stored cognitions, studies of learning may overstate bias if they do not account for motivated responding.

We reassess the extent to which people learn in a motivated manner by simultaneously measuring motivated learning and motivated responding in three separate studies. Building on the design in Kahan et al. (2017)—people are presented data from a social scientific study and asked to report the conclusion supported by the data—we incentivize a random set of respondents to honestly report the conclusion they actually think the data supports. Incentives reduce congeniality bias in reporting the study's result, suggesting that a portion of what appears to be motivated learning is in fact motivated responding. Correspondingly, we find no evidence that respondents selectively recall the data they were presented. However, there is considerable heterogeneity in responsiveness to the incentive treatment, depending on respondents' initial position on the issue under study. Moreover, incentivizing respondents to report uncomfortable conclusions supported by a study comes at a price: respondents become more skeptical about the study's credibility. In all, our findings suggest that motivated learning is less common than what previous work suggests, but that motivated evaluation of information remains common.

Motivated Learning, Reporting, and Interpretation

Psychological motivations are powerful in shaping how people process information. Broadly, people are motivated to both reach conclusions that are accurate and that are congenial to their beliefs (see Kunda 1990). Most research in political psychology focuses on the latter, which Kunda calls "directional" motivation. A great deal of research shows that directional motivations, like partisanship and ideology, bias evaluations of policy arguments (e.g., Lord et al. 1979; Taber and Lodge 2006; Bolsen et al. 2014) and of leaders (e.g., Bartels 2002; Lebo and Cassino 2007; Kim et al. 2010).

Motivated Learning

While most research on motivated reasoning focuses on subjective attitudes, motivated reasoning may also influence factual learning. Suppose, for instance, that there exist some data that support an unequivocal conclusion relevant to public policy. And we ask an individual with an existing opinion on said policy to learn the conclusion supported by the data. The extent to which directional goals outweigh accuracy goals will be correlated with the probability the individual learns that the data support a conclusion that is congenial to them, even if this conclusion is incorrect. We refer to this phenomenon as *motivated learning*.

Motivated learning has received recent attention. For instance, Nyhan and Reifler (2010) find that partisans ignore correct information that contradicts their ideological worldview.[2] Similarly, Jerit and Barabas (2012) find that partisans selectively learn party-relevant factual information. The authors present partisans with facts that reflect either positively or negatively on

[2]Nyhan and Reifler test motivated learning using two-sided information flows. They instill false beliefs in respondents via misleading news stories and then attempt to reduce misperceptions with corrective stories. While this design lends external validity to their study, a one-sided context without contradictory information would afford a stricter test of motivated learning.

Democrats or Republicans. They find that partisans are more likely to learn congenial than uncongenial facts, e.g., Democrats are more likely to learn about the success of the Troubled Asset Relief Program than the size of the trade deficit. A major challenge in such experiments is cleanly manipulating information congeniality, while holding constant other attributes, such as topic and difficulty. We now consider a recent study that does just that.[3]

Kahan et al. (2017) cleverly repurpose a "covariance detection task" (see Gilovich 1991) to test whether people engage in motivated reasoning when processing policy-relevant data. Respondents see a 2×2 table with data on the relationship between banning concealed carry and rates of crime. The table's column headings are manipulated so that the data either support the conclusion that banning concealed reduces crime or the conclusion that banning concealed carry increases crime. When asked what result the data support, respondents are more likely to answer correctly when the data support a congenial claim than when the data support an uncongenial claim. That is, liberal Democrats are more likely to report the correct result when banning concealed carry reduces crime than when banning concealed carry increases crime, while the reverse is true among conservative Republicans. Kahan et al. use the terms *motivated numeracy* and *motivated cognition* to describe the phenomenon.

The finding is consistent with several other studies that show that people evaluate congenial and uncongenial claims differently, using different evidentiary standards and investing different amounts of effort in processing the information. When evaluating congenial claims, people tend not to be as thorough in searching for evidence (e.g., Kruglanski and Webster 1996; Nickerson 1998), and evaluate available evidence more superficially and less skeptically (e.g., Chaiken and Maheswaran 1994). On the other hand, when evaluating uncongenial claims, people are more skeptical and invest greater processing effort (e.g., Ditto and Lopez 1992; Ditto et al.

[3]Many observational studies suggest motivated learning indirectly, by documenting partisan bias in factual beliefs (Bartels 2002; Shani 2006; Jerit and Barabas 2012), though other observational studies find little partisan bias in factual beliefs (e.g., Gaines et al. 2007; Blais et al. 2010).

1998; Dawson et al. 2002a). For instance, in the covariance detection task, respondents are more likely to only partially consider the data if doing so leads them to think that the data support a congenial conclusion (Dawson et al. 2002b).

Building on this psychological literature, Kahan et al. (2017) argue that our natural tendency to learn from data heuristically (i.e., using mental shortcuts) results in bias. For instance, when presented with tabular data, people tend to only consider the most salient datum in the table, for instance, the largest number, to deduce the conclusion supported by the data (Gilovich 1991, p. 31). If heuristic processing yields a congenial answer, people tend to stop processing, concluding that the data supports their beliefs. If heuristic processing instead yields an uncongenial result, people tend to look at the data more carefully to make sure they are correct. For instance, activating directional goals can motivate people to overcome the common error of neglecting 'cell D' in a 2×2 table (Mata et al. 2015a,b). This imbalance in scrutiny produces a *congeniality effect*: people learn congenial facts more readily than uncongenial facts. Kahan et al. argue that this effect increases with numeracy, because only respondents with sufficient numerical ability are capable of learning the correct result by considering all four cells.

Aside from asymmetric scrutiny, selective perception may also contribute to motivated learning. A long line of research suggests that expectation structures cognition (Hastorf and Cantril 1954; Bechlivanidis and Lagnado 2013; Kahneman 2013). In our hurry to learn from data, for example, we misperceive data in ways that are consistent with our prior beliefs. For instance, when trying to distill information from a contingency table, people potentially misread the column or row labels in a way that suggests a congenial result. Alternatively, one might misread the numbers in the table. In all, either due to selective perception or an imbalance in scrutiny, people are thought to be more likely to learn correctly when the data are congenial than when they are uncongenial.

Motivated Responding

Motivated responding concerns what survey respondents report on surveys, rather than what they have learned or know.[4] In particular, it occurs when people with same underlying beliefs giving congenial answers more often than uncongenial answers when asked about their beliefs. In all, it contends that survey responses to factual questions reflect a mix of what people believe and what they wish to be true (Luskin et al. 2013; Prior et al. 2015).

People engage in motivated responding for a variety of reasons. Some deliberately misreport as a way to express their attitudes. For example, a survey respondent who vehemently opposes President Barack Obama may not admit knowing that the unemployment rate declined between 2008 and 2016 on a survey. The respondent may instead report a rise in unemployment to express their opposition. This kind of expressive self-presentation has been described as "cheerleading" (e.g., Gerber and Huber 2009). Others may misreport just to be consistent within a survey, ensuring that later answers do not contradict their earlier ones (e.g., Sears and Lau 1983; Lau et al. 1990; Wilcox and Wlezien 1993; Palmer and Duch 2001). Yet others may engage in motivated responding to indicate their disbelief in information. For instance, in Kahan et al. (2017), a respondent may pick the congenial answer even after figuring out that the data support an uncongenial conclusion as a way to express their disbelief in the putative data.

In addition to actively misreporting what they believe, motivated responding may take more passive forms. For example, a respondent may withhold their beliefs by selecting "Don't Know" or skipping a question. Alternately, respondents who don't know the correct answer may report a congenial answer as their best guess. People may also engage in motivated responding without being consciously aware of it. For example, when asked a factual question in a survey,

[4]It is possible, even likely, that motivated responding also operates outside the survey context, explaining which beliefs people reveal in discussions with social networks, for example. However, in this paper, we focus on motivated responding in surveys.

respondents may scan their memory for a longer time to come up with instances of congenial beliefs than uncongenial beliefs. In all, a variety of reasons exist why people may engage in motivated responding. Our study does not disentangle the various reasons behind motivated responding. Instead, our aim is to merely estimate the bias due to motivated responding in estimates of motivated learning derived from ordinary survey instruments.

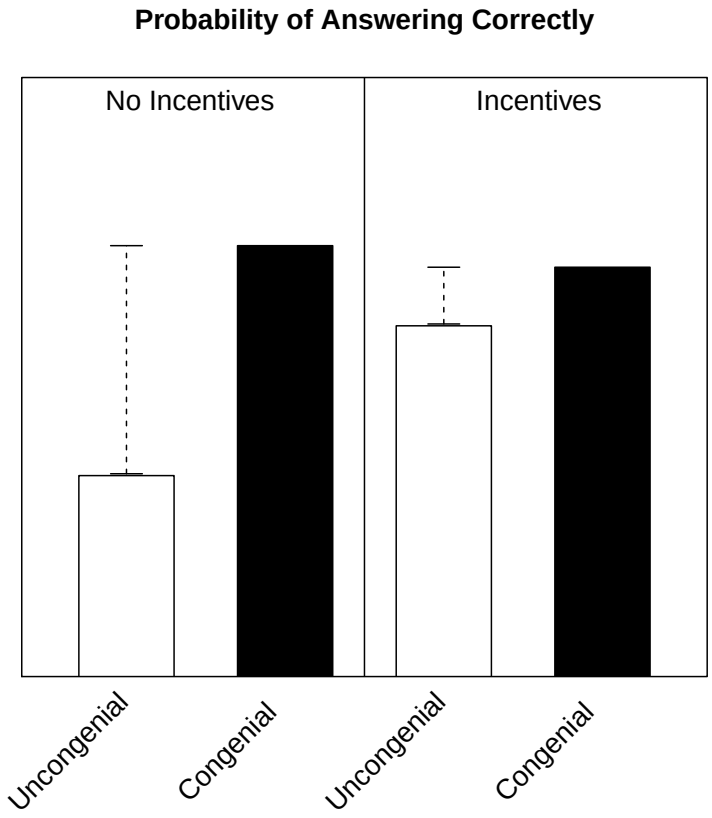
While both motivated learning and responding fall under motivated reasoning, there are two major differences. First, motivating responding is about the survey response process, as opposed to learning. Motivated responding affects what people *say* they have learned, not what they actually learn. Second, and relatedly, we suspect that part of motivated responding is just "cheap talk" that people engage in to publicly protect their core attachments and beliefs. And to the extent that these public pronouncements are shallow, based not in what people deeply believe but what people are prepared to say publicly, these reported 'beliefs' are unlikely to shape respondents' attitudes and behavior. On the other hand, uncongenial beliefs, which respondents are reluctant to express, may nonetheless influence attitudes and behavior. It is therefore important to distinguish between beliefs and instrumental or shallow responses.

A pair of recent studies find evidence that partisans engage in motivated responding on factual questions with political implications. Bullock et al. (2015) and Prior et al. (2015) find that partisans incorrectly report congenial beliefs, even when they know or could have inferred a more accurate answer. Both studies uncover motivated responding by boosting respondents' accuracy motivation. Bullock et al. (2015) do so by offering respondents bonus payments for correct answers and for admitting their ignorance (by marking 'don't know'), while Prior et al. (2015) use a combination of bonus payments for correct answers, and textual appeals for responding honestly. Importantly, neither of the treatments provide any additional information. Therefore, any change in responses can be attributed to a change in respondents' motivation, rather than a change in their knowledge. Both studies find that, consistent with motivated responding, accuracy incentives substantially reduce partisan bias—by as much as half.

Because studies of knowledge and learning rely on similar survey instruments, discovery of motivated responding on questions about stored knowledge suggests that estimates of motivated learning may also be inflated. To uncover motivated responding, we borrow a design feature of Prior and Lupia (2008), Bullock et al. (2015), and Prior et al. (2015), offering a random set of respondents small monetary incentives to accurately report their beliefs. To minimize the possibility that incentives affect processing of information, we present information about incentives *after* respondents have seen the information and can no longer revisit it. These incentives increase respondents' motivation to honestly answer the which conclusion they think the data support. Respondents who would ordinarily offer a congenial answer as their best guess or as an act of political expression may be nudged to think more carefully or be more truthful. We therefore hypothesize that this treatment will attenuate the congeniality effect that Kahan et al. (2017) and others observe.

More specifically, we hypothesize that incentives will increase the probability of answering correctly when the correct answer is uncongenial. If respondents learn the uncongenial result correctly, and then knowingly report an incorrect answer, incentives should encourage them to reveal their true belief. If respondents offer an incorrect, congenial response as their best guess, incentives should increase the probability that they will guess more evenhandedly. In both scenarios, incentives should increase correctness. However, we do not expect incentives to significantly increase correctness in the congenial condition, because here the congenial answer is the correct answer. We present our hypothesized pattern of results in a graphical format in Figure 1. The vertical axis represents the probability of answering correctly, and the dashed lines indicate the size of the congeniality effect. Incentives reduce the congeniality effect by increasing the probability of answering correctly in the uncongenial condition.

Figure 1: Hypothesized Pattern of Results. Probability of correctly reporting study's result by experimental condition. Dashed vertical lines indicate congeniality effect, i.e., difference in probability of answering correctly between congenial and uncongenial conditions. We hypothesize that incentives will reduce the congeniality effect by increasing the probability of answering correctly in the uncongenial, but not the congenial, condition.



Motivated Interpretation

Directional motivations affect not just what people learn and report but also how credible people think a study is. Previous work suggests that people are more likely to question a study's credibility when its results are uncongenial than when they are congenial (e.g., Lord et al. 1979; Kunda 1990; Ditto and Lopez 1992). This phenomenon likely stems from the more general tendency to spend greater time and effort scrutinizing and refuting uncongenial claims than congenial claims, known as disconfirmation bias (e.g., Edwards and Smith 1996; Taber and Lodge 2006). We therefore expect a congeniality effect on subjective study ratings: respondents will rate the

same study as more convincing and well executed when its result is congenial than when it is not. And we expect this congeniality effect to be more pronounced when we incentivize respondents to report the correct study result, because respondents who admit an uncongenial result should be even more likely to report that the study behind the result is unconvincing.[5]

Research Design

To distinguish between motivated learning and motivated responding, we conduct three experiments that build upon the original design of Kahan et al. (2017). In particular, we add an orthogonal manipulation, offering participants a small monetary incentive to accurately report what they have learned. The 2×2 design enables us to test whether incentives reduce the congeniality effect, which has been interpreted as evidence for motivated learning. In addition to measuring the outcome used by Kahan et al., we also measure subjective ratings of the study to examine whether the study's congeniality influences its perceived credibility. Below we elaborate on each of these extensions.

In Study 1, we asked respondents to read a summary of a hypothetical study on gun control. A preface to the study described its purpose: a city government is trying to decide whether or not to ban private citizens from carrying concealed weapons and wants to know if doing so would increase or decrease crime. After the preface, the study was briefly summarized: researchers compared changes in annual crime rates in cities that had banned concealed carry with changes in annual crime rates in cities that had not banned concealed carry. A 2×2 contingency table with the study's putative results came next.

[5]We do not incentivize study ratings, because unlike the question about study's result, ratings are inherently subjective. Therefore, the logic of providing incentives to be accurate on subjective questions is unclear. Correspondingly, even if we were to provide incentives, we would not be able to interpret the results unambiguously.

