Information Exposure and Belief Manipulation in Survey Experiments*

Zikai Li†
Robert Gulotty‡

Working paper. Click here to download the latest version.

Abstract

Researchers use survey experiments to study the effect of informational beliefs on individual attitudes and behaviors. When the manipulation of interest is a belief about some fact, such studies must assume respondents receive the information and, further, that their beliefs change in the intended manner. These assumptions are often not addressed in survey experimental designs. We suggest that researchers collect and analyze post-treatment measures of both treatment compliance and belief change. With these measures, researchers can model the reception and causal effect of information and perform placebo tests of their proposed theoretical mechanisms. We demonstrate the utility of our framework by re-analyzing three prominent survey experiments in political science and with an original study on how changes in a factual belief can affect downstream judgments. We find that accounting for treatment compliance and belief change can better connect experimental studies and the substantive theories of politics they seek to test.

Keywords: factual information, survey experiments, beliefs, manipulation checks, noncompliance.

---

*We are grateful to Ryan Brutger, James Druckman, Andy Eggers, Narrelle Gilchrist, Anton Strezhnev, and Arthur Yu for their feedback on earlier drafts of the paper and to Zhijie Huang for research assistance. The original survey for this paper was approved by the Social and Behavioral Sciences Institutional Review Board at the University of Chicago (Protocol Number: IRB22-1240).

†PhD Student, Department of Political Science, University of Chicago; zkl@uchicago.edu

‡Associate Professor, Department of Political Science, University of Chicago; gulotty@uchicago.edu

1 Introduction

Beliefs play a central role in connecting material conditions to political attitudes and behavior, particularly those beliefs that pertain to factual information about the world and the society in which people live. Studies of the causal effects of such information use survey experiments to randomly assign respondents to facts embedded in a questionnaire. By controlling exposure to information, such studies hope to manipulate factual beliefs and test whether those manipulations affect downstream attitudes or behavior. However, these experiments produce credible estimates of the effects of exposure to information only if participants consume the information as the researchers intend; that is, they comply with the experimental protocol. Further, when studying the effect of information, researchers assume that exposure to the questionnaire moves the participants' factual beliefs. They assume, often implicitly, not only attentiveness but also credulity—that participants take all facts embedded in the questionnaire as evidence and update their beliefs accordingly.

In reality, not all information is taken as presented and respondents can fail to retain or internalize claims they read in a survey.[1] Respondents in survey experiments are often not only inattentive but also incredulous or indifferent. Some respondents may doubt the veracity of factual claims in surveys, perhaps because of the general erosion of trust in authority figures or because of experience with deception on survey platforms (Kennedy, Tyson, and Funk 2022; Boynton, Portnoy, and Johnson 2013). Moreover, even among attentive and credulous respondents, it is not obvious that exposure to a single piece of evidence should be sufficient to move beliefs.

In what follows, we begin by identifying the two assumptions briefly described above—that participants receive an information treatment as the researcher intends and that treated individuals move their beliefs in a certain direction and/or by a certain magnitude. These two assumptions are violated when a respondent fails to be

[1]According to data collected by the US Department of Education, 54% of American adults aged 16 to 74 lack literary proficiency (Rothwell 2020).

exposed to the information, or, upon being exposed, is unmoved by the information. We review informational survey experiments in papers recently published in top political science journals and find that they often do not measure the success of information exposure or belief change; for a handful of studies that do measure information exposure, the compliance rate is highly variable. We show how accounting for these two assumptions affects the interpretation of the experimental findings relative to the goals of the researcher in recovering the effects of information. In particular, if these assumptions do not hold, survey experiments still recover an intention-to-treat effect. Such a quantity is relevant, for example, if the researcher is evaluating the relative efficacy of messages or designing optimal policy. However, the intention-to-treat effect is only a lower bound for the effect of information treatment or changes in factual beliefs on downstream attitudes and behavior when the treatment effect is monotonic. When the effect is not monotonic, can mask important heterogeneity in the effect of information.

Survey experiments can be designed to more reliably recover the effects of information. Contemporary practice is to avoid bias by excluding information gathered after an experimental manipulation and to profile results by respondent attentiveness (Montgomery, Nyhan, and Torres 2018; Berinsky, Margolis, and Sances 2014). In this paper, we argue for incorporating measures of information recall and factual beliefs to account for the heterogeneity in exposure and belief change. With these measures, it is possible to extend the typical instrumental variable approach for non-compliance, where exposure to treatment is instrumented for by the random assignment of treatment, to the problem of changes in beliefs. In general, conditioning on post-treatment outcomes risks introducing bias that undermines the benefits of an experiment; however, under certain assumptions, post-treatment measures can be used to isolate those respondents that are exposed to the information and those for whom the information changes their beliefs about the exposed facts. The procedure can thus recover the more theoretically relevant effects of information exposure and belief change.

In our empirical demonstrations, we reanalyze three experiments using data from

Brutger et al. (2022a) with an IV approach to account for noncompliance in information reception. Here we use treatment-relevant manipulation checks, which are still relatively uncommon but have become increasingly popular in the field (Kane and Barabas 2019), as a way to measure exposure to treatment. In addition, we show how a placebo test can serve as evidence against the theoretical relevance of the intention-to-treat effect (ITT) (Eggers, Tuñón, and Dafoe 2023).

We then demonstrate the benefits of accounting for prior beliefs and measuring belief changes with a survey experiment in Taiwan. This application takes up the relationship between economic and political integration, namely, the effect of international trade on the diffusion of support for democracy (Mattingly et al. 2022; Magistretti and Tabellini 2022). In particular, we examine the effect that manipulating respondents' factual beliefs regarding interdependence with mainland China has on the prioritization between economic development and democracy.[2] Our approach focuses on those who are both attentive and moved by the information. Contrary to our expectation of a "backlash" effect, we find when the information produces an upward correction in the respondents' belief about the export dependence, respondents assign greater weight to economic development relative to democratic elections, a position consistent with rhetoric from the Chinese government.

2 How Do Your Participants Process Information?

Our focus is on experimental studies of how factual aspects of the external world, as understood by individuals, affect behavior and attitudes (Druckman 2022; Haaland, Roth, and Wohlfart 2023; Mutz 2021; Naoi 2020). For some cases, such as causal beliefs about economic relationships or the most effective response to climate change, facts are contested or complicated and the question is what would occur if there were

[2]In the paper, we use the terms China, mainland China and the People's Republic of China (PRC) interchangeably, and Taiwan and the Republic of China (ROC) interchangeably, to refer to the post-1949 PRC and ROC governments, respectively, and their de facto jurisdictions.

consensus.[3] In other cases, the facts may change, and the question is about how people would respond if they were made aware of it. In either case, factual beliefs form the link between material conditions and political behavior and attitudes.

It may seem that political scientists should leave the question of how individuals process factual information to cognitive scientists (Gul and Pesendorfer 2008). However, even those uninterested in neural computations or subjective psychology often theorize and make assumptions about whether and how individuals process information. In this section, we identify two sets of assumptions that are implicit in survey experiments that use information treatments. The first pertains to "basic compliance," i.e., that the respondents do pay attention to the information and understand the information in the manner intended by the researchers. The second pertains to the direction and magnitude of belief change: our studies implicitly assume a direction in which individual beliefs should move in light of exposure to the information treatment and the minimum magnitude of such movements.

The following sections characterize these two types of assumptions and argue that researchers can evaluate both empirically. For the first type, they can perform treatment-relevant manipulation checks to measure information reception (Kane and Barabas 2019). For the second, they can try to measure the target belief for manipulation, both before and after an information treatment. They can incorporate these new variables into an instrumental-variable design to gain additional insight into how belief changes map onto variations in downstream attitudes and behavior.

2.1 Information Provision and Reception

Our first goal is to study the effect of the *reception* of some experimentally manipulated information, but a prior question is to determine whether or not individuals are actually exposed. The reception of information is not equivalent to the provision of it, just as

[3]For example, recognizing most voters do not have a strong belief about the distributional effects of trade, Rho and Tomz (2017) use a survey experiment to offer information about these causal relationships to study the prevalence of egoistic policy preferences.

the prescription of a drug may not be equivalent to the actual use of it. When the theoretical treatment is the information itself rather than the provision of it, failing to read or understand the information is an example of non-compliance.

In the presence of non-compliance, the typical difference-in-means estimator recovers an intention-to-treat effect, or ITT (Angrist, Imbens, and Rubin 1996). This quantity may be of direct interest in a number of contexts across political science and policy analysis that seek an optimal informational intervention.[4] For example, in American and comparative politics, researchers studying ethnic discrimination toward candidates for office may design a vignette to study the effects of racially salient campaign materials on vote choices (Hainmueller and Hangartner 2013). It is unnecessary to distinguish cases of non-compliance from compliance to know what materials are more or less effective. Similarly, in studies of support for war or rebellion, the intention-to-treat effect may be of direct relevance.

In such studies, the researchers' goal can be to design treatments that closely resemble real life. For instance, making campaign fliers, or public messages to help adjudicate the efficacy of various informational programs. In such cases, researchers may benefit from using an adaptive experimental design that allows them to learn the most effective arm in the space of various information treatments (Offer-Westort, Coppock, and Green 2021; Villar, Bowden, and Wason 2015). However, when the goal is theoretical understanding of a treatment effect, effectiveness is not a direct goal. As Druckman (2022) puts it, "sound treatments do not depend on their mundane realism but rather on whether the relevant independent variable changes" (p. 54).

When the goal is not policy relevance, the issue of compliance has direct theoretical consequences. Consider the use of survey experiments on the topic of the nuclear taboo (Press, Sagan, and Valentino 2013). There, the survey presented news stories that are "designed to vary the relative military utility of nuclear weapons" (p. 196) to study the strength of the taboo. Reading these news stories is supposed to highlight the practical

[4]There is still the question of whether the specific intervention speaks to similar interventions of the same type.

utility of nuclear weapons, a consideration ruled out by the logic of the nuclear taboo. The story and surrounding information are merely a communication device to help the respondent learn about the information. If the vignette were too long or convoluted for respondents to work through, the lack of an effect of the military utility of nuclear weapons would not constitute support for a taboo. It could be that the respondent did not understand the information at all, and so would have changed their evaluations and positions on use, but did not get the chance.

In the end, ignoring non-compliance amounts to using the intention-to-treat effect as a proxy for the treatment effect of interest. If the problem is information non-reception, the use of the ITT as a proxy produces results with the same sign but potentially lower magnitude than the treatment effect of interest. In the following, we discuss one way to get closer to the treatment effect of theoretical interest.

2.2 Manipulation Checks: "No Harm in Checking," but What Is the Benefit?

One approach to addressing non-compliance in survey experiments is to include factual manipulation checks (Kane and Barabas 2019). In contrast to subjective manipulation checks, which ask the respondents what they think of the manipulation of interest, or instructional manipulation checks, which evaluate attentiveness more generally, treatment-relevant factual manipulation checks evaluate objective questions about the main elements of the experiment. This requires researchers to be explicit about the intended interpretation of the information treatments. For example, in replicating Press, Sagan, and Valentino (2013), Brutger et al. (2022a) determine whether respondents demonstrate basic recall of the part of the news story that varies between the treated and placebo groups. The participants' performance in this task can serve as a measure of compliance with the information treatment, i.e., of whether the information is received.

What distinguishes manipulation checks from other post-treatment outcomes is that the researchers only ask the participants what the text provided to them earlier in the survey said vis-à-vis some aspect of the world, not what the participants themselves know and/or believe about it. Kane and Barabas (2019) find that posing such questions does not affect outcomes, offering a low-cost diagnostic for the study, and suggest it is possible to use the outcomes of these checks to help interpret experimental findings. However, once one estimates a passing rate it is not clear how to incorporate the result into the study itself. In Section 3, we argue that, with a few more assumptions, an instrumental-variable approach can give us additional analytical leverage over the results of treatment-relevant manipulation checks. We demonstrate this in our re-analysis of two survey experiments in Section 4.

Past studies have used manipulation checks, but they are rarely incorporated into the analysis. Table 1 shows the shares of papers published in the *American Journal of Political Science*, the *American Journal of Political Science*, and the *Journal of Politics* between 2019-2023 that deploy survey experiments with information treatments. Section A.1 in the appendix provides more details on the search procedure. We categorize qualifying papers into those with manipulation checks and those without. For those with manipulation checks, we further check whether the manipulation check is "treatment-relevant" (Kane and Barabas 2019) in that it asks about the aspects of the informational treatments that are directly relevant to the authors' explanatory variable of interest. If it does, we code the study as having a treatment-relevant check. Table 1 shows the results of this review.