Following Kahan et al. (2017), we manipulated the conclusion supported by the study by switching the column labels. In Table 1, cities that banned concealed carry were more likely to experience a crime *decrease* relative to cities that did not ban concealed carry. This result can be learned by comparing the ratios of cities in the first row (75:223 or about 1:3) and the second row (21:107 or about 1:5). Flipping the column labels produces the opposite result—the data now indicate that cities that banned concealed carry saw *increases* in crime (see Table 2 below).

Table 1: Oppose Concealed Carry

	Increase in crime	Decrease in crime
Cities that did ban carrying concealed handguns in public	223 cities	75 cities
Cities that did not ban carrying concealed handguns in public	107 cities	21 cities

Table 2: Support Concealed Carry

	Decrease in crime	Increase in crime
Cities that did ban carrying concealed handguns in public	223 cities	75 cities
Cities that did not ban carrying concealed handguns in public	107 cities	21 cities

After presenting the summary of the study, we asked respondents whether cities with a ban were more likely to experience an *increase* or *decrease* in crime than cities without a ban. This question serves as our primary dependent variable. Note that it is strictly factual in nature. It simply asks which of two descriptions is consistent with the data. The question does not ask the respondents to assess a causal claim, evaluate gun control, or indicate their faith in the study. Respondents could not access the study description and table when picking which of the conclusions were supported by the data. At the end of the survey, respondents were debriefed and informed that the data were not real.

To measure motivated responding, we independently manipulated respondents' motivation to give the answer they thought was correct. We offered a random set of respondents a small

'nudge,' an additional $0.10 for the correct answer. Keeping the amount small has the virtue of not raising respondents' suspicions—some respondents may take a larger amount as cue that uncongenial answer is the right one. As to whether the small nudge is sufficient, we strictly cannot say, but note that Prior et al. (2015) uncover about same amount of motivated responding by emphasizing the importance of accuracy without any extra money as they do by offering another $1 for the correct answer. To ensure that incentives did not affect how respondents processed the contingency table in the treatment condition, we withheld any information about the incentive until after they had seen the table, and could no longer return to it. The control group was not offered incentives.

After measuring the primary dependent variable, we asked respondents to rate how convincing and how well done they found the concealed carry study. Each rating was measured on a 0-10 scale. We also asked respondents to recall the numbers in the contingency table at the end of the survey, in order to test whether respondents are more likely to remember congenial data than uncongenial data. To minimize respondent disengagement, we offered an additional $0.05 for each number recalled correctly.

In Study 2, we readministered the concealed carry task and added another task following it. In the second task, respondents were presented with a study on the impact of raising the minimum wage. Again, respondents were asked to indicate its result based on tabular data. The minimum wage task was very similar to the concealed carry task in design, with two important differences (aside from the change in topic). First, with the intention of making it easier to learn the correct result, we replaced cell frequencies with percentages in the table. The data suggest that the change had the intended effect, as there was a large increase in the percentage of correct responses. Second, we manipulated the study's congeniality by switching the row labels instead of the column labels in the table (see Supporting Information (SI) Section 1.1). We believe this is a cleaner manipulation as it holds constant the increase-to-decrease ratio in each row, and simply changes the policy associated with each ratio. While lowering the task's difficulty might change

the congeniality effect observed, we do not expect our changes to affect the degree of motivated responding. Lastly, randomization in the second task was conducted independently of the first, but the sequence of the two tasks was fixed.

In Study 3, we readministered both the concealed carry and minimum wage tasks on a more representative sample. Following our hypothesis that incentives influence responses in the uncongenial condition, we presented all respondents with an uncongenial version of the concealed carry task (based on their pre-treatment attitudes) and randomized incentives as in Studies 1 and 2. This simpler design allows us to conserve resources while testing our central theoretical claim that incentives increase correctness in the uncongenial condition. Additionally, we replicated the full 2×2 minimum wage task. The purpose of Study 3 was to gather confirmatory evidence and probe the generalizability of the estimates in Studies 1 and 2.

In each study, in order to identify respondents that would find each study's result congenial or uncongenial, we measured attitudes toward banning concealed carry and raising the federal minimum wage before the tasks (experiments). We expect respondents who oppose concealed carry to find a *decrease* in crime because of a concealed carry ban to be congenial, and an *increase* in crime uncongenial; we expect the opposite among respondents who support support concealed carry. The same logic applies to the minimum wage task. We measured party identification, political ideology, and demographics prior to the experiments in each study. SI Section 1.1 contains a complete description of the three studies and the complete wording of each task and question. In Studies 2 and 3, we omitted recall questions and ratings of the minimum wage study due to concerns about the length of the survey.

Data

We recruited respondents from Amazon's Mechanical Turk (MTurk) in Studies 1 and 2.[6] We recruited 'workers' for both studies by advertising a 'task' of completing a short survey on 'how people learn.' To assess whether our findings generalize beyond samples recruited on MTurk, we recruited respondents via Qualtrics in Study 3. While it is not a "gold standard," the Qualtrics sample is more representative of the general population, and respondents appear to be less attentive and detail-oriented than MTurk workers. Study 1 was fielded in December 2013–January 2014, Study 2 in March 2015–April 2015, and Study 3 in August 2016. For details of the recruited samples and how the samples compare to established benchmarks, see Table SI 1.

While none of the samples are nationally 'representative,' we can still learn a great deal from them. Multiple studies find that MTurk samples yield high-quality data and are more heterogeneous and representative than other common convenience samples, such as student samples (Buhrmester et al. 2011; Berinsky et al. 2012; Paolacci and Chandler 2014). And recently, Mullinix et al. (2015) replicate a broad array of experiments across different samples and find that treatment effects are broadly similar across samples. More generally, treatment effects vary across samples only when they are strongly conditioned by covariates that vary heavily across samples. We do not expect partisans on MTurk to differ from other partisans with respect to motivated reasoning. Multiple studies using MTurk samples find partisan bias in both stored knowledge (e.g., Chambers et al. 2014; Ahler and Sood; Chambers et al. 2015; Bullock et al. 2015) and political judgments (e.g., Arceneaux and Vander Wielen 2013; Lyons and Jaeger 2014; Crawford and Xhambazi 2015; Crawford et al. 2015; Thibodeau et al. 2015). In all, we think it likely that our treatment effects would be similar to studies using a nationally representative sample.

[6] MTurk is a micro-task market: workers complete small tasks, such as surveys, for money. For details of how samples are recruited on MTurk and general characteristics of the market, see Buhrmester et al. (2011) and Berinsky et al. (2012).

Given the theoretical expectation that we should only observe motivated learning among respondents with sufficient numerical ability to complete the covariance detection task, we screened for high-numeracy respondents using a numeracy quiz in Study 1. The numeracy quiz was composed of the five easiest questions in Weller et al. (2012) (see SI Section 1.2 for exact items). We invited respondents answering four or more items correctly to participate in the full study. We use a threshold of four because Kahan et al. (2017) find that the median respondent answers four items correctly on the full nine-item scale. In Studies 2 and 3, we invited all respondents to complete the main task, irrespective of numeracy, to ensure that our findings in Study 1 were not contingent on the relatively numerate MTurk sample. In SI Section 1.3, we track the number of respondents in the enrollment, screening, allocation, and analysis phases of the concealed carry experiment by study. We find that low- and high-numeracy respondents are similar in terms of party identification, ideology, and demographics (see Table SI 1 in SI Section 1.4), but we also present their results separately in SI Section 2.1.

In Study 1, we recruited 1,207 respondents and invited 785 (65% of sample) who passed the numeracy quiz to participate in the full survey. Our main analyses include 686 respondents (87% of screened sample) reporting a position on a concealed carry ban, which is necessary to code congeniality. Of them, 34% opposed concealed carry (i.e., favored ban) and 66% supported concealed carry (i.e., opposed ban). In Studies 2 and 3, we recruited another 947 and 1,062 respondents, respectively. Of those indicating a position, similar percentages to those in Study 1 opposed concealed carry: 36% in Study 2 and 40% in Study 3. The vast majority of respondents were in favor of raising the federal minimum wage: 85% in Study 2 and 65% in Study 3.[7]

[7]Attitudes in our study are similar to Americans' attitudes in two nationally representative surveys. A CBS/New York Times Survey from January 2013 finds that 34% of Americans favor "a federal law requiring a nationwide ban on people other than law enforcement carrying concealed weapons" (including 19% of Republicans a 52% of Democrats). And an Associated Press/GfK Poll from January 2015 finds that 77% favor raising the federal minimum wage (from $7.25/hour).

Results

We begin by presenting results from the concealed carry task, first pooling data from Studies 1 and 2, followed by results from Study 3. We separate out Study 3 because the concealed carry task only included the uncongenial condition. We follow it with results from the minimum wage task, and end with describing impact of treatment on respondents' subjective study ratings.

If people learn in a motivated manner, the percentage of respondents answering correctly when the study's result is congenial should be greater than when the study's result is uncongenial. Respondents who *oppose* concealed carry should be more likely to answer correctly if the data support the conclusion that crime is more likely to decrease in cities with concealed carry bans than in cities without such bans. Among respondents who *support* concealed carry, the reverse should be true. We thus examine whether the congeniality manipulation increases the probability of answering correctly.[8]

Before analyzing data from the covariance detection tasks, we check to see if partisanship, ideology, and demographics are balanced across the experimental conditions. The average p-value of cross-condition comparisons is .42 in Study 1, .47 in Study 2, .56 in Study 3, and .48 overall (see Table SI 2 in SI Section 1.5). We also confirm that the first experimental task did not affect behavior in the second (see Table SI 3 in SI Section 1.6). In all, the data suggest that randomization was successful, so we now move on to analyzing data from the first experiment.

[8]Kahan et al. (2017) define congeniality on the basis of party identification and ideology. However, overlap between a composite of partisanship and ideology and attitude toward concealed carry is considerably short of 100%. Across the three studies, 47% of self-described liberal Democrats oppose a ban on concealed carry, and 15% of self-described conservative Republicans favor such a ban. We therefore opt for coding congeniality in terms of the attitude most directly related to the data being. Recoding congeniality on the basis of party identification and ideology, following Kahan et al., results in a substantively similar congeniality effect (see SI Section 2.2).

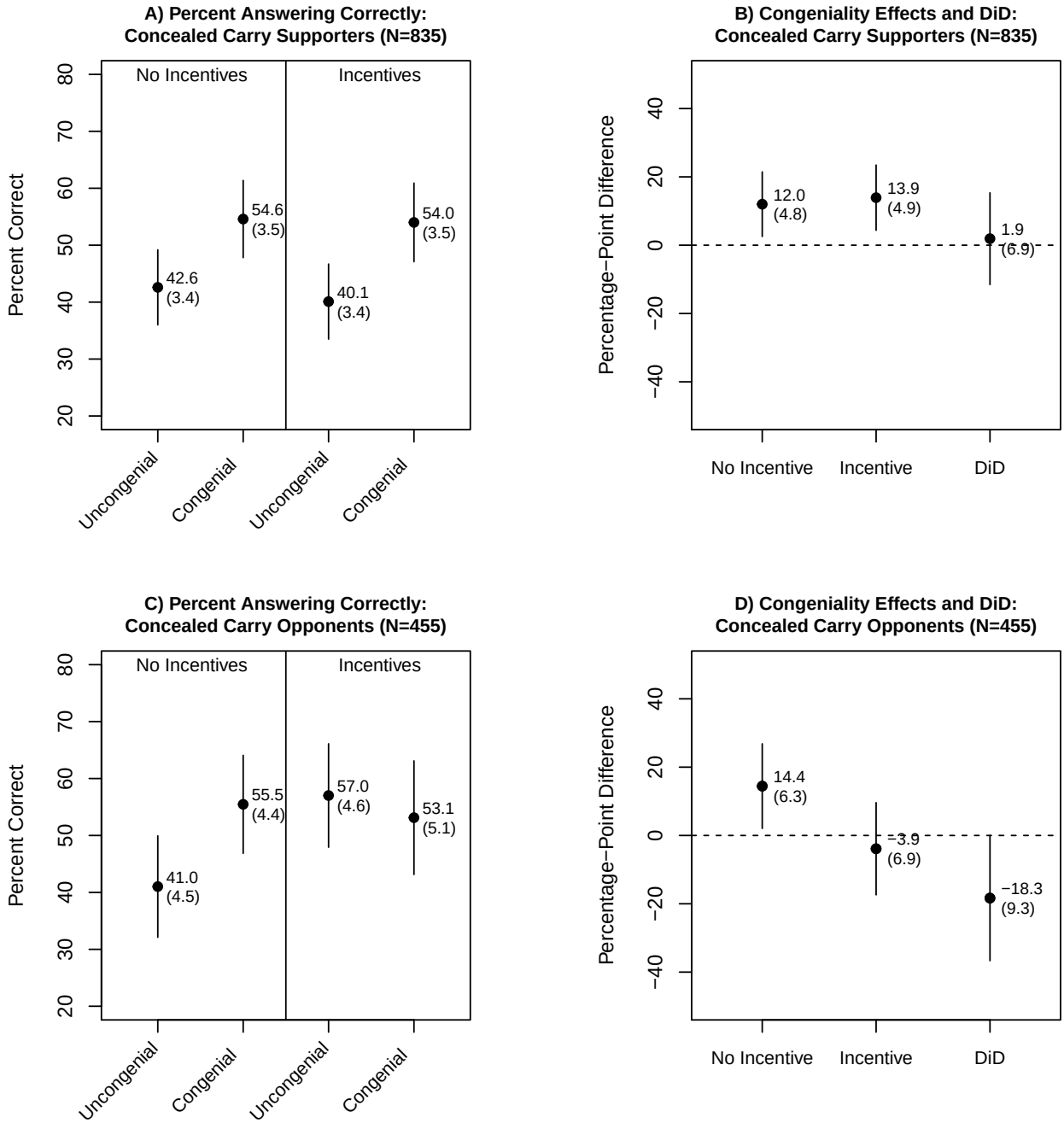
Figure 2 plots the percentage of respondents answering correctly in the concealed carry task by experimental condition across Studies 1 and 2.[9] We first consider the percentage correct among concealed carry supporters in the absence of incentives (Panel A, left). When the result is uncongenial (i.e., pro-ban), only 42.6% of respondents mark the right answer. When the result is congenial (i.e., anti-ban), the percentage increases to 54.6%. Thus, simply changing the study's result from uncongenial to congenial (by swapping column headers) increases the probability of answering correctly by 12.0 percentage points (plotted in Panel B). The pattern is similar among concealed carry opponents in the `No Incentives` condition (Panel C, left). When the result is uncongenial (i.e., anti-ban), 41.0% of the respondents answer correctly. When the result is congenial (i.e., pro-ban), the number increases to 55.5%. The congeniality effect is 14.4 percentage points (see Panel D). Thus, in the absence of incentives, the congeniality effects among both concealed carry supporters and opponents is statistically significant.