Table 1. The shares of APSR, JoP, and AJPS papers with no manipulation checks (MCs), any MCs, and treatment-relevant MCs. See Section A.1 in the appendix for a list of the papers reviewed.

Category	Count
No mention of MC	43
Treatment-relevant (TR) MC	9
Any MC (excl. TRMC)	15
Total	67

Of the 67 papers we reviewed, only 9 have a treatment-relevant manipulation check and 15 have a manipulation check that we classify as not directly related to the treatment of theoretical interest: five are subjective manipulation checks while the rest are often attention checks. The remaining 45 papers do not mention the use of a manipulation check in either the main paper or the appendix. Of the experiments in the five papers that have manipulation checks, the median is 66%. A 70% compliance rate, for example, can dilute the ITT by 30%.[5] Apart from efficiency loss, these figures mask wide variation, even within such a small sample: The lowest passing rate is 29% and the highest 93%. Unaddressed, this variation in compliance rates can confound comparisons of results from across experiments or studies.

2.3 Belief Change, or the Lack Thereof

Our second and primary goal is to study the downstream effects of experimentally induced changes in an individual's belief about a fact. Upon reception, whether an information treatment affects individual beliefs about a fact depends on a variety of factors, including a person's existing knowledge and beliefs.

Consider an analogy from the health sciences, specifically the study of vitamin D's impact on calcium absorption. This process can be influenced by exposure to ultraviolet (UV) light, with variation in UV light exposure affecting the body's storage of vitamin D, which in turn, affects calcium absorption. However, the impact of UV light exposure depends on patient characteristics; a patient with already high levels of vitamin D might not show significant changes. Conversely, a person deficient in vitamin D might show marked changes, making UV light exposure more effective. In a similar vein, new information can induce attitudinal and behavioral change only when an individual has a "deficiency" in that particular information before exposure. Information that is already known may be redundant.

[5]This assumes the non-compliers in the treatment group exhibit a similar outcome structure as those in the placebo group but, if compliant, would react to the treatment in the same manner as those who do receive the information. This also assumes the placebo has no effect on participants.

Moreover, just as the efficacy of UV light exposure depends on an individual's ability to synthesize vitamin D—for instance, people with melanated skin may synthesize vitamin D less efficiently under the same UV light exposure—the persuasiveness of new information also varies across individuals. This variation is dictated by their pre-existing belief structures. Some of these structures and beliefs may result in some individuals being less persuadable than others by a given piece of information.[6]

The consequence of redundancy, prior beliefs and incredulity is to introduce heterogeneity in the treatment effect of information. In the case of medicine, the standard approach would be to measure the levels of vitamin D prior to exposure to UV light. Factual beliefs can be more complicated. When an individual receives a piece of information that is credible but presents a characterization of the world that differs from their expectation, they should adjust. The magnitude of this movement, however, depends on a complex interaction among an individual's epistemological system, the relationship between the source of the information and this system, and their prior interactions with the external world that gave rise to their prior belief.

The issue of belief manipulation arises frequently in survey research. Recall the study of the nuclear taboo, where the main empirical question is whether Americans' attitudes toward the use of nuclear weapons "are driven by consequentialist considerations of military utility" (p. 188). If we do not measure whether and how the participants' perception of that utility changes, a lack of an effect could be driven by a variety of logics. It could be that individuals do update their beliefs about the military utility of nuclear weapons but their policy positions on use are driven by non-consequentialist logics like a taboo. Alternatively, it could be that the information in the vignette, while understood, was not enough to change evaluations of the military uses of nuclear weapons, even once one accounts for information reception. Some respondents may have internalized the military utility of nuclear weapons prior to the study, and so did not respond

[6]For more on motivated reasoning see Coppock (2022), Druckman and McGrath (2019), Little, Schnakenberg, and Turner (2022), and Taber and Lodge (2006). Our framework assumes that study participants do not update beliefs in the opposite direction intended by the researcher.

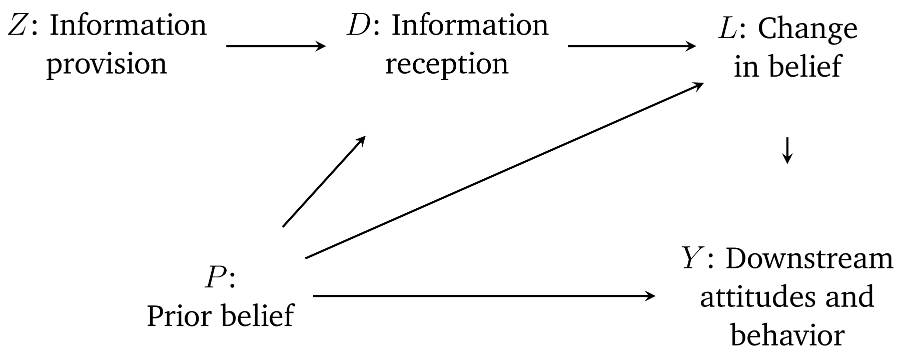

Figure 1. A directed acyclic graph (DAG) illustrating the pathways connecting information, beliefs, and attitudes and behavior.

to the redundant information. Others may be so convinced as to the military disutility of nuclear weapons that a single source of information contradicting their prior is insufficient to move their beliefs.

In general, researchers should not assume respondents align their beliefs with the information treatment even conditional on successful reception. First, in an increasingly contentious information environment, respondents may hold varied second-order beliefs about the credibility of the "messengers," whether researchers themselves or the sources they cite. Second, information that appears highly at odds with a respondent's existing beliefs may be discounted (Butler et al. 2017; Christensen 2023; Druckman and McGrath 2019).

Figure 1 schematizes the roles information reception and belief change play in a typical survey experiment with an information intervention. Our goal is to draw researchers' attention to the heterogeneity in information reception and belief change both within and across survey experiments. Failing to account for this heterogeneity can complicate our learning from experimental findings.

3 Accounting for Noncompliance and Heterogeneous Updating

An instrumental-variable approach can address both non-compliance regarding information reception and, under stronger assumptions, the issues related to heterogeneous updating. For information reception, researchers can measure compliance using treatment-relevant factual manipulation checks (TRMC; Kane and Barabas 2019). For heterogeneous updating, they can, whenever possible, elicit prior and posterior beliefs. In both cases an instrumental-variable design can bring empirical estimates closer to the theoretical quantity of interest. In such a design, the treatment assignment becomes the instrument with which we seek to induce changes in the information and beliefs individuals hold, assuming that information reception and belief change are the more theoretically relevant explanatory variables.

Under certain assumptions, the IV model recovers the effect of information provision on those who receive the information, the average effect on the treated (ATT). The assumptions are random assignment, stable unit treatment value, IV relevance, monotonicity, and the exclusion restriction (Imbens and Rubin 2015). Random assignment and stable unit treatment value are plausibly satisfied in the survey experiment context, and IV relevance is testable.

The IV approach also requires assuming there are no defiers, individuals who are more likely to take up treatment when assigned to control, or less likely to take up treatment when assigned to treatment. In the case of survey experiments, assignment to the information exposure arm in the survey should not decrease the actual reception of the information. The analogy to experiments with heterogeneous updating is straightforward. Following the logic laid out by Yu (2023), subjects can be divided into the persuaded, never persuaded, already persuaded and the dissuaded. Here too we must make the assumption that offering "correct" information does not paradoxically cause people to move their beliefs away from the information. There are two additional sets of assumptions that need to be met for an IV analysis to be informative, one about the exclusion restriction and the other about measurement errors in post-treatment checks. We discuss them below.

3.1 The Exclusion Restriction

In the IV setting, the exclusion restriction requires that the instrument affect the outcome only via the explanatory variable of interest Angrist, Imbens, and Rubin (1996). For the purposes of non-compliance in experiments, this requires assignment to the informational treatment to only affect exposure to the information. This can break down if, for example, there is a compound treatment for which the researcher cannot isolate the key varying information and measure its reception.

Under this assumption, studies that measure information reception can use that information to evaluate their hypothesized mechanisms. If the exclusion restriction assumption holds and the study includes a treatment-relevant manipulation check, there should be no effect of treatment among those that fail the manipulation check. That is, if we trust that failure to pass a manipulation check means that individuals fail to discern the information in the first place, then any observed effect would be evidence against the theorized mechanism.

In Brutger et al.'s (2022a) replication of Press, Sagan, and Valentino (2013), for example, the treatment vignettes list the expected probabilities of success of conventional and nuclear attacks, respectively, and the information that varies between the treatment and control groups is whether the probability of success for a conventional strike is the same as or substantially lower than a nuclear strike. The manipulation check for both the treatment and control groups then asks the respondents whether the vignette describes the probability of success for a conventional strike as similar to, higher than, or lower than that for a nuclear strike. If those assigned to treatment—which describes the probability of success for a nuclear strike as much higher than that for a conventional strike—fail this manipulation check, we should not expect this key information to have registered in the participants' minds. Because the vignettes for the treatment and the control are otherwise the same, the exclusion restriction should be satisfied unless we expect the manipulation of the relative probabilities to change the preferences of those assigned to treatment even when they fail to register the gap.

If the treatment does affect the outcome among those who fail the treatment-relevant manipulation check, we can draw two conclusions. First, the instrumental variable design is invalid. Second, it raises questions about the theoretical interpretation of the ITT. The researcher has manipulated some other consequential information beyond what she aims to manipulate (Dafoe, Zhang, and Caughey 2018). With a treatment-relevant manipulation check, we can conduct a population placebo test (Eggers, Tuñón, and Dafoe 2023) to assess the plausibility that the treatment affects the outcome primarily via the reception of the information we seek to convey. We demonstrate this in Section 4.2.

The exclusion restriction is satisfied under the theory of belief change represented in Figure 1. The exclusion restriction rules out effects from the delivery of treatment other than informational beliefs, including affective responses, or if the treatment also contains non-measured factual information. However, the framework does allows the treatment to affect related beliefs ("nuisance belief") related to the outcome. Such an effect would only constitute a violation of the exclusion restriction (Dafoe, Zhang, and Caughey 2018) if the treatment manipulates these beliefs simultaneously. Our framework thus differs from Acharya, Blackwell, and Sen (2018) in that we focus on the total effect of a manipulation of interest and the imperfection of such a manipulation while Acharya, Blackwell, and Sen address estimating the contribution of mediators to the causal process *provided* the first-stage manipulation is successful. Combining Acharya, Blackwell, and Sen's framework with ours, researchers can further decompose the effect of a belief manipulation into main effects and indirect ones that go through changes in nuisance beliefs. At a minimum, this requires that the information treatment be narrow in scope. The researcher's measure should be able to characterize respondents'

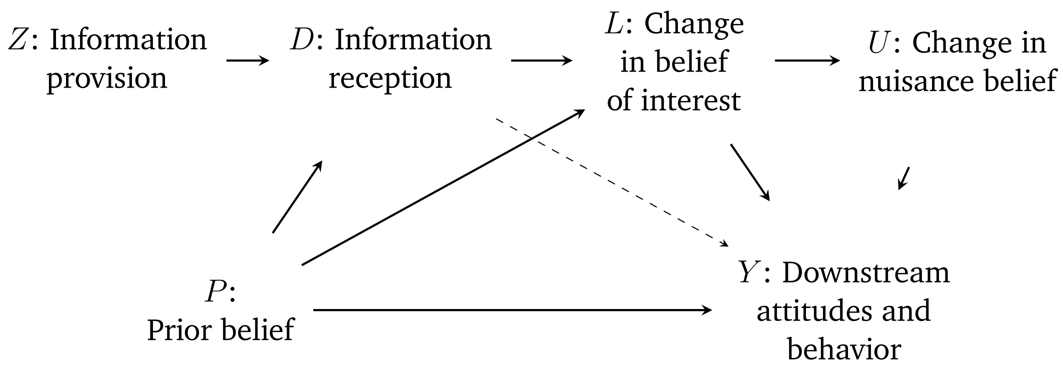

Figure 2. A DAG illustrating an informational experiment with second-order belief manipulation.

beliefs about this information fully.[7]

Consider an experiment on the public-opinion basis of democratic peace (Tomz and Weeks 2013; Dafoe, Zhang, and Caughey 2018). Suppose researchers assign participants to different vignettes about hypothetical countries. They manipulate only the regime types of these countries in the vignettes and seek to study how this affects the respondents' willingness to endorse the use of force against the countries. One concern about such a design is that, by learning that a country is a democracy, participants might also change their belief about the location of this country and this may affect their preferences on the use of force (Dafoe, Zhang, and Caughey 2018). This should be a concern, however, only when the beliefs about regime type and location change simultaneously in response to the information that a country is a democracy, which contradicts the concern that participants might change their belief about a country's location upon learning about its regime type. As Figure 2 shows, when a nuisance belief (belief about a country's location in this example), U, is affected by the belief of interest (belief about a country's regime type), L, but is downstream with respect to it, estimation using an instrumental-variable design is still valid. In the interpretation, however, researchers may need to point out the possibility of the indirect channel ($L \to U \to Y$)

[7]Our framework does not speak to experiments in which researchers simultaneously perturb many features of the informational text to study the effects of latent constructs of interest.

if this channel is not of theoretical interest.