We next examine the extent to which incentives reduce these congeniality effects, which Kahan et al. (2017) and other scholars take as evidence of motivated learning. As we note earlier, the incentive treatment was administered in such a way that it did not affect how respondents (initially) processed the data—incentives were revealed after the respondents had seen the data and could not go back to it. If we were successful in administering the incentive treatment in the way we intended to, respondents should be as good at recalling the data in the `No Incentives` condition as in `Incentives` condition. We tested this hypothesis with the recall questions at the end of the survey. Data suggest that incentives had no impact on the accuracy of recall (see Tables SI 5 and SI 6 in SI Section 2.3.) Thus, it is unlikely that any treatment effects we see are explained by greater attention to the data when incentives are offered.

Examining Panel B of Figure 2, we see that offering incentives to concealed carry supporters does not reduce bias. The congeniality effect is 13.9 percentage points with incentives, which

[9]We subset high-numeracy respondents in Study 2 to ensure commensurability with Study 1. For concealed carry task results separated out by study, see Figures SI 5 and SI 6 in Section 2.4.

Figure 2: Concealed Carry Task Results (Studies 1 and 2 Pooled). Panels on the left display percentage of concealed carry opponents (Panel A) and supporters (Panel C) correctly indicating study result by experimental condition. Panels on the right display congeniality effect by incentive condition, as well as difference-in-differences (DiD), among concealed carry opponents (Panel B) and supporters (Panel D). Vertical lines indicate 95 percent confidence intervals. Only respondents who passed the numeracy screener and indicated a position on concealed carry are included (686 in Study 1 and 604 in Study 2).



is almost indistinguishable from the congeniality effect in the absence of incentives (difference-in-differences is 1.9). Since the congeniality effect remains substantial regardless of incentive condition, it appears that concealed carry supporters learn in a motivated manner.

Data from opponents of concealed carry, however, tell quite a different story (Figure 2, Panel D). Here, it appears that motivated responding masquerades as motivated learning. Incentives lower the congeniality effect from 14.4 percentage points to an insignificant -3.9 percentage points. The difference-in-differences is -18.3 percentage points and statistically significant (s.e. = 9.3, $p = .05$). Incentives completely wipe out the bias in answering the question about the study's result. Moreover, consistent with our hypothesis, the reduction in bias is entirely due to an increase in correctness in the uncongenial condition (16.0 percentage points, s.e. = 6.4, $p < .05$), rather than any change in the congenial condition.[10]

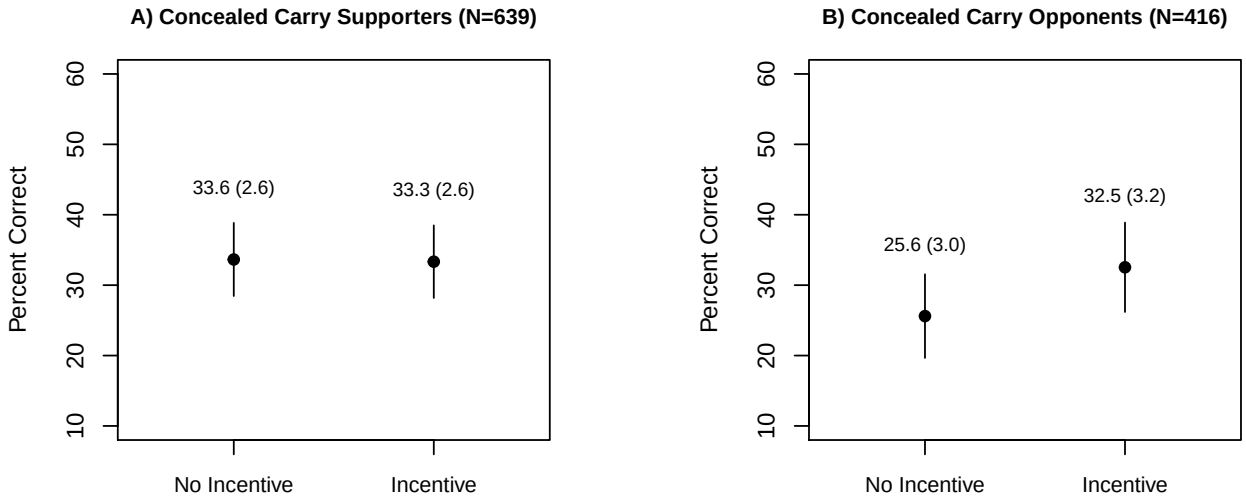
Results from Study 3 are similar to results from Studies 1 and 2 for both concealed carry supporters and opponents. Recall that Study 3 only included the uncongenial version of the concealed carry task. Figure 3 displays the percent correct among concealed carry supporters (Panel A) and opponents (Panel B) by incentive condition. Overall, Study 3 respondents had more trouble correctly identifying the study's result than respondents in Studies 1 and 2. For instance, only 33.6% of concealed carry supporters correctly identified the uncongenial study's results without incentives, while 42.6% did so in Studies 1 and 2. More relevant for our purposes, we see that concealed carry supporters are again essentially immune to the incentive treatment. The percentage correct among this group is almost identical when we offer accuracy incentives (33.3%). There is no evidence of motivated responding here.

Concealed carry opponents, on the other hand, once again exhibit a pattern of motivated responding. While only 25.6% answer correctly without incentives, 32.5% answer correctly when offered incentives, resulting in a treatment effect of 6.9 percentage points (s.e. = 4.4, $p < .06$). While the magnitude of the effect is smaller than in Studies 1 and 2, it is still non-trivial. In

[10]For equivalent logistic regressions, see Tables SI 7 and SI 8 in SI Section 2.5.

all, the evidence suggests that motivated responding introduces substantial amounts of bias in estimates of motivated learning.

Figure 3: Concealed Carry Task Results (Study 3). Panels display percentage of concealed carry opponents (Panel A) and supporters (Panel B) correctly indicating study result by experimental condition. Vertical lines indicate 95 percent confidence intervals.

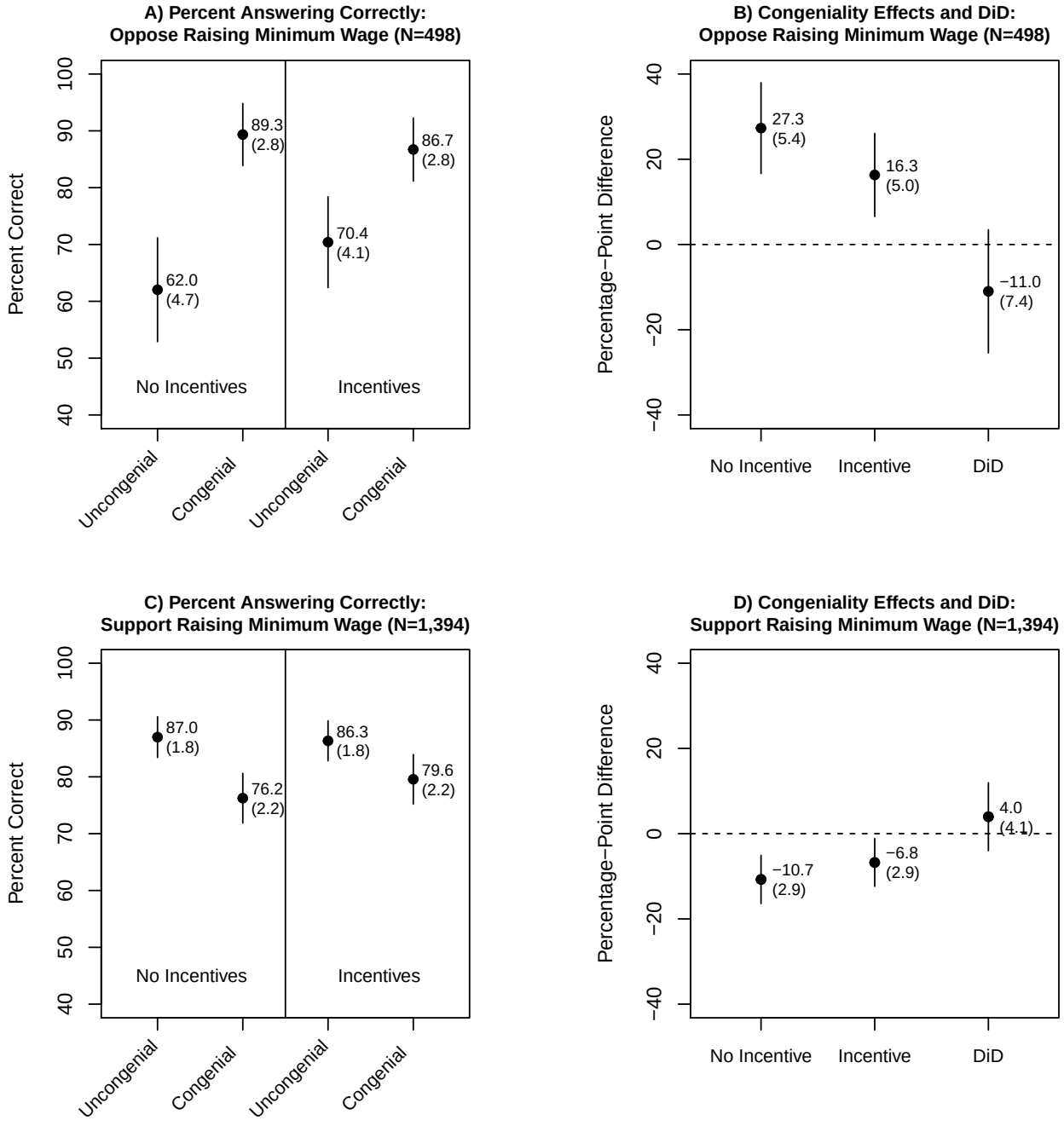


We now turn our attention to the minimum wage task, which we included in Studies 2 and 3, to probe the degree of motivated learning and responding on a different issue, using a slightly different experimental design. The overall percent correct was high in this task (87% in Study 2 and 77% in Study 3), which is unsurprising given our decision to make this task easier on respondents than the concealed carry task — we had replaced frequencies with percentages. Figure 4 summarizes the results from this experiment, pooling respondents in Studies 2 and 3.[11]

Our results from the minimum wage task are more consistent with motivated responding than motivated learning, on balance. Among opponents of raising the minimum wage, we see the familiar pattern that opponents of concealed carry display in the first task. The percentage of

[11]In Study 2, only 129 respondents (15%) oppose raising the minimum wage. Pooling them with opponents in Study 3 yields a large enough sample to analyze. We also obtain substantively similar results we see when analyze each study separately (see Figures SI 7 and SI 8 in SI 2.4).

Figure 4: Minimum Wage Task Results (Studies 2 and 3). Panels on the left display percentage of opponents (Panel A) and supporters (Panel C) of raising the federal minimum wage correctly indicating study result by experimental condition. Panels on the right display congeniality effect by incentive condition, as well as difference-in-differences (DiD), among opponents (Panel B) and supporters (Panel D). Vertical lines indicate 95 percent confidence intervals.



these respondents correctly identifying the result of the minimum wage study is 62.0% in the uncongenial condition and 89.3% in the congenial condition (Panel A). This dramatic congeniality effect is significantly attenuated by incentives. Specifically, it decreases from 27.3 to 16.3 percentage points, which is a 40% reduction (see Panel B). Again, this reduction is due to an increase in correctness in the uncongenial condition. Incentives increase the percent correct by 8.4 points in the uncongenial condition, but their effect is null in the congenial condition.

Supporters of raising the minimum wage do not behave in a manner consistent with motivated learning (Panel C). In fact, the congeniality effect is a significant -10.7 percentage points without incentives, indicating that respondents are actually *less* likely to correctly report a congenial result than an uncongenial result. In this case, the theoretical expectation for incentives is unclear, because the bias in the control condition is neither consistent with directionally motivated learning nor responding. We do not expect incentives to significantly alter behavior if there is no bias to reduce. Indeed, incentives do little to affect responses in either the uncongenial or congenial condition, so the congeniality effect remains negative with incentives (Panel D).[12] Nevertheless, this finding suggests that a congeniality bias in factual learning is yet less common than the literature suggests. In sum, across the two experimental tasks in our three studies, we see some evidence of motivated learning and a great deal of evidence for motivated responding.

[12]One possible explanation for this finding is that behavior in this task was affected by the previous task on concealed carry, since we did not randomize task order. We explore this possibility in SI Section 1.6 but find little support for it. Another, related possibility is that respondents mistook the row labels in the second task to be parallel to the row labels in the first task (i.e., cities enacting the given policy are always in the first row). Properly testing this hypothesis (e.g., by fully randomizing row and column labels, and task order) is beyond the scope of this study.

Biased Study Ratings

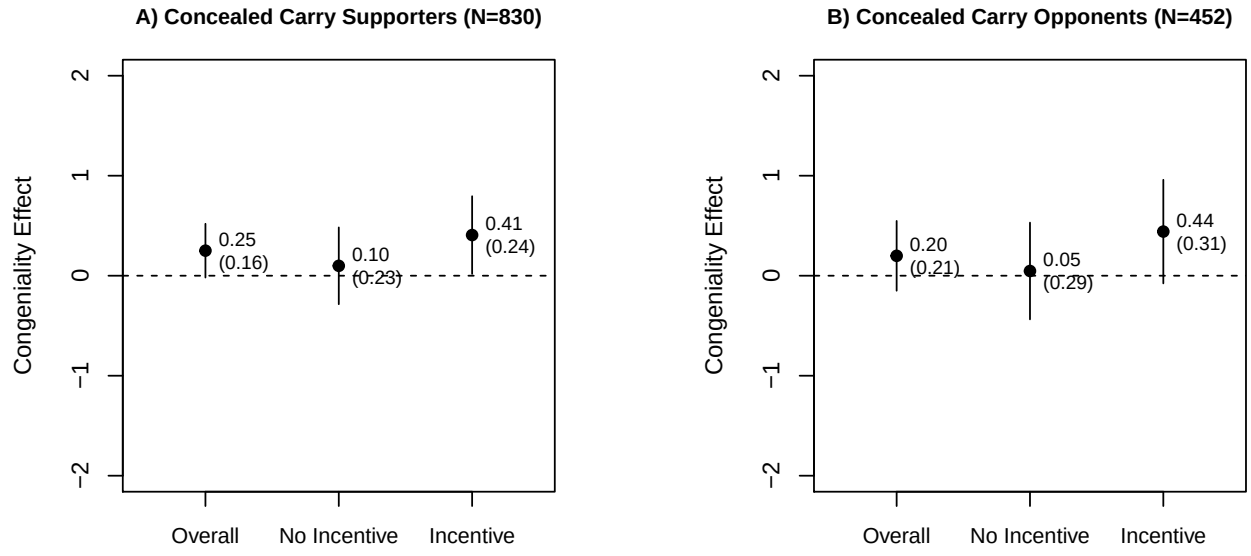
Does the congeniality of the hypothetical study's result affect how positively respondents rate the study? Recall that after the concealed carry task in both Study 1 and 2, we had asked respondents to rate how "well done" and "convincing" they found the study on a 0–10 scale (see SI Section 1.1 for wording). We find that the two ratings are strongly correlated, so we average them into a single rating (Cronbach's α = .82).