In some situations, researchers may still be concerned that their information treatment affects a variable that (1) is not downstream to the belief of interest and (2) affects the outcome. In such situations, we would recommend that researchers pair an instrumental-variable design with a sensitivity analysis to test how sensitive their results are to various degrees of violation of the exclusion restriction (Cinelli and Hazlett 2019; Strezhnev, Kelley, and Simmons 2021). Strezhnev, Kelley, and Simmons's (2021) approach allows researchers to specify theoretical priors about the size of the effect of the instrument on the outcome that does not go through the belief of interest ($D \to Y$ in Figure 2) and check how much the substantive results would change were the priors true. Further, researchers may benchmark the values of the effect $D \to Y$ using the first-stage effect of the instrument on the belief of interest ($D \to L$).

In general, researchers interested in questions that match our framework in Figure 1 should not be discouraged from using an instrumental-variable approach with either information reception or belief change as the endogenous variable merely out of concern about violating the exclusion restriction. As we argued above, when the information subject to manipulation is narrow in scope, the exclusion restriction should be plausible. Even when it is not, with recent advances in sensitivity analysis, researchers may conduct theoretically motivated probes of how the estimate of the effect of interest might change when the exclusion restriction is violated. That is, learning is possible even when the possibility of bias looms (Little and Pepinsky 2021), and an instrumental-variable approach could allow researchers to gain from their data more learning vis-à-vis their theoretical priors.

When researchers believe the exclusion restriction is violated but are reluctant to use an instrumental-variable design with a sensitivity analysis, the only alternative may be to focus on the difference in means. But if the exclusion restriction is indeed violated, they cannot interpret the difference-in-means as anything beyond the ITT (Dafoe, Zhang, and Caughey 2018). Thus, when the exclusion restriction is violated and the researcher reports only the difference-in-means, it is unclear what we should learn from the estimate if the treatment of theoretical interest is information reception or belief change.

3.2 Measurement Error

Researchers may be concerned that the measures of exposure and belief change are noisy and would compromise the validity of an instrumental-variable analysis. Few designs in the social sciences can escape measurement errors. The question we face is whether the approach we propose here improves upon the status quo. We argue that, under reasonable assumptions about the measurement errors and the exclusion restriction (see Section 3.1), the instrumental-variable estimates are still better connected to the theoretical quantities of interest. We show this for the estimation of the treatment effect using a treatment-relevant manipulation check as the compliance measure.

Denote the IV estimator as $\hat{\theta}$. Under SUTVA, exclusion restriction, and ones-sided noncompliance and in the absence of measurement errors, it has the following property (Imbens and Rubin 2015):

$$\begin{aligned}
\mathbb{E}(\hat{\theta}) &= \frac{\mathbb{E}(Y_i(Z_i = 1)) - \mathbb{E}(Y_i(Z_i = 0))}{\mathbb{E}(D_i(Z_i = 1)))} && (1)\\
&= \mathbb{E}[Y_i(D_i = 1 \mid D_i(Z_i = 1) = 1) - Y_i(D_i = 0 \mid D_i(Z_i = 1) = 1)] && (2)\\
&= \text{Average treatment effect on the treated} = \theta && (3)
\end{aligned}$$

Proposition 1. *Under SUTVA and exclusion restriction, $\hat{\theta}$ is an unbiased estimator of θ if, in the measurement of treatment uptake among those assigned to treatment, those who fail to take up treatment are as likely to be mismeasured as those do take up treatment.*

Proof. See A.2 in the appendix. □

Thus, if the measurement errors are symmetric between those with $D_i = 1$ and those with $D_i = 0$, the IV estimator is unbiased. When the treated are more likely to select the wrong answer than the untreated are to happen to select the correct answer, the IV estimator is biased upwards. For this scenario to realize, those respondents who are attentive to and comprehend the information provision would need to be somehow substantially less attentive to the manipulation check. This is unlikely when manipulation checks are relatively simple and short compared to the information provision itself (the vignette), a condition that the replications we conduct in this paper satisfy.

If untreated respondents are more likely to happen to select the correct answer than treated participants are to select the wrong answer, the IV estimate is attenuated relative to the truth. But unless the difference in these two types of errors is larger than the true noncompliance rate, the IV estimator is still a better approximation of the target quantity than the ITT. If we consider the ITT as an estimator for the average treatment effect of information reception, then the assignment indicator is a proxy for information reception. But it is likely a noisy proxy if we hold moderate confidence in the results of typical manipulation checks. This means a sizable percentage of respondents are assigned to the treatment information but nevertheless do not receive it, and are thus mismeasured as treated when we use the assignment indicator as a proxy for treatment status. The attenuation this produces is smaller than the IV estimator only if the percentage of such mismeasured observations in the entire sample is smaller than the difference in the error rates of the manipulation check as a measurement of treatment status, which is unlikely.

Measurement errors pose an inferential threat to the approach we propose in this paper, as it does to many statistical analyses in the social sciences. The analysis above lays out the assumptions that need to be satisfied for the manipulation-check IV estimator to better approximate the theorized treatment effect than the ITT even in the presence of measurement errors. We believe these assumptions are plausible in carefully designed experiments with simple manipulation checks, but researchers nevertheless need to assess them using their substantive expertise.

4 Re-analysis of Three Experiments

4.1 Accounting for Noncompliance in Information Reception

In this section, we first reanalyze two studies, Press, Sagan, and Valentino (2013) and Nicholson (2012), using replication data from Brutger et al. (2022a). We choose to re-analyze the experiments in Brutger et al. (2022a) because they included treatment-relevant manipulation checks that are simple and straightforward. These two studies are also well-suited to our purposes because the information under manipulation is clearly defined and narrow in scope, making it easier to map empirical results to theoretical hypotheses. In their replication of Press, Sagan, and Valentino's (2013) experiment on support for using nuclear weapons, Brutger et al. manipulate the relative probabilities of success of using conventional and nuclear strikes against an adversary so that the success rate of a nuclear strike is higher than a conventional strike for the treatment group and the same as a conventional strike for the placebo group.