In Figure 5, we plot the effect of the congeniality manipulation on this average rating among concealed carry supporters (Panel A) and opponents (Panel B). We estimate the overall congeniality effect and then disaggregate it by incentive condition. Consistent with our expectations, congeniality affects ratings, but more so when we incentivize respondents. Among concealed carry supporters, the overall congeniality effect is a marginally significant .25 ($t = 1.53$, $p = .06$). Drilling down, we find a null effect without incentives and a significant increase of .41 with incentives ($t = 1.78$, $p = .04$). Similarly, among concealed carry opponents, the overall effect is .20, which is in the hypothesized direction but not significant. And again, the effect only appears in the `Incentive` condition ($t = 1.41$, $p = .08$).[13]

One possible explanation for this pattern is that when we incentivize respondents to admit having learned an uncongenial fact, they express displeasure via the study ratings. If so, we should see a greater congeniality effect among respondents who correctly report the study's result than among incorrect respondents. Indeed, we find that the congeniality effect in the `Incentive` condition only occurs among respondents answering correctly. Among concealed carry supporters answering correctly, the congeniality effect on study ratings in the `Incentive`

[13]Pooling concealed carry supporters and opponents, we find that ratings increase from 4.79 if the study is uncongenial to 5.02 if the result supported by the study is congenial ($t = 1.79$, $p = .04$). With incentives, the effect is .42 ($t = 2.30$, $p = .01$), and in the absence of incentives, the effect is only .08 ($t = .46$, $p = .32$). All t-tests reported here are one-tailed.

Figure 5: Congeniality Effect on Study Ratings. Figure displays effect of congeniality manipulation on average study ratings, which were measured on a 0-10 scale, among concealed carry opponents (A) and supporters (B). Data are pooled from Studies 1 and 2. Effects are calculated overall and disaggregated by incentive condition. Vertical lines indicate 90 percent confidence intervals.



condition is 1.05 ($t = 3.31$, $p < .001$). Among concealed carry opponents answering correctly, effect in the `Incentive` condition is 1.39 ($t = 3.43$, $p < .001$). These findings are merely suggestive, because answering correctly is of course an endogenous variable.

Discussion

Increasing affective polarization (see, e.g., Iyengar et al. 2012; Iyengar and Westwood 2014) has brought long-standing concerns over motivated reasoning to the fore. These concerns were amply highlighted during the recent presidential election, with numerous stories of people quickly latching on to congenial information, whatever the source, making the rounds. Motivated learning is related to such incidents, but it refers to something even more alarming: people coming across the same information supporting an unambiguous conclusion and yet walking away with different beliefs about what this information supports.

Motivated learning is particularly troubling because it has the potential to upend the benefits bestowed by the information age—easy access to reliable, trustworthy, objective information about a variety of politically relevant topics. Motivated learning means that people may possess different facts even after being exposed to the same information supporting an unambiguous conclusion, and may reach preferences that are very different from what they would had if they learned in an unbiased manner (Gilens 2001; Hochschild 2001). Given that motivated learning facilitates disagreement over what the facts are, it also implies that the possibility of democratic deliberation and compromise is slimmer still (Shapiro and Bloch-Elkon 2008; Muirhead 2013).

Our findings confirm that motivated learning occurs in some cases, but also suggest that estimates of motivated learning are upwardly biased. The concealed carry experiments suggest that supporters of concealed carry likely learn in a motivated manner. Manipulating the congeniality of the data through a minor change has a significant impact on the probability of reporting the correct answer among these respondents. And the accuracy incentives fail to change this tendency. On the other hand, a portion of what is thought to be motivated learning is instead motivated responding. When other respondents are offered a mere $0.10 to report their beliefs accurately, estimates of motivated learning decline sharply in some cases. And given that incentives could not have affected how respondents initially processed the data—respondents could not go back and look at the data after being informed about the incentives—incentives very likely identify the artifactual component of the evidence for motivated learning.

However, there are other potential explanations for why money may reduce estimates of motivated learning. The lure of making additional money may cause respondents to choose the answer that they believe the experimenter favors, rather than the one they think is right. Or, respondents may take monetary incentives as a cue that the congenial answer is incorrect. In both cases, the decline would be artifactual. We explore both possibilities, finding little empirical support for either (see SI 2.6, and in particular tables SI 9 and SI 10). On balance, the data suggest that incentives reduced bias in estimates of motivated learning, rather than increase it.

It is possible that the data, even accounting for incentives, still overstate the extent to which people learn (or mislearn) in a biased manner. It is likely that a non-trivial proportion of respondents in the experimental tasks simply tune out because they find the wording too complex, or because they are disinterested in the question. Such respondents may pick an answer by taking a blind guess or by relying on their priors. They are aware that they haven not really learned the result supported by the study. Other respondents may use cheap heuristics to deduce the correct result. And some of these respondents also likely know the fallibility of inference, and prorate their certainty in what they have learned accordingly. A simple correct/incorrect scoring does not capture either of these concerns, instead treating each answer as evidence of learning a particular result. To assess these concerns, we asked people how confident they were about the answers they gave after they had selected their answers in Study 3. Only 13% of respondents are certain of their answer in the concealed carry task without incentives. Even fewer, 10%, are certain and incorrect. This suggests that the proportion of people who become confidently misinformed due to motivated learning, which is the gravest concern, is not very large.[14]

Two other pieces of evidence suggest that motivated learning is less common than what conventional estimates reveal. Firstly, we find that respondents recall the data in a largely unbiased manner. Secondly, among those who support increasing the minimum wage, the congeniality effect is *negative*—respondents were more likely to report the correct result in the uncongenial condition than in the congenial condition. This is exactly the opposite of what the theory of motivated learning would have predicted, which suggests that motivated learning may not be a feature of learning more generally.

Lastly, data from the minimum wage task suggest that when the task is made easier, motivated learning all but disappears. This may happen because when the truth is transparent and easy to grasp, even people who are otherwise prone to motivated reasoning have trouble

[14]In the minimum wage task, 23% of respondents are certain of their answer, while only 4% are certain and incorrect. We present these results in more detail in Table SI 11 in SI Section 2.7.

denying it (see also Bisgaard 2015, who shows that when economic conditions are unambiguous, as in the 2008 recession, factual beliefs about the economy converge). The finding is consistent with bounded rationality, a mechanism proposed for motivated learning. When little effort is required, motivated learning is perhaps not as much an issue. This also suggests that treatments designed to teach people how to infer data correctly from the contingency table ought to prove efficacious. So should treatments that give people more time to learn, and incentivize attention.

More work is, however, needed to understand why incentives reduced the congeniality effect among those who oppose concealed carry, and not among those who support it. One possibility is that supporters of concealed carry have a stronger affective reaction to the issue and experience a greater directional pull toward their preferred conclusion. Another possibility is that this group of respondents differs on traits, such as need for cognition, that likely affect the extent to which people learn in a biased learning (e.g., Arceneaux and Vander Wielen 2013).

The fact that incentives reduce estimates of motivated learning has broad implications for survey measurement. It suggests that in order to precisely measure certain variables, such as factual beliefs with partisan implications, accuracy incentives may be the best way to obtain unbiased estimates. In extending this method to experiments on learning, our study contributes to the burgeoning literature on motivated responding, which combines the insights of survey satisficing with literature on group-based reasoning and affect (Bullock et al. 2015; Prior et al. 2015).

To put the results in perspective, however, note that estimates of motivated learning and motivated responding come from a unique task. The task of interpreting a contingency table is unusual for many people. It is also the case that the data in the table is arranged so that common heuristics (e.g., focusing on the upper-left cell) will always lead to the wrong answer. While news media occasionally present studies using tabular or graphical formats, it is not the norm. And in such cases, it is likely that training people in how to interpret tables or incentivizing how attentively they process the data will significantly reduce bias.

Finally, the results suggest that incentivizing accurate reporting among respondents comes at a price. Respondents rate the study as less well done and less convincing when incentivized to report the uncongenial fact. Thus, while it is possible to increase accuracy in reported factual beliefs, doing so may increase bias in interpretations of these facts. Our finding is consistent with other research showing that even when people agree on factual information, they nevertheless tend to interpret the information in a motivated manner (e.g., Gaines et al. 2007; Bisgaard 2015). While the observed reduction in motivated learning is welcome news for the prospect of democratic accountability, reducing differences between groups in one place may increase differences between them in another place.

References

Ahler, D. and Sood, G. The parties in our heads: Misperceptions about party composition and their consequences.

Arceneaux, K. and Vander Wielen, R. J. (2013). The effects of need for cognition and need for affect on partisan evaluations. *Political Psychology*, 34(1):23–42.

Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, 24(2):117–150.

Bechlivanidis, C. and Lagnado, D. A. (2013). Does the "why" tell us the "when"? *Psychological Science*, 24(8):1563–1572.

Berinsky, A. J., Huber, G. J., and Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, 20(3):351—368.

Bisgaard, M. (2015). Bias will find a way: Economic perceptions, attributions of blame, and partisan-motivated reasoning during crisis. *The Journal of Politics*, 77(3):pp. 849–860.

Blais, A., Gidengil, E., Fournier, P., Nevitte, N., Everitt, J., and Kim, J. (2010). Political judgments, perceptions of facts, and partisan effects. *Electoral Studies*, 29(1):1–12.

Bolsen, T., Druckman, J. N., and Cook, F. L. (2014). The influence on partisan motivated reasoning on public opinion. *Political Behavior*, 36(2):235–262.

Buhrmester, M., Kwang, T., and Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1):3–5.

Bullock, J. G., Gerber, A. S., Hill, S. J., and Huber, G. A. (2015). Partisan bias in factual beliefs about politics. *Quarterly Journal of Political Science*, 10(4):519–578.

Chaiken, S. and Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology*, 3(66):460–473.

Chambers, J. R., Swan, L. K., and Heesacker, M. (2014). Better off than we know: Distorted perceptions of incomes and income inequality in america. *Psychological Science*, 25(2):613–618.

Chambers, J. R., Swan, L. K., and Heesacker, M. (2015). Perceptions of U.S. social mobility are divided (and distorted) along ideological lines. *Psychological Science*, 26(4):413–423.

Crawford, J. T., Kay, S. A., and Duke, K. E. (2015). Speaking out of both sides of their mouths: Biased political judgments within (and between) individuals. *Social Psychological and Personality Science*, 6(4):422–430.

Crawford, J. T. and Xhambazi, E. (2015). Predicting political biases against the Occupy Wall Street and Tea Party movements. *Political Psychology*, 36(1):111–121.

Dawson, E., Gilovich, T., and Regan, D. T. (2002a). Motivated reasoning and performance on the Wason selection task. *Personality and Social Psychology Bulletin*, 28(10):1379–1387.

Dawson, E., Gilovich, T., and Regan, D. T. (2002b). Motivated reasoning and susceptibility to the "Cell A" bias. Manuscript submitted for publication.

Ditto, P. H. and Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, 4(63):568—584.

Ditto, P. H., Scepansky, J. A., Munro, G. D., Apanovitch, A. M., and Lockhart, L. K. (1998). Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology*, 1(75):53–69.

Edwards, K. and Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology*, 71(1):5.

Gaines, B. J., Kuklinski, J. H., Quirk, P. J., Peyton, B., and Verkuilen, J. (2007). Same facts, different interpretations: Partisan motivation and opinion on Iraq. *Journal of Politics*, 69(4):957–974.

Gerber, A. S. and Huber, G. A. (2009). Partisanship and economic behavior: Do partisan differences in economic forecasts predict real economic behavior? *American Political Science Review*, 103(3):407–426.

Gilens, M. (2001). Political ignorance and collective policy preferences. *American Political Science Review*, 95(2):379–396.

Gilovich, T. D. (1991). *How We Know What Isn't So: The Fallibility of Human Reason in Everyday Life.* The Free Press, New York.

Hastorf, A. H. and Cantril, H. (1954). They saw a game: A case study. *The Journal of Abnormal and Social Psychology*, 49(1):129.

Hochschild, J. L. (2001). Where you stand depends on what you see: Connections among values, perceptions of fact, and political prescriptions. In Kuklinski, J. H., editor, *Citizens and Politics: Perspectives from Political Psychology*. Cambridge University Press, New York.

Iyengar, S., Sood, G., and Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly*, 76(3):405–431.

Iyengar, S. and Westwood, S. J. (2014). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*.

Jacobson, G. (2010). Perception, memory, and partisan polarization on the Iraq War. *Political Science Quarterly*, 1(125):31–56.

Jerit, J. and Barabas, J. (2012). Partisan perceptual bias and the information environment. *Journal of Politics*, 74(3):672–684.

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, 8(4):407–424.

Kahan, D. M., Peters, E., Dawson, E. C., and Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, Forthcoming.

Kahneman, D. (2013). The marvels and flaws of intuitive thinking. In Brockman, J., editor, *Thinking: The New Science of Decision-making, Problem-solving, and Prediction*. Harper Collins.

Kim, S., Taber, C. S., and Lodge, M. (2010). A computational model of the citizen as motivated reasoner: Modeling the dynamics of the 2000 presidential election. *Political Behavior*, 32(1):1–28.

Kruglanski, A. W. and Webster, D. M. (1996). Motivated closing of the mind: "seizing" and "freezing". *Psychological Review*, 103(2):263–283.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3):480–498.

Lau, R. R., Sears, D. O., and Jessor, T. (1990). Fact or artifact revisited: Survey instrument effects and pocketbook politics. *Political Behavior*, 12(3):217–242.

Lebo, M. J. and Cassino, D. (2007). The aggregated consequences of motivated reasoning and the dynamics of partisan presidential approval. *Political Psychology*, 28(6):719–746.

Lord, C. G., Ross, L., and Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11):2098–2109.

Luskin, R. C. and Bullock, J. G. (2011). "Don't know" means "don't know": DK responses and the public's level of political knowledge. *The Journal of Politics*, 73(2):547–557.

Luskin, R. C., Sood, G., and Blank, J. (2013). The waters of Casablanca: Political misinformation (and knowledge and ignorance). Unpublished manuscript.

Lyons, J. and Jaeger, W. P. (2014). Who do voters blame for policy failure? Information and the partisan assignment of blame. *State Politics & Policy Quarterly*.

Mata, A., Garcia-Marques, L., Ferreira, M. B., and Mendonça, C. (2015a). Goal-driven reasoning overcomes cell D neglect in contingency judgements. *Journal of Cognitive Psychology*, 27(2):238–249.

Mata, A., Sherman, S. J., Ferreira, M. B., and Mendonça, C. (2015b). Strategic numeracy: Self-serving reasoning about health statistics. *Basic and Applied Social Psychology*, 37(3):165–173.

Muirhead, R. (2013). The case for party loyalty. In Levinson, S., Parker, J., and Woodruff, P., editors, *Loyalty*. New York University Press, New York.

Mullinix, K. J., Leeper, T. J., Druckman, J. N., and Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, 2(2):109–138.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2):175–220.

Nyhan, B. (2010). Why the "death panel" myth wouldn't die: Misinformation in the health care reform debate. *The Forum*, 8(1):Article 5.

Nyhan, B. and Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2):303–330.

Palmer, H. D. and Duch, R. M. (2001). Do surveys provide representative or whimsical assessments of the economy? *Political Analysis*, 9(1):58–77.

Paolacci, G. and Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, 23(3):184–188.

Pasek, J., Sood, G., and Krosnick, J. A. (2015). Misinformed about the Affordable Care Act? Leveraging certainty to assess the prevalence of misperceptions. *Journal of Communication*, 65(4):660–673.

Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, 16:101–127.

Prior, M. and Lupia, A. (2008). Money, time, and political knowledge: Distinguishing quick recall and political learning skills. *American Journal of Political Science*, 52(1):169–183.