The Elite Messaging replication builds on Nicholson's (2012) investigation of support for an immigration policy based on endorsements from politicians within or outside of the party one leans toward. Brutger et al. introduced treatments that varied the partisan identity of the endorsing politician. Here we restrict our analysis to treatment arms where the politicians are fictional and the situations are hypothetical because using prominent real-world politicians manipulates not just the partisan identity of the endorser but also other attributes of them. The information under manipulation is thus the identity of the politician endorsing the message.

Table 2. Treatment recall rates by treatment status and study (Brutger et al. 2022a)

Study	Treatment Status	Correct Recall
Nuclear Weapons	0	0.56
	1	0.58
Elite Messaging	0	0.67
	1	0.56

In both replications, Brutger et al. added treatment-relevant manipulation checks

Table 3. A cross tab of information assignment and exposure

		Recall Status X_i	
		Correct Recall	Incorrect Recall
Assignment Z_i	Treatment	$\mathbb{E}(Y_i \mid Z_i = 1, X_i = 1)$	$\mathbb{E}(Y_i \mid Z_i = 1, X_i = 0)$
	Placebo	$\mathbb{E}(Y_i \mid Z_i = 0, X_i = 1)$	$\mathbb{E}(Y_i \mid Z_i = 0, X_i = 0)$

to test the respondents' recall of the aspect of the treatment or placebo vignettes that were relevant to the theories. This aspect should also be the only feature that varies between the vignettes. In the Nuclear Weapons study, this question asked whether the vignette that the respondents had seen said nuclear weapons had a higher, equal, or lower success rate. In the Elite Messaging replication, the question asked about the partisan identity of the actor in the vignettes that had been presented to the respondents.

As Table 2 shows, the passing rates for the manipulation checks, although similar across treatment and control groups, are not high. Researchers who do include treatment-relevant manipulation checks in survey experiments often only use the data to show descriptive statistics of the passing rates and check whether they differ with respect to the treatment arm or some other covariate of interest. Under particular assumptions, researchers could extract more information from results of manipulation checks using an instrumental-variable analysis.

The variable that represents whether respondents answer the manipulation checks correctly, upon some recoding, can serve as a measure of information reception in our framework (Figure 1). In Brutger et al.'s replications, the manipulation checks test whether the respondents can recall the information that researchers seek to manipulate in the vignettes. Table 3 categorizes the means in the outcome of interest by the respondents' assignment and recall status. Using the interaction between `treatment status` and `correct recall`, we can create a new variable that measures the receipt of the information treatment (Table 4). With this variable as the endogenous variable of interest and the provision of treatment as the instrument (Z), the IV estimator recovers the average treatment effect on the treated (Imbens and Rubin 2015).

Figure 3 shows the results. For the Nuclear Weapons study, the IV estimates are

Table 4. Information reception based on treatment status and correct recall

Assignment status (Z_i)	Correct recall (X_i)	Information reception (D_i)
0	0	0
0	1	0
1	0	0
1	1	1

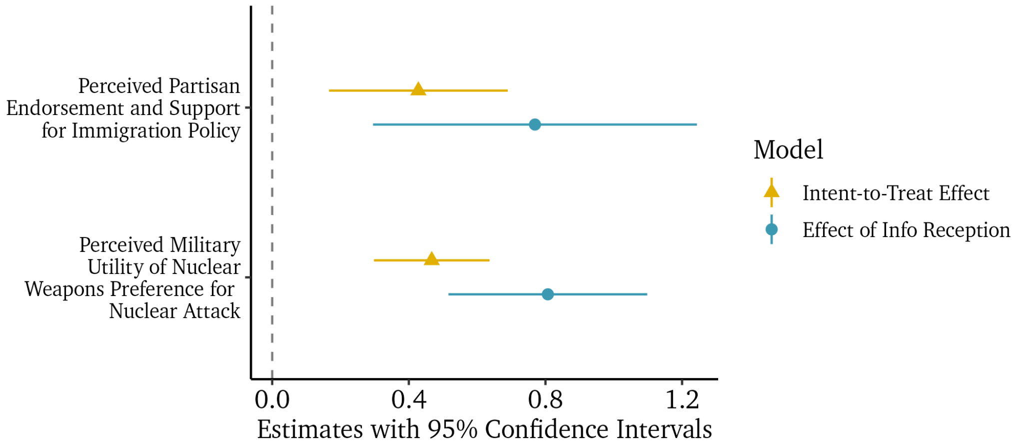

Figure 3. ITT versus effect of information reception for the Elite Messaging study (N = 278) and the Nuclear Weapons study (N= 524) in (Brutger et al. 2022a). Confidence intervals are calculated with HC2 standard errors. No control variables are included. Table A2 shows the full results in numerical format.

larger than the ITTs, indicating the information treatment had substantively larger effects on respondents who received the information. For the Elite Messaging study, the IV estimate is noisy—the number of respondents that correctly recalled the treatment was limited.[8]

4.2 Placebo Tests Using Results of Manipulation Checks

The validity of the IV approach rests on whether the exclusion restriction is satisfied, which cannot be tested. When it is violated, estimation of the effect of information reception is unwarranted, and so is any interpretation of the ITT in terms of the researcher's theoretical expectations. In this section, we show how we can conduct a simple population placebo test (Eggers, Tuñón, and Dafoe 2023), again using the results of

[8]We show the results of a placebo test, which we introduce in the next section, in Section A.3 in the appendix. The results suggest the exclusion restriction is plausible.

the manipulation checks, to probe the plausibility of the exclusion restriction. For the purpose of exposition, we use Brutger et al.'s replication of Mutz and Kim (2017), for which the placebo test indicates the existence of a channel other than the theoretical mechanism hypothesized by the authors. For the two studies to which we applied the IV approach above, the placebo test does not indicate such a violation (See Section A.3). This approach is useful not just for probing whether one should proceed with an IV estimation but also for assessing the substantive conclusions researchers draw from the ITT estimates.

The key idea here is that an important premise of an informational experiment is that the treatment only affects outcomes among those who change their factual beliefs. There should be no effect in the subsample that fails the treatment-relevant manipulation check. If, on the other hand, we do find such an effect, we should lower our confidence in the hypothesized mechanism.

Table 5. Treatment arms in the ingroup favoritism experiment

Condition 1	US gains 10 jobs, other country gains 1,000
Condition 2	US gains 10 jobs, other country loses 1,000
Condition 3	US gains 1,000 jobs, other country gains 10

In their replication of Kim and Mutz's study on the role of ingroup favoritism in American attitudes toward a trade policy, Brutger et al. changed the relative gains in jobs for the US in the vignettes they presented to the respondents (Table 5). The manipulation check asks whether the vignette says the US gains more, less, or the same as the other country. In their analysis, Brutger et al. code Condition 3 as the treatment condition and the other two conditions as the control. We disagree with this coding because both Conditions 2 and 3 present a trade deal in which the US gains relative to the country and, if the theoretically relevant feature is the relative gain for the US, they should both be coded as the treatment condition [9]. The manipulation check, however, is

[9]This does not affect the validity of Brutger et al.'s (2022a) framework and findings, which are independent of the theoretical interpretation of this particular experiment. In their book, Brutger et al. (2022b) clarifies the substantive interpretation to be aligned with the coding decision instead of the original theory.

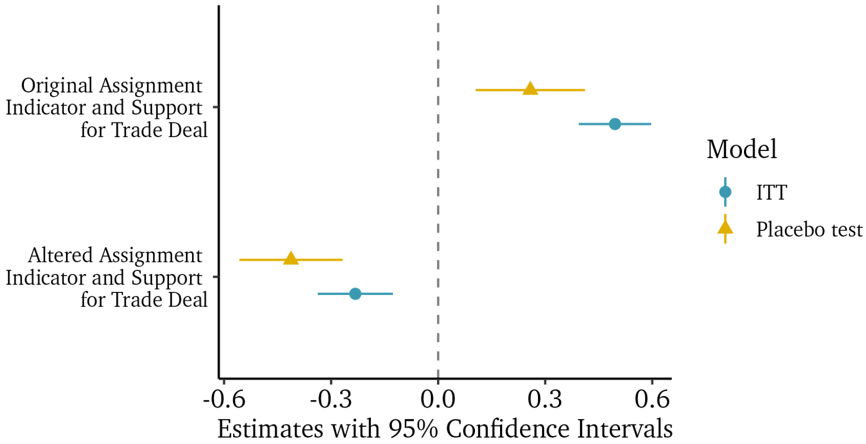

Figure 4. ITT, effect of information reception, and estimates from placebo test of Ingroup Favoritism study (N=1507) in Brutger et al. (2022a). Confidence intervals are calculated with HC2 standard errors. No control variables are included. Table A3 shows the full results in numerical format.

coded in line with the theory being tested in that if a respondent assigned to Condition 2 chooses "The US gains more," the recall is coded as correct. We discovered this after our first placebo test, in which we restricted our analysis to those with incorrect recall across the three conditions but left the coding of the indicator of treatment assignment unchanged. The estimate thus targets $\mathbb{E}(Y_i \mid Z_i = 1, X_i = 0) - \mathbb{E}(Y_i \mid Z_i = 0, X_i = 0)$ in Table 3. As the coefficient from the placebo test (in the upper right in Figure 4) shows, the assignment to treatment has a positive effect even for those who fail to recall the information hypothesized to be theoretically relevant.

We then recode the indicator of treatment assignment by coding Condition 1, where the relative gain for the US is negative, as 0 and the other two conditions, where the US gains relative to the other country, as 1. If ingroup favoritism is indeed driving trade attitudes, we should expect the indicator of treatment assignment—in this case, the provision of the information that the US would benefit more—to have no effect on trade attitudes for those who failed the manipulation checks. However, as Figure 4 shows, the ITT is negative after we recode the treatment indicator and the treatment indicator has an even larger effect on trade attitudes for those who failed to answer correctly which country would benefit more. This indicates that relative gain for the US is not the primary driver behind the respondents' support for a trade policy.

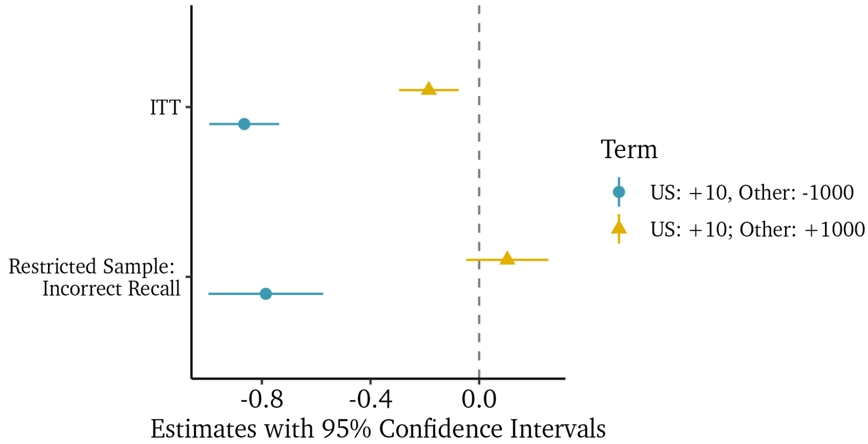

Figure 5. Analyses with a factorial coding of treatment assignment of Ingroup Favoritism study (Brutger et al. 2022a). Baseline: US: +1000; Other: +10. Confidence intervals are calculated with HC2 standard errors. N: 1507 for ITT estimation and 668 for placebo test. No control variables are included. Table A3 shows the full results in numerical format.

Part of the problem here is that responses may vary within grouped treatments. In a further analysis, we create a factorial variable to indicate which condition a respondent is assigned to and perform several analyses with "Condition 3" (US: + 1,000 jobs; Other Country: -1,000) as the reference level. Figure 5 shows the results. The reference level is the one in which the US gains 1,000 jobs while the other country gains only 10. As the figure shows, relative to the reference condition, respondents assigned to "US: +10; Other: -1,000" are much less likely to favor the trade deal. Further, respondents assigned to this condition dislike the trade deal even more than those assigned to a condition in which the relative loss for the US is large ("US: +10; Other: +1,000"). More interestingly, those assigned to "US: +10; Other: -1,000" but cannot recall that the US gains relative to the other country still dislike the trade deal compared to those assigned to the reference condition but similarly cannot recall which country gains relatively. This suggests the difference in the respondents' attitudes toward the trade deals presented in the two conditions is not driven primarily by considerations of the relative gain or loss for the US; rather, they may be driven by considerations of overall welfare improvement for both countries but with slightly larger weights for American job gains. The results from these placebo tests also cast doubt on another study with similar variation in the treatment arms (Mutz and Lee 2020).[10]

Although estimates from placebo tests should not be interpreted at face value, the assumptions needed for them to be informative about the theorized mechanism are mild. For the placebo tests we conducted to be informative, we need to assume: Conditional on having received the treatment-relevant information or its placebo counterpart, the ability to answer manipulation checks correctly should not differ between the treatment and placebo groups. This is only violated when the treatment-relevant information somehow has a perverse effect on participants' ability to recall the information relative to the placebo information. Even if such an effect exists, for the placebo test to be uninformative, the perverse effect would have to be large enough to explain away the placebo effect.