Prior, M., Sood, G., and Khanna, K. (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.

Sears, D. O. and Lau, R. R. (1983). Apparently self-interested political preferences. *American Journal of Political Science*, 27(2):223–252.

Shani, D. (2006). Knowing your true colors: Can knowledge correct for partisan bias in political perceptions? Presented at the Annual Meeting of the Midwest Political Science Association, Chicago, IL.

Shapiro, R. Y. and Bloch-Elkon, Y. (2008). Do the facts speak for themselves? Partisan disagreement as a challenge to democratic competence. *Critical Review*, 20(1-2):115–139.

Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, 60(3):556–576.

Taber, C. S. and Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3):755–769.

Thibodeau, P., Peebles, M. M., Grodner, D. J., and Durgin, F. H. (2015). The wished-for always wins until the winner was inevitable all along: Motivated reasoning and belief bias regulate emotion during elections. *Political Psychology*, 36(4):431–448.

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., and Peters, E. (2012). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making*, 26(2):198—212.

Wilcox, N. and Wlezien, C. (1993). The contamination of responses to survey items: Economic perceptions and political judgments. *Political Analysis*, 5(1):181–213.

Supporting Information

1 Additional Details about the Sample and Research Design

1.1 Question Wording

1.1.1 Gun Attitudes and Ownership (Studies 1, 2, and 3)

Do you favor or oppose a federal law requiring a nationwide ban on people other than law enforcement officers carrying concealed weapons?

- Ban people other than law enforcement officers from carrying concealed weapons
- Do not ban people other than law enforcement officers from carrying concealed weapons
- Unsure

How important is the issue to you?

- Extremely important
- Very important
- Somewhat important
- Not too important
- Not important at all

Do you or someone else in your household own a gun?

- Yes
- No

1.1.2 Minimum Wage Attitudes (Studies 2 and 3 Only)

The federal minimum wage is currently $7.25 per hour. Do you favor or oppose raising the federal minimum wage?

- Favor raising the federal minimum wage
- Oppose raising the federal minimum wage
- Neither favor nor oppose raising the federal minimum wage

How important is the issue to you?

- Extremely important
- Very important
- Somewhat important
- Not too important
- Not important at all

1.1.3 Concealed Carry Task (Studies 1, 2, and 3)

A city government is trying to decide whether to pass a law banning private citizens from carrying concealed handguns in public. Government officials were unsure whether the law would be more likely to decrease crime by reducing the number of people carrying weapons or increase crime by making it harder for law-abiding citizens to defend themselves from violent criminals.

Researchers completed a study of two groups of cities to answer that question. The study involved comparing changes in annual crime rates for one group of cities that had banned concealed handguns with changes in annual crime rates for a second group of cities that had not banned concealed handguns.

In each group, the number of cities in which the crime rate decreased and the number of cities in which the crime rate increased are recorded in the table below. The exact number of cities in each group is not the same, but this does not prevent assessment of the results.

We would like to know whether cities that enacted a ban were more likely to have a decrease or increase in crime than cities without bans.

[Display if randomly assigned to pro-gun condition:]

	Increase in crime	Decrease in crime
Cities that did ban carrying concealed handguns in public	223 cities	75 cities
Cities that did not ban carrying concealed handguns in public	107 cities	21 cities

[Display if randomly assigned to anti-gun condition:]

	Decrease in crime	Increase in crime
Cities that did ban carrying concealed handguns in public	223 cities	75 cities
Cities that did not ban carrying concealed handguns in public	107 cities	21 cities

[Page Break]

What result does the study support? [**Incentive Treatment**:] We will give you a bonus of $0.10 for the correct answer.

- Cities that enacted a ban on carrying concealed handguns were more likely to have a decrease in crime than cities without bans.
- Cities that enacted a ban on carrying concealed handguns were more likely to have an increase in crime than cities without bans.

[Page Break]

On a scale of 0 to 10, where 0 is "very poorly done" and 10 is "very well done," how well do you think the study on concealed handguns was done?

On a scale of 0 to 10, where 0 is "completely unconvincing" and 10 is "completely convincing," how convincing is this study as evidence of the effect of banning concealed handguns on crime?

1.1.4 Minimum Wage Task (Studies 2 and 3 Only)

A city government is trying to decide whether to pass a law increasing the minimum wage. Government officials were unsure whether the law would be more likely to decrease jobs or increase jobs.

Researchers completed a study of two groups of states to answer that question. The study involved comparing changes in jobs for one group of states that had increased the minimum wage with changes in jobs for a second group of states that had not increased the minimum wage.

In each group, the percentage of states in which jobs decreased and the percentage of states in which jobs increased are recorded in the table below. The exact number of states in each group is not the same, but this does not prevent assessment of the results.

We would like to know whether states that increased the minimum wage were more likely to have a decrease or increase in jobs than states without minimum wage increases.

[Display if randomly assigned to pro-raise condition:]

	Increase in jobs	Decrease in jobs
States that did not increase minimum wage	16%	84%
States that did increase minimum wage	37%	63%

[Display if randomly assigned to anti-raise condition:]

	Increase in jobs	Decrease in jobs
States that did increase minimum wage	16%	84%
States that did not increase minimum wage	37%	63%

[Page Break]

What result does the study support? [**Incentive Treatment**:] We will give you a bonus of $0.10 for the correct answer.

- States that increased the minimum wage were more likely to have a decrease in jobs than states without minimum wage increases.
- States that increased the minimum wage were more likely to have an increase in jobs than states without minimum wage increases.

1.2 Numeracy Screener

We include our numeracy scale below. It is a shortened version of the scale used by Kahan et al. (2017) and developed by Weller et al. (2012). We include the five easiest items that Weller et al. find, in order to identify low-numeracy respondents. Two concerns vitiate commensurability of our scores and those of Kahan et al.: guessing and item sampling. Research suggests that lucky guessing is a minor concern on open-ended items (Luskin and Bullock 2011). And if we assume that the items are Guttman scaled, i.e., a person who answers a more difficult item correctly will also answer easier items correctly, our threshold will be exactly the same as the median score in Kahan et al.. We take Guttman scaling on numeracy items to not only be plausible but likely.

1. If we roll a fair, six-sided die 1,000 times, on average, how many times would the die come up as an even number?

   Answer: 500

2. There is a 1% chance of winning a $10 prize in the Megabucks Lottery. On average, how many people would win the $10 prize if 1,000 people each bought a single ticket?

   Answer: 10

3. If the chance of getting a disease is 20 out of 100, this would be the same as having a _____% chance of getting the disease.

   Answer: 20

4. If there is a 10% chance of winning a concert ticket, how many people out of 1,000 would be expected to win the ticket?

   Answer: 100

5. In the PCH Sweepstakes, the chances of winning a car are 1 in a 1,000. What percent of PCH Sweepstakes tickets win a car?

   Answer: .1

1.3 Respondents Enrolled, Screened, Allocated, and Analyzed by Study

Figures SI 1, SI 2, and SI 3 are flow diagrams displaying the number of respondents in each phase of Studies 1, 2, and 3, respectively. We display the number of respondents assessed for eligibility, invited to participate, assigned to an experimental condition in the concealed carry task, and included in our main analyses.

Figure SI 1: Flow Diagram of Concealed Carry Experiment in Study 1

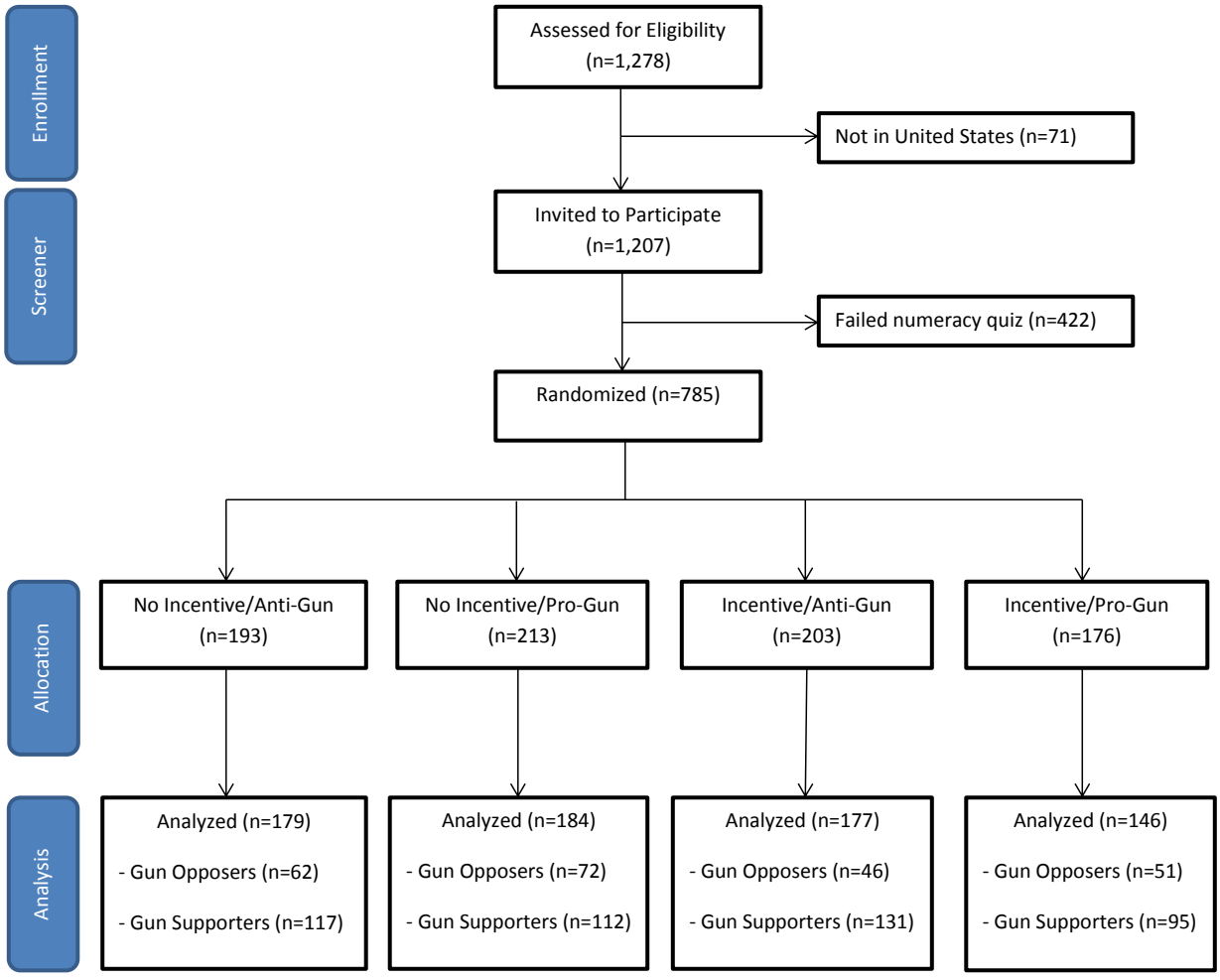


Figure SI 2: Flow Diagram of Concealed Carry Experiment in Study 2

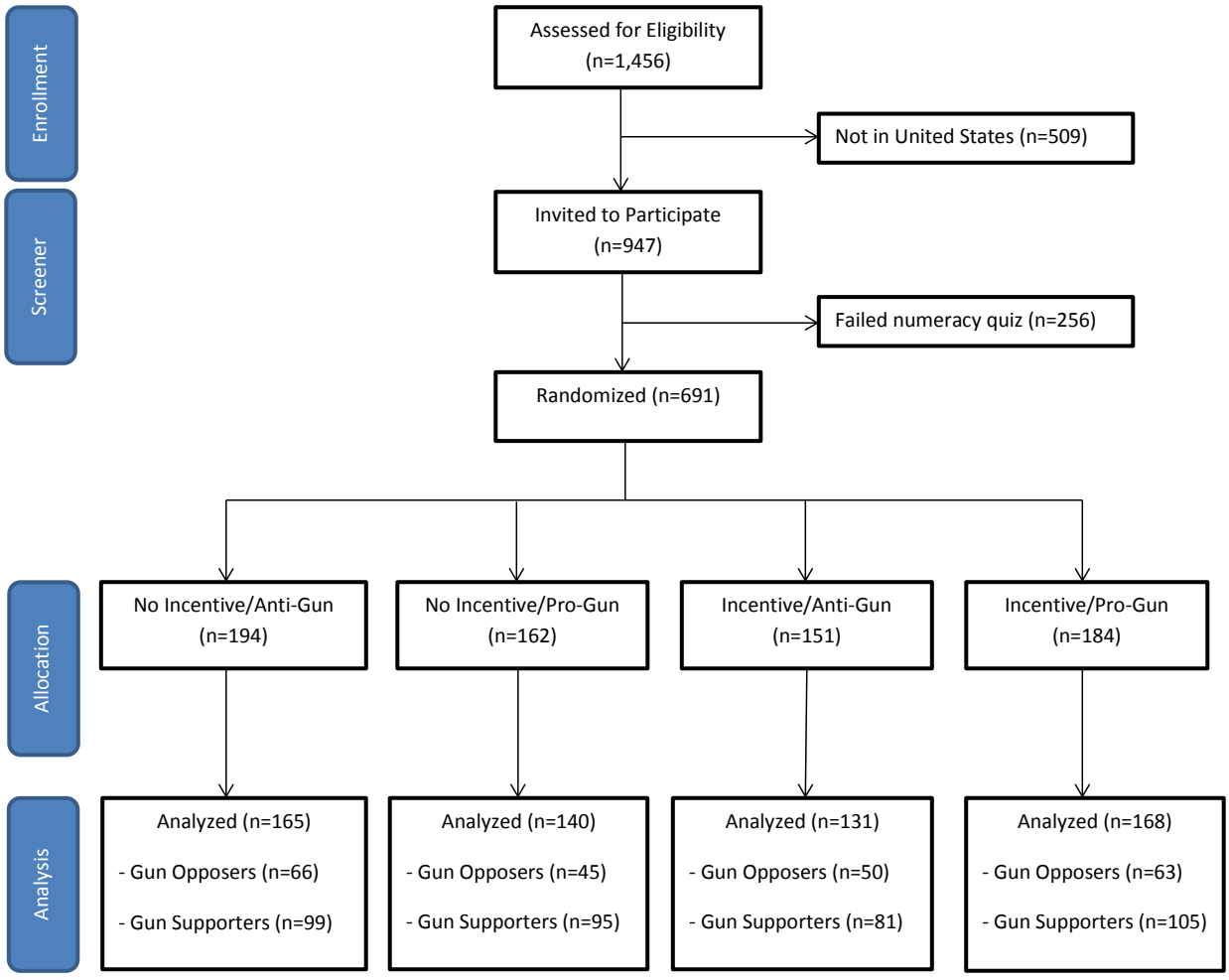
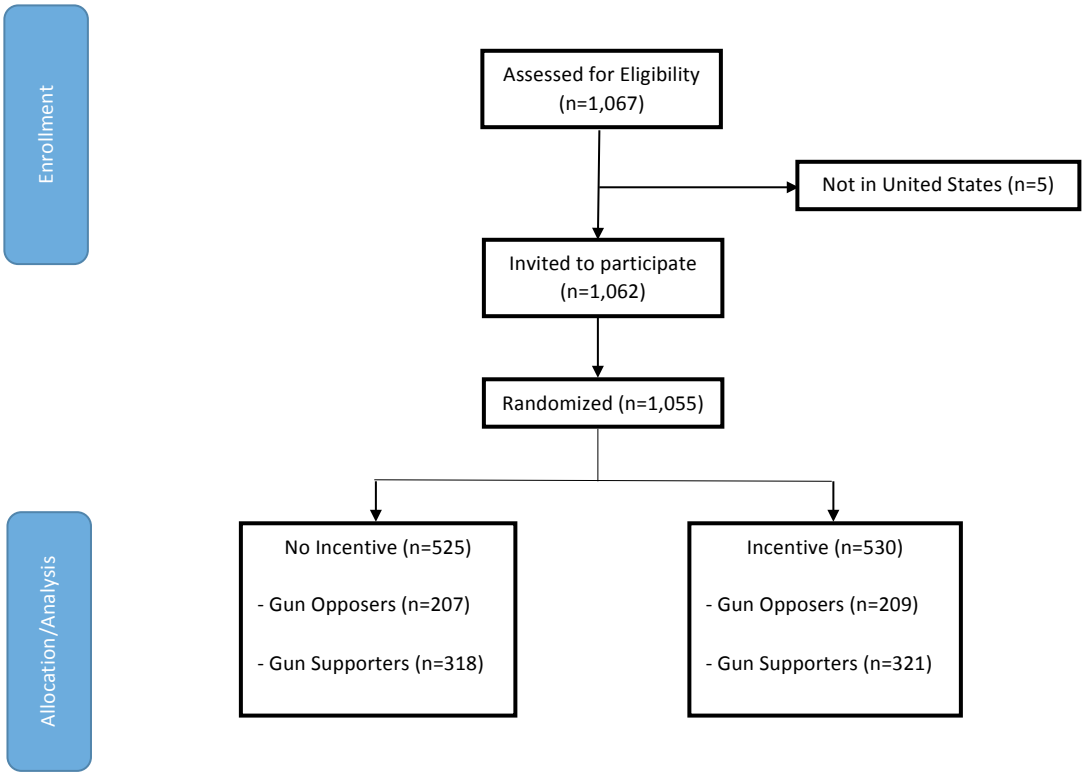


Figure SI 3: Flow Diagram of Concealed Carry Experiment in Study 3



1.4 Sample Characteristics

To shed light on the quality of our respondent samples, we compared marginals on key variables to the numbers from better samples in Table SI 1. We used the 2012 American National Election Study to compare partisanship and ideology, and the U.S. Census 2015 American Community Survey for demographics. Like other MTurk samples, the Study 1 and 2 samples are more young, white, male, and liberal than the general population. Study 3 is more representative in terms of party and ideology, but more educated, female, and white than the general population. In Studies 2 and 3, we can compare high-numeracy respondents with the entire sample (55% of sample). As before, we define high-numeracy as answering four or more items correctly in the five-item scale. In Study 2, there are no appreciable differences by numeracy. However, high-numeracy respondents in Study 3 are more educated, white, and male than low-numeracy respondents.

Table SI 1: Sample Characteristics by Study

	Study 1	Study 2	Study 2 High-Num.	Study 3	Study 3 High-Num.	National Sample
Democrat	65%	60%	62%	51%	50%	46%
Republican	21%	24%	22%	35%	39%	39%
Liberal	51%	50%	52%	23%	22%	27%
Conservative	16%	17%	16%	29%	28%	39%
HS or Less	9%	12%	10%	25%	17%	41%
Some College	40%	41%	41%	39%	37%	31%
College Degree	40%	39%	39%	26%	33%	18%
Post-Graduate	11%	9%	10%	10%	13%	10%
18-29 years old	57%	55%	55%	10%	11%	22%
30-44 years old	33%	33%	34%	34%	33%	25%
45-64 years old	9%	11%	10%	44%	47%	34%
65+ years old	1%	1%	1%	13%	10%	19%
Female	32%	38%	34%	66%	57%	52%
White	75%	73%	73%	76%	84%	62%
Black	5%	6%	5%	10%	5%	12%
Hispanic/Latino	5%	7%	7%	7%	4%	18%
Asian	11%	10%	11%	3%	4%	5%
Other/Mixed	4%	5%	5%	3%	3%	3%

1.5 Covariate Balance

Table SI 2 presents estimates of the extent to which covariates were balanced across conditions in the concealed carry task, separately by study. Each row reports the F-statistic and associated p-value from a linear regression of the indicated covariate on the congeniality treatment, incentive treatment, and interaction between the two treatments. In Study 3, which only contained the uncongenial version of the concealed carry task, the only regressor is the accuracy treatment.

Table SI 2: Covariate Balance across Experimental Conditions in Concealed Carry Task

	Study 1 (N=785)	Study 2 (N=947)	Study 3 (N=1,062)
Partisanship (7-point)	$F = .45, p = .72$	$F = .88, p = .45$	$F = .04, p = .85$
Ideology (5-point)	$F = .30, p = .83$	$F = 1.29, p = .28$	$F = .002, p = .96$
Education (4-point)	$F = 2.42, p = .07$	$F = .78, p = .50$	$F = 2.51, p = .11$
Age (continuous)	$F = 1.34, p = .26$	$F = .56, p = .64$	$F = 1.00, p = .32$
Gender (binary)	$F = 1.53, p = .20$	$F = .47, p = .71$	$F = 1.04, p = .31$
Non-White (binary)	$F = .64, p = .42$	$F = 1.40, p = .24$	$F = .06, p = .81$

1.6 Assessing Spillover

Because the concealed carry task always preceded the minimum wage task in Studies 2 and 3, we assess the extent to which randomization in the former task affected behavior in the latter task. We estimate treatment effects in both tasks simultaneously via logistic regression. The outcome variable indicates correctness in the minimum wage task (1 = correct). The regressors are the congeniality treatment, the incentive treatment, and their interaction, in both tasks. We allow the minimum wage treatments to vary by concealed carry condition.

Table SI 3 displays the results for supporters of raising the minimum wage in Studies 2 and 3, and opponents in Study 3. In each regression, none of the coefficients reach significance, except for the congeniality treatment in the minimum wage task. Importantly, the concealed carry task treatments all have null effects, confirming that the first task did not spill over into the second task. Moreover, the null interactive effects indicate that the concealed carry treatments did not alter the effects of the minimum wage treatments in any of the samples.

Table SI 3: Logistic Regression of Minimum Wage Correctness on Experimental Treatments in Both Tasks

	Coefficient (SE)	z value	Pr(>|z|)
Study 2: Supporters of Raising Minimum Wage (n=818)			
Intercept	1.556 (.389)	4.006	.000
Concealed Carry Congeniality	.611 (.656)	.932	.352
Concealed Carry Incentives	.777 (.719)	1.081	.280
CC Congeniality × CC Incentives	.717 (1.350)	.532	.595
Minimum Wage Congeniality	.927 (.716)	1.295	.195
Minimum Wage Incentives	1.045 (.830)	1.259	.208
MW Congeniality × MW Incentives	-.667 (1.256)	-.531	.595
CC Congeniality × MW Congeniality	-1.767 (.965)	-1.830	.067
CC Incentives × MW Congeniality	-1.443 (1.035)	-1.394	.163
CC Congeniality × CC Incentives × MW Congeniality	-.157 (1.627)	-.096	.923
CC Congeniality × MW Incentives	.450 (1.412)	.319	.750
CC Incentives × MW Incentives	-1.534 (1.117)	-1.374	.170
CC Congeniality × CC Incentives × MW Incentives	-1.545 (1.964)	-.786	.432
CC Congeniality × MW Congeniality × MW Incentives	-.735 (1.789)	-.411	.681
CC Incentives × MW Congeniality × MW Incentives	1.209 (1.618)	.747	.455
CC Congen. × CC Incent. × MW Congen. × MW Incent.	2.856 (2.458)	1.162	.245
Study 3: Supporters of Raising Minimum Wage (n=918)			
Intercept	1.655 (.303)	5.466	.000
Concealed Carry Incentives	-.060 (.421)	-.141	.888
Minimum Wage Congeniality	-.961 (.376)	-2.56	.011
Minimum Wage Incentives	-.032 (.414)	-.077	.939
MW Congeniality × MW Incentives	.452 (.531)	.852	.394
CC Incentives × MW Congeniality	.155 (.535)	.290	.772
CC Incentives × MW Incentives	-.032 (.577)	-.055	.956
CC Incentives × MW Congeniality × MW Incentives	-.348 (.750)	-.464	.643
Study 3: Opponents of Raising Minimum Wage (n=262)			
Intercept	.211 (.326)	.648	.517
Concealed Carry Incentives	.482 (.462)	1.043	.297
Minimum Wage Congeniality	2.519 (.679)	.679	.0002
Minimum Wage Incentives	.792 (.480)	1.649	.099
MW Congeniality × MW Incentives	-1.443 (.900)	-1.602	.109
CC Incentives × MW Congeniality	-1.015 (.920)	-1.103	.270
CC Incentives × MW Incentives	-.846 (.647)	-1.308	.191
CC Incentives × MW Congeniality × MW Incentives	.889 (1.182)	.752	.452

Note: CC and MW indicate concealed carry and minimum wage experiments, respectively.

2 Additional Empirical Results

2.1 Does Numeracy Condition Treatment Effects?

We confined Study 1 to high-numeracy respondents based on the theoretical argument and main result of Kahan et al. (2017) that polarization only occurs among this group. However, in Studies 2 and 3, we invited all respondents to participate fully, regardless of numeracy, to compare the behavior and background characteristics of high- and low-numeracy respondents. In theory, respondents may differ by numeracy in multiple ways. First, high-numeracy respondents are probably more educated, which we see in Study 3 (see Table SI 1). Second, they may hold different attitudes. For example, high-numeracy respondents may exhibit greater political sophistication, ideological constraint, party identification, or other traits that predispose them to motivated learning, motivated responding, or both.

Therefore, in Table SI 4, we compare the behavior of low- and high-numeracy respondents in the concealed carry experiment. Note that our numeracy scale is truncated, so our ability to fully answer this question is limited. We find that the substantive pattern of results is similar among both groups, though the magnitudes of treatment effects change. Each cell in Table SI 4 displays the incentive treatment effect in percentage points, which is percent correct in the `Incentive` condition minus the percent correct in the `No Incentive` condition (standard errors in parentheses). Each row contains a type of respondent, and each column contains a different study and congeniality condition. Among concealed carry supporters, neither high- nor low-numeracy respondents exhibit significant treatment effects, in either the uncongenial or congenial condition. (The effects are positive among low-numeracy respondents in Study 2, but have large standard errors.) Among concealed carry opponents, we see the familiar pattern of positive treatment effects in the uncongenial, but not the congenial, condition. These effects occur among high-numeracy respondents in Study 2, and both high- and low-numeracy respondents in Study 3 (though they are only statistically significant among the latter group).

Table SI 4: Incentive Effects in Concealed Carry Task by Study, Condition, and Respondent Numeracy

	Study 1		Study 2		Study 3
Respondents	**Uncongenial**	**Congenial**	**Uncongenial**	**Congenial**	**Uncongenial**
CC Supporters					
High-Numeracy	-1.5 (6.3)	-3.3 (6.9)	-4.2 (7.3)	2.8 (7.1)	2.9 (5.8)
	n=248	n=207	n=180	n=200	n=261
Low-Numeracy	-	-	13.3 (11.3)	13.3 (11.0)	-2.4 (4.9)
			n=72	n=78	n=378
Combined	-	-	.6 (6.2)	5.9 (6.0)	-0.3 (3.7)
			n=252	n=278	n=639
CC Opponents					
High-Numeracy	9.1 (9.1)	-8.0 (9.5)	25.4 (9.4)	3.0 (9.4)	4.5 (7.6)
	n=123	n=108	n=108	n=116	n=159
Low-Numeracy	-	-	-10.1 (16.0)	-11.7 (17.9)	8.0 (5.4)
			n=38	n=32	n=257
Combined	-	-	17.1 (8.1)	-0.3 (8.2)	6.9 (4.4)
			n=146	n=148	n=416

Note: 'CC Supporters' and 'CC Opponents' refer to concealed carry supporters and opponents, respectively.

Since numeracy is not a prerequisite to understand which result is more congenial, in response to a concern raised by a reviewer, we also checked whether low-numeracy respondents tended to pick the congenial answer more often than the uncongenial one. We term this behavior "response bias" and operationalize it as the odds of choosing the congenial answer. Low-numeracy respondents indeed exhibit such bias. In Study 2, response bias is 1.25 without incentives, meaning respondents are 25% more likely to pick the congenial answer. High-numeracy response bias is a very similar 1.22. Where the two groups differ is in their responsiveness to incentives. Low-numeracy respondents continue to exhibit a response bias of 1.13 with incentives, while bias among high-numeracy respondents reverses itself to .82, indicating that as expected, they become more likely to pick the *uncongenial* answer.

In Study 3, incentives reduce response bias among both low- and high-numeracy respondents alike, paralleling the results in the last column of Table SI 4. Incentives reduce low-numeracy response bias from 3.68 to 2.41 and high-numeracy response bias from 2.04 to 1.68. Interestingly, it is low-numeracy respondents who exhibit more response bias overall in Study 3.

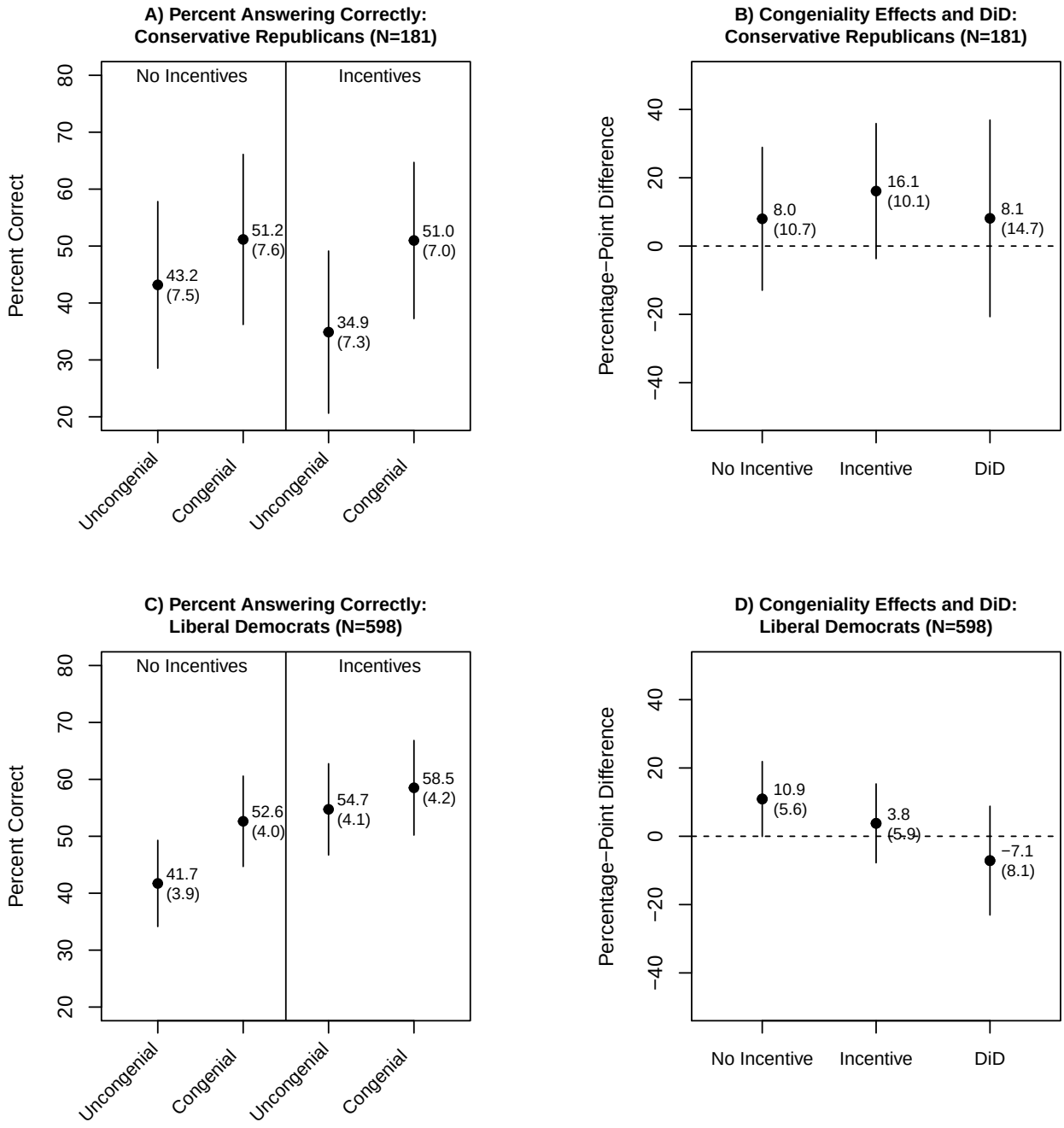
2.2 Subsetting Respondents by 'Ideological Worldview'

Kahan et al. (2017) operationalize congeniality using respondents' 'ideological worldview' instead of their position on concealed carry. The authors conceive of 'ideological worldview' as a combination of partisanship and ideology, and measure it by simply multiplying the two. At one end of the scale are "Conservative Republicans," and at the other end, "Liberal Democrats." And while we are skeptical, it is possible that ideological worldview establishes what information is congenial to a respondent, more so than relevant attitudes. To explore this concern, we construct a similar variable, categorizing respondents as either conservative Republicans or liberal Democrats, depending on their self reported party identification (including leaners) and three-point ideology (conservative, moderate, or liberal). Using this variable, we re-analyze the concealed carry task in Studies 1 and 2.

Before we present the results, a caveat: positions on concealed carry are only weakly related to ideological worldview, especially among low-numeracy respondents. In Study 2, for example, the correlation is .41 among high-numeracy respondents and only .16 among low-numeracy respondents. In fact, this weak relationship may be why Kahan et al. do not observe bias among low-numeracy respondents. We prefer to subset respondents by their issue position, precisely because the overlap between issue positions and ideological worldview is not 100%.

With the caveats above, we rerun our main analyses using this new variable to subset respondents, presenting the results in Figure SI 4. The substantive pattern of results is very similar to our main results in Figure 2. Conservative Republicans exhibit congeniality effects with and without incentives. The magnitudes are substantial, but the effects are marginally significant due to the small size of this group in Studies 1 and 2. Liberal Democrats, on the other hand, exhibit a significant congeniality effect of 10.9 points without incentives, which is reduced to an insignificant 3.8 points with incentives. While we view this conditioning variable as indirectly related to our outcome of interest, it is reassuring that we are able to replicate the main result of Kahan et al.. And again, incentives only work on respondents on the political left.

Figure SI 4: Concealed Carry Task Results using Kahan et al. (2017) Conditioning Variable (Studies 1 and 2). Panels on the left display percentage of conservative Republicans (Panel A) and liberal Democrats (Panel C) of raising the minimum wage correctly indicating study result by experimental condition. Panels on the right display congeniality effect by incentive condition, as well as difference-in-differences (DiD), among conservative Republicans (Panel B) and liberal Democrats (Panel D). Vertical lines indicate 95 percent confidence intervals.



2.3 No Evidence of Selective Perception

One proposed mechanism for why people learn in a motivated manner is that they simply misperceive the data. Thus, we next examine the extent to which errors in cognition are explained by congeniality of the data. Given our hypothesis that motivated responding explains a large portion of what appears to be motivated learning, we also conjectured that most respondents do not perceive data in a selective manner.

To test this hypothesis, we examined how well respondents did across various conditions in recalling the numbers in the contingency table (in Study 1 only). Prior studies using similar tasks suggest that errors are most common in the top-left cell in the contingency table (Gilovich 1991; Dawson et al. 2002b). We therefore first examine whether the congeniality manipulation changed estimates of the number of cities in this cell. Since the incentive treatment occurred only after respondents saw the contingency table, we do not expect incentives to affect recall. We therefore focus on the effect of congeniality on data recall in the absence of incentives.

There were 223 cities in the upper-left cell in both conditions (see Tables 1 and 2). If respondents misperceive (or misremember) this value in a way that makes the study's result more congenial to them, then we would expect them to overestimate this number in the uncongenial condition, relative to the congenial condition.[15] However, we find that mean estimates do not differ appreciably by congeniality (without incentives). The mean number of cities recalled was 213.9 (s.e. = 5.5) in the uncongenial condition and 212.5 (s.e. = 4.9) in the congenial condition.

We next checked whether estimates of the 107 cities in the lower-left cell varied by condition. Here we would expect estimates to be *lower* in the uncongenial condition than in the congenial condition. However, recall was again fairly consistent across conditions: 103.1 (s.e. = 2.7) and 105.2 (s.e. = 4.7) in the uncongenial and congenial conditions, respectively. Estimates

[15]Overestimating the upper-left cell inflates the crime increase:decrease (decrease:increase) ratio for cities that banned concealed handguns in the anti-gun (pro-gun) version of the table.

of the other two cells in the table were also unaffected by the congeniality manipulation. In summary, we do not find a congeniality effect on the recall task in the absence of incentives.

Finally, we simultaneously test the effect of congeniality, incentives, and their interaction on recall. We examine recall estimates of the number of cities in the upper-left and lower-left cells of the contingency table. As Tables SI 5 and SI 6 below show, we find no significant effects for either cell. We conclude that respondents do not perceive the data selectively. This analysis also serves as a placebo test, indicating that incentives affect reporting without affecting recall.

Table SI 5: Linear Regression: Upper-Left Cell Recall Estimates

	Coefficient (SE)	t value	Pr(>|t|)
Intercept	213.92 (5.58)	38.31	.00
Congeniality	-1.39 (8.08)	-.17	.86
Incentive	-1.92 (7.97)	-.24	.81
Congeniality × Incentives	2.66 (11.79)	.23	.82

Table SI 6: Linear Regression: Lower-Left Cell Recall Estimates

	Coefficient (SE)	t value	Pr(>|t|)
Intercept	103.09 (3.40)	30.34	.00
Congeniality	2.10 (4.91)	.43	.67
Incentive	1.08 (4.85)	.22	.82
Congeniality × Incentives	1.92 (7.17)	.27	.79

2.4 Covariance Detection Task Results by Study

We present the concealed carry results separately for Studies 1 and 2 in Figures SI 5 and SI 6, respectively. Our findings are substantively similar across studies. In both studies, concealed carry supporters exhibit sizable congeniality effects, with and without incentives. And in both studies, concealed carry opponents exhibit a congeniality effect in the control condition that is wiped out by incentives. Due to the reduced power of splitting up studies, the difference-in-differences is marginally significant but substantively large: 17.1 percentage points in Study 1 and 22.4 percentage points in Study 2. Moreover, incentives increase correctness only in the uncongenial condition, which is consistent with our hypothesis, in both studies.

Figure SI 5: Concealed Carry Task Results from Study 1. Panels on the left display percentage of concealed carry supporters (Panel A) and opponents (Panel C) correctly indicating study result by experimental condition. Panels on the right display congeniality effect by incentive condition, as well as difference-in-differences (DiD), among concealed carry supporters (Panel B) and opponents (Panel D). Vertical lines indicate 95 percent confidence intervals.

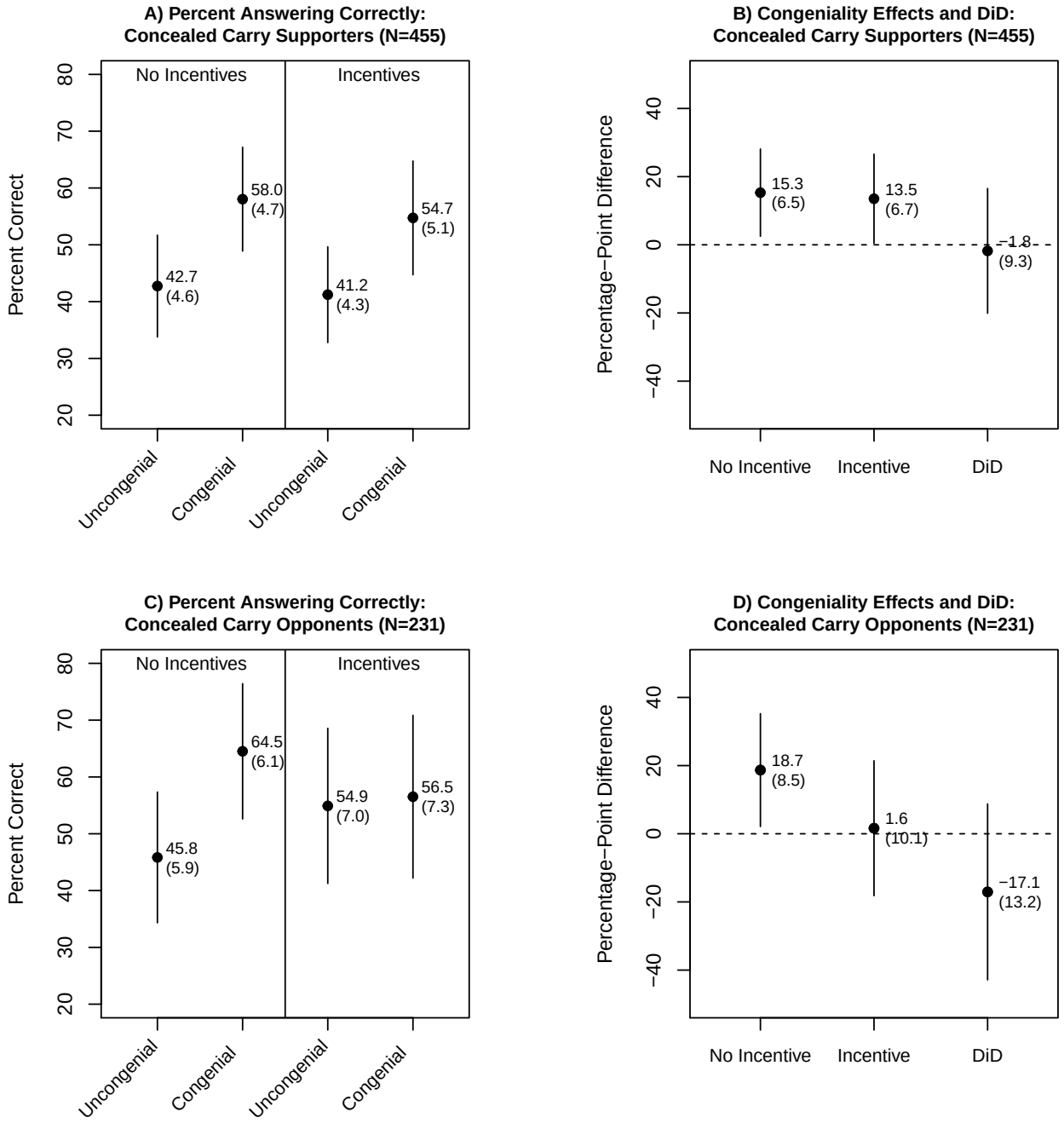
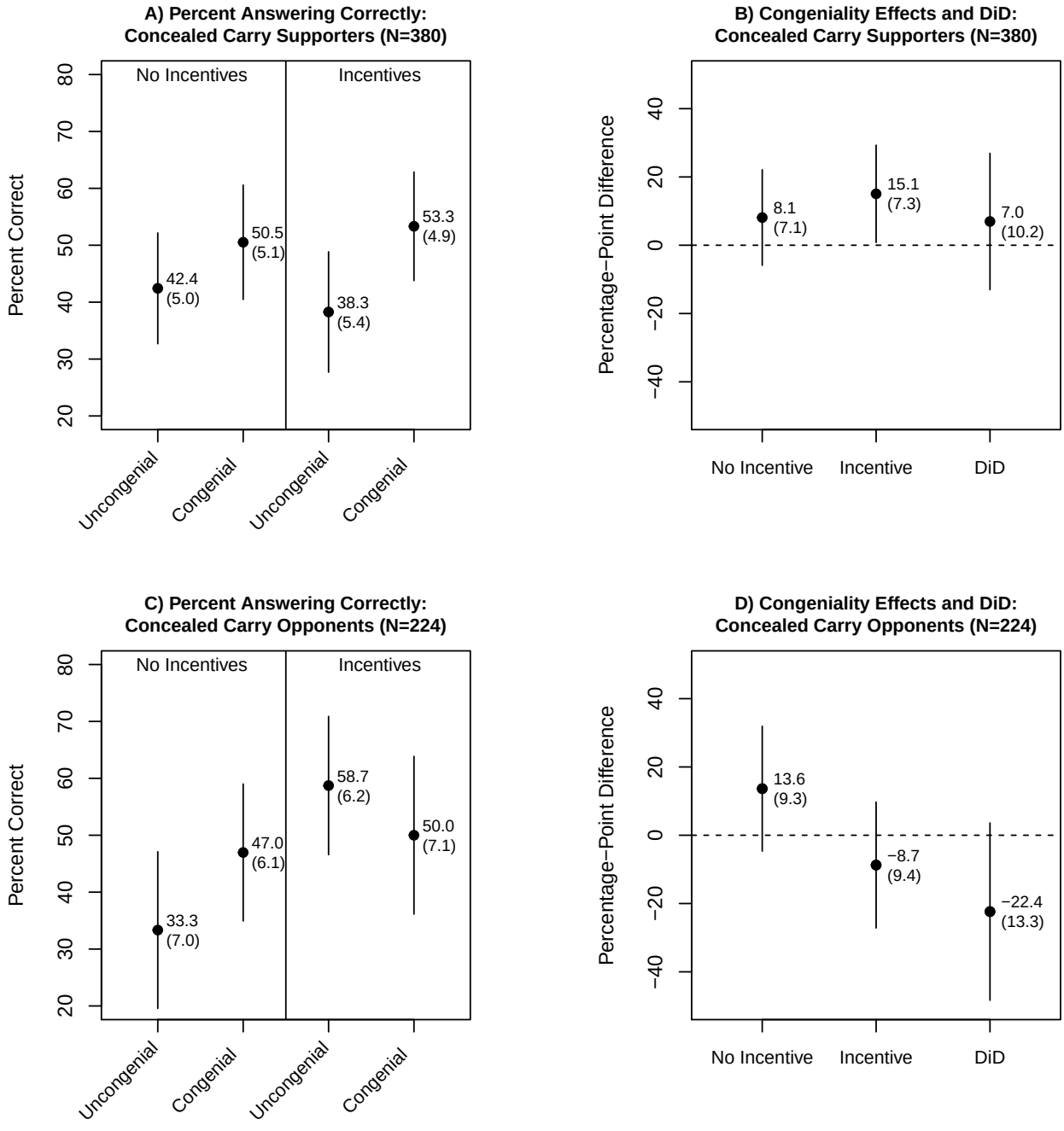


Figure SI 6: Concealed Carry Task Results from Study 2. Panels on the left display percentage of concealed carry supporters (Panel A) and opponents (Panel C) correctly indicating study result by experimental condition. Panels on the right display congeniality effect by incentive condition, as well as difference-in-differences (DiD), among concealed carry supporters (Panel B) and opponents (Panel D). Vertical lines indicate 95 percent confidence intervals.



We present the results of the minimum wage task separately for Studies 2 and 3 in Figures SI 7 and SI 8, respectively. In both studies, supporters of raising the minimum wage exhibit a negative congeniality effect in the absence of incentives, indicating they are more likely to answer correctly in the *uncongenial* condition than in the congenial condition. This pattern is inconsistent with motivated learning. Among opponents of raising the minimum wage, whom we can only analyze in Study 3, there is a 31-point congeniality effect without incentives. The incentive treatment reduces this effect by 14 percentage points (or 46%). This large decrease is consistent with motivated responding.

Figure SI 7: Minimum Wage Task Results (Study 2). Panel A displays percentage of minimum wage raise supporters correctly indicating study result by experimental condition. Panel B displays congeniality effect by incentive condition, as well as difference-in-differences (DiD), among minimum wage raise supporters. Vertical lines indicate 95 percent confidence intervals.

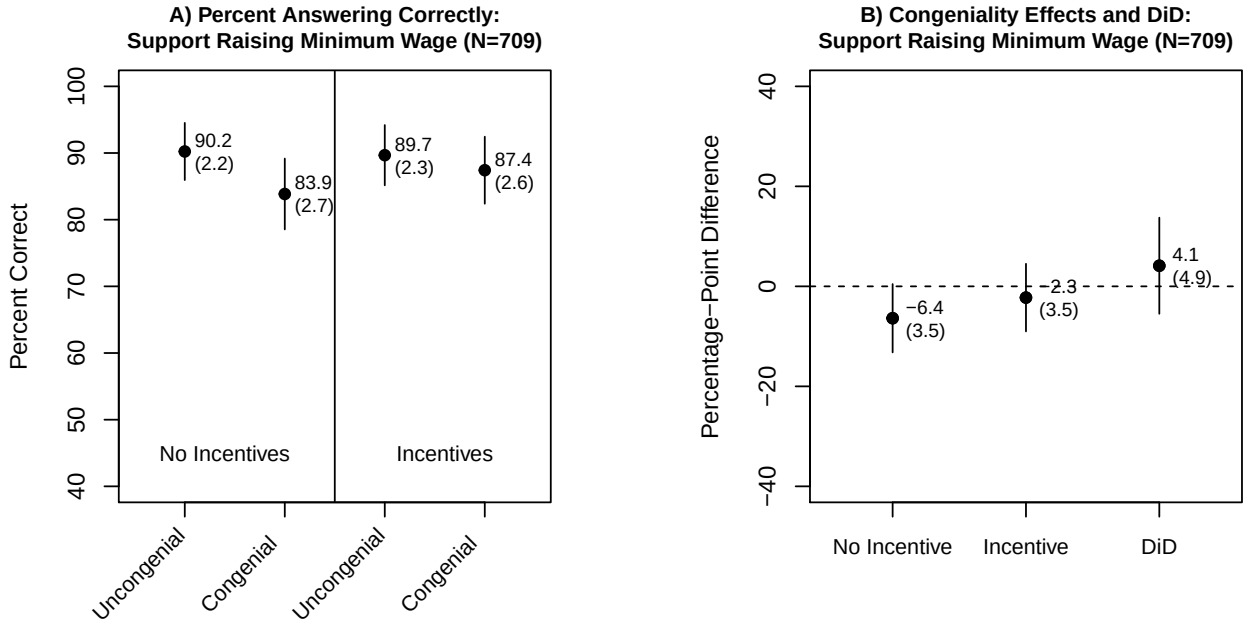
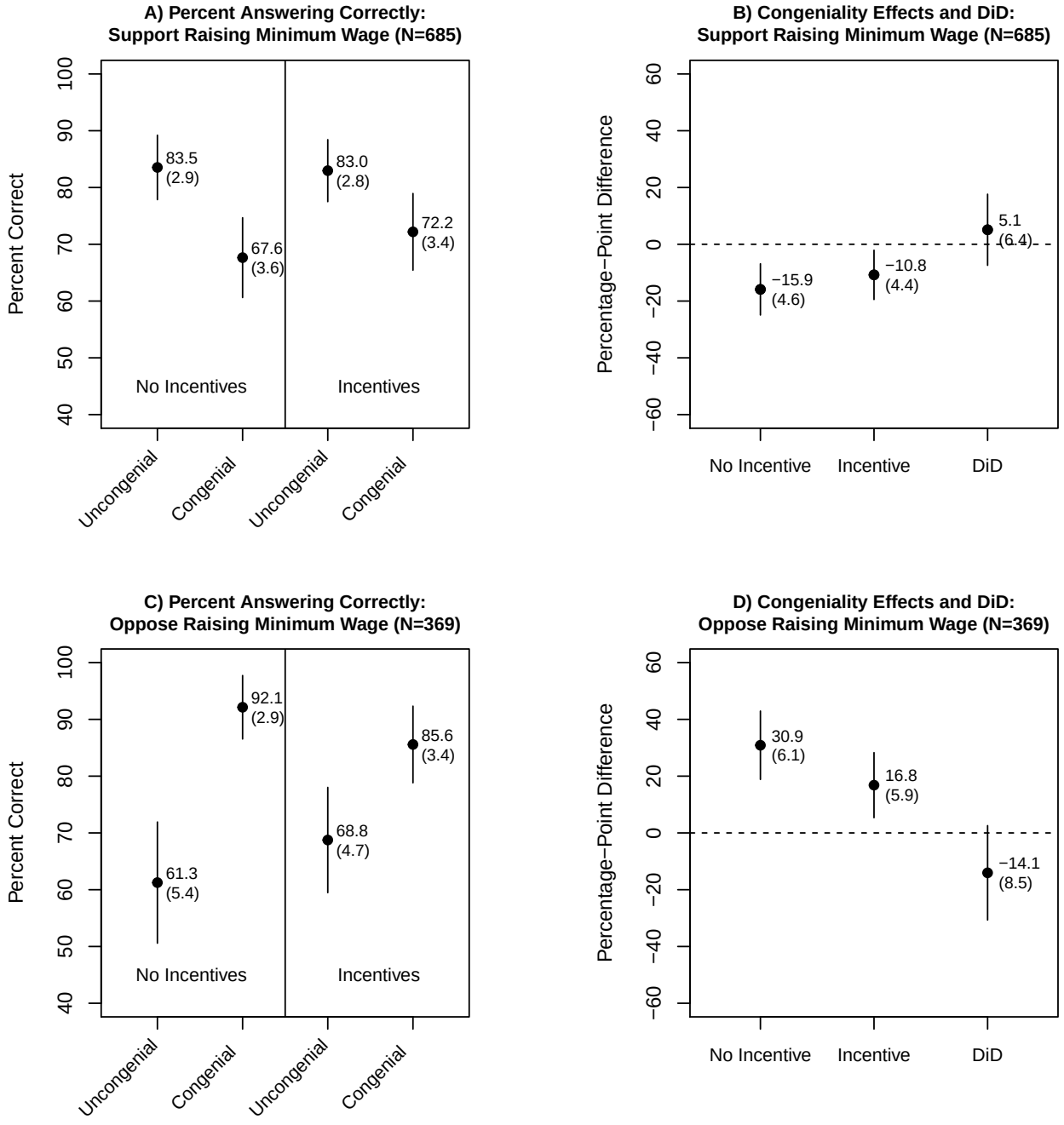


Figure SI 8: Minimum Wage Task Results (Study 3). Panels on the left display percentage of supporters (Panel A) and opponents (Panel C) of raising the minimum wage correctly indicating study result by experimental condition. Panels on the right display congeniality effect by incentive condition, as well as difference-in-differences (DiD), among supporters (Panel B) and opponents (Panel D). Vertical lines indicate 95 percent confidence intervals.



2.5 Concealed Carry Task Results as Logistic Regressions

For another succinct way to look at the results, we estimated the effects of the treatments using a logistic regression. The dependent variable is a dummy variable indicating whether or not the respondent correctly identified the result of the concealed carry study. Pooling our data from Studies 1 and 2, we run separate logistic regressions for concealed carry supporters (Table SI 7) and opponents (Table SI 8). Among concealed carry supporters, there is a significantly positive congeniality effect (odds ratio = 1.62), while the incentive effect is null. Among concealed carry opponents, both the congeniality and incentive treatments are significantly positive. And the interaction between the treatments is large and negative, indicating that the incentives significantly reduce the congeniality effect.

Table SI 7: Logistic Regression on Concealed Carry Supporters (Studies 1 & 2)

	Coefficient (SE)	Odds Ratio	z value	Pr(>|z|)
Intercept	-.298 (.138)	.742	-2.169	.030
Congeniality Treatment	.483 (.196)	1.620	2.462	.014
Incentives Treatment	-.103 (.196)	.902	-.525	.600
Congeniality × Incentives	.079 (.280)	1.083	.284	.777

Table SI 8: Logistic Regression on Concealed Carry Opponents (Studies 1 & 2)

	Coefficient (SE)	Odds Ratio	z value	Pr(>|z|)
Intercept	-.363 (.188)	.696	-1.931	.054
Congeniality Treatment	.583 (.259)	1.791	2.251	.024
Incentives Treatment	.646 (.267)	1.907	2.420	.016
Congeniality × Incentives	-.740 (.380)	.477	-1.946	.052

2.6 Assessing Experimenter Demand Effects

To earn a little extra money, do respondents report what they think the experimenter believes is true instead of what they believe to be true? To answer that, we need to conjecture about respondents' beliefs about the experimenter conducting the study. We consider two hypotheses about what respondents may think in this regard. The first is that respondents believe that most social scientists are liberal, and therefore infer that the result favoring gun control will always be treated as correct. Another is that respondents interpret financial incentives as a cue that the uncongenial option is the right one. We investigate both possibilities.

If respondents believe the experimenter is liberal, incentives should encourage respondents to select the anti-gun (i.e., pro-ban) answer, irrespective of the study's congeniality. To investigate this, we pooled opponents of concealed carry from Studies 1 and 2 as they had responded appreciably to the incentive treatment (see Figure 2). In Table SI 9 below, we display the percentage of these respondents who give the congenial (i.e., anti-gun/pro-ban) and uncongenial response (i.e., pro-gun/anti-ban) by condition. Without incentives, respondents are more likely to give the congenial response than the uncongenial response (59% vs. 41%). As in SI Section 2.1, we quantify this "response bias" by computing the odds of selecting the anti-gun response. Response bias is 1.44 in the uncongenial condition and 1.25 in the congenial condition. When respondents are offered incentives, response bias reverses itself in the uncongenial condition. It falls to .75, indicating that respondents are now more likely to select the pro-gun answer. Thus, incentives do not cause respondents to blindly choose the anti-gun answer, or the pro-gun answer, for that matter. Instead, they become less likely to pick the congenial answer and more likely to pick the correct answer.

Among concealed carry supporters, incentives have no discernible effect on response bias. These respondents are always more likely to pick the congenial (i.e., pro-gun) answer than the uncongenial (i.e., anti-gun) answer. Bias is a bit greater in the uncongenial condition than in the congenial condition, resulting in poorer performance in the uncongenial condition. For example,

in the absence of incentives, response bias is 1.35 in the uncongenial condition and 1.20 in the congenial condition. More importantly, incentives hardly change response bias in either the uncongenial or congenial condition. Thus, contrary to the scenario outlined above, incentives do not encourage supporters of concealed carry to select any particular response.

Table SI 9: "Response Bias" among Concealed Carry Opponents (Studies 1 and 2)

Experimental Condition	N	% Selecting Anti-Gun Answer	% Selecting Pro-Gun Answer	Odds of Anti-Gun Answer
No Incentives, Uncongenial	117	59.0%	41.0%	1.44
No Incentives, Congenial	128	55.5%	44.5%	1.25
Incentives, Uncongenial	114	43.0%	57.0%	.75
Incentives, Congenial	96	53.1%	46.9%	1.13

Table SI 10: "Response Bias" among Concealed Carry Supporters (Studies 1 and 2)

Experimental Condition	N	% Selecting Anti-Gun Answer	% Selecting Pro-Gun Answer	Odds of Pro-Gun Answer
No Incentives, Uncongenial	216	42.6%	57.4%	1.35
No Incentives, Congenial	207	45.4%	54.6%	1.20
Incentives, Uncongenial	212	40.1%	59.9%	1.49
Incentives, Congenial	200	46.0%	54.0%	1.17

Do respondents interpret the incentive as a cue that the uncongenial answer is correct, no matter what version of the table they see? It doesn't appear so. Table SI 9 shows that incentives only change behavior in the uncongenial condition. In the congenial condition, respondents are no more likely to pick the uncongenial (incorrect) answer with incentives than without them. And Table SI 10 shows that concealed carry supporters always exhibit a response bias in favor of the congenial answer, and incentives do not significantly reduce this bias. All of the considerations above contradict the idea that incentives introduce demand effects whereby respondents falsify their factual beliefs to earn extra money. Instead, the most plausible interpretation of our results is that incentives reduce the prevalence of motivated responding, thereby increasing correctness in the uncongenial condition.

2.7 How Confident are Respondents in Their Answers?

In Study 3, after the respondents had marked their answer about the conclusion supported by the study, we asked the respondents "how confident are you of your answer to this question." The respondents could pick between the following options: Not confident at all, Not very confident, Moderately confident, Very confident, and Certain. Table SI 11 below displays the percentage of incorrect and correct answers by respondents' subjective confidence in their answer. Few people are certain of their answers, particularly in the concealed carry task. Only 13% of respondents report that they are certain of their answer, and the vast majority of these certain answers are incorrect. Still, only 10% of respondents overall report being certain of an incorrect answer, mitigating the concern that motivated learning leads to being misinformed. Probably because the minimum wage task was relatively much easier, average certainty about the answer is substantially higher in the minimum wage task. One in five answers are certain and correct, and only 4% are certain and incorrect.

Table SI 11: Percentage of Incorrect and Correct Answers by Confidence (Study 3)

	Concealed Carry Task		Minimum Wage Task	
Confidence Rating	Incorrect	Correct	Incorrect	Correct
Not at all confident	2%	1%	1%	1%
Not very confident	8%	5%	2%	6%
Moderately confident	27%	3%	11%	24%
Very confident	22%	9%	6%	26%
Certain	10%	3%	4%	20%

Note: Cells display overall percentages, so cells for a task add up to 100%.