In cases similar to the experiment reanalyzed in this section, the exclusion restriction is violated, and thus, an IV approach would be inappropriate without a sensitivity analysis. But it also means that the theory that gives rise to the experimental design might not be valid regardless of the magnitude and certainty of the ITT estimate.

5 An Original Survey Experiment in Taiwan

In this section, we demonstrate the utility of accounting for heterogeneity with respect to belief change with a survey experiment conducted in Taiwan. The experiment pertains to one of the most important changes in global politics: China's rising share of global trade and its implications for democracy. On the one hand, trade and economic integration have been touted as promoters of democracy, as more trade from more democratic to less democratic nations can transmit political values (Magistretti and Tabellini 2022). Scholars and policymakers in the West once hoped that trade and other economic engagements with China would lead the country toward political liberalization and democratization. However, there is a risk that trade with China would

[10]More recent work by Mutz, Mansfield, and Kim (2021) and Lobo and Brutger (2023) suggest that racially based heterogeneity in treatment effects may partially explain the inconsistent effects of relative job gains on support for trade agreements.

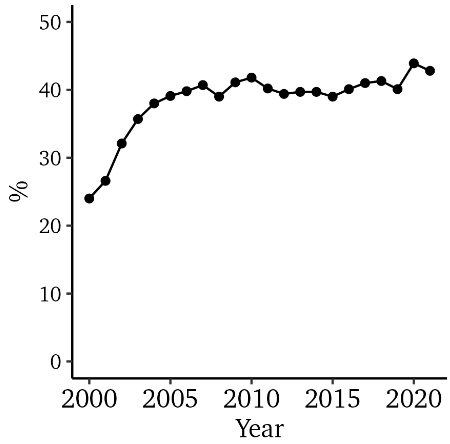

Figure 6. Export from Taiwan to mainland China (incl. Hong Kong) as a percentage of total export; in monetary value. *Source*: National Statistics, Republic of China (Taiwan)

also have consequences on the political institutions of the democratic world or even other authoritarian states. Trade with China could raise the salience of its economic success as an authoritarian state, which may, in turn, generate a rally-around-the-flag effect or other forms of resistance among recipients of that trade (Foot 2020). Alternatively, trade may give China economic tools that advance its own model of politics (Krishnarajan, Doucette, and Andersen 2022).

The democratic costs of economic interdependence with China might be most acute in Taiwan, which not only trades intensively with China but also faces an existential threat to its democracy. Since Taiwan and mainland China loosened restrictions on unofficial exchanges, cross-strait trade has expanded significantly (Fig. 6). The PRC has used tariffs and other economic sanctions against perceived moves toward independence and away from the One China "1992 Consensus."[11] The tension between economic development and democracy (Lin 2016) has become particularly salient under the latest administration, which has taken steps to divert trade from China toward Southeast Asia (through the "New Southbound Policy") while characterizing Taiwan as being on "democracy's first line of defense." In the following, we examine whether shifts in Taiwanese perceptions of economic dependence on mainland China affect the

[11]The 1992 Consensus was a verbal agreement reached in a meeting of semi-official representatives from Taiwan and mainland China in 1992.

prioritization of economic development versus democratic institutions.

5.1 Experimental Design

To measure respondents' prior belief about the trade dependence of Taiwan on mainland China, we ask all respondents to provide their best guess about the share of Taiwan's export that went to mainland China (including Hong Kong) in 2021; here we provide the respondents with the share of Taiwan's export that went to the United States as a benchmark. Then all respondents are asked a battery of demographic questions. After that, a randomly selected group of respondents are shown the true figure, as well as the extent, in percentage, to which it is larger than their guess and the extent to which it is larger than the share of Taiwanese export that went to the US.

Then, after asking a battery of occupational questions, we measure the posterior belief about the export share with a different question format. This time, we ask respondents to use slides to indicate their best guesses about the shares of Taiwanese exports that went to mainland China, the US, Japan, South Korea, the European Union, and all other economies, respectively. We require all figures to add up to 100%. We take the value for mainland China to be the posterior belief about Taiwan's export dependence on mainland China. For our main outcome of interest, we measure respondents' relative weighting of economic development and democracy using the following question with a five-point Likert scale:

> Between
>
> 1) sustained economic development and
>
> 2) free and fair elections,
>
> some think the former is more important, while others think the latter is more important. Which do you think is more important for Taiwan?

We use stratified randomization by dividing respondents into strata based on partisan leaning, political knowledge, and whether the respondents believe the national interests and security of Taiwan are more aligned with those of mainland China or the US. The National Taiwan University Web Survey (NTUWS) team administered the survey in Taiwan to a sample of around 1,036 respondents from November 29, 2022 to December 12, 2022 (Chang and Tseng 2023). The platform has a relatively strict screening procedure and requires users to register accounts using their mobile phone numbers. The sample of respondents approximates the voting-age population in Taiwan in terms of gender, sex, education, party affiliation, and region of residence.[12]

5.2 Estimation

For expository purposes, we use binary and factorial codings for our instrument and report first-stage and reduced-form results for both. We aim to contrast the results for the former with those for the latter. The binary coding ($\texttt{Binary}_i$) indicates whether the respondent is assigned to the instrument. The two-factor coding ($\texttt{2factor}_i$) divides those assigned to the instrument into two categories: whether their prior belief is less than 39% (upward correction) or not (downward or no correction).[13]

In our regressions, we use the following coding procedure for non-binary discrete variables. For covariates we use for blocking, we create stratum indicators (12 in total). For the dependent variable, a five-point measure, we use the deviation from the mean divided by the range, i.e., $(X_i - \overline{X}_i)/4$. This puts the dependent variable on a 0-1 scale. For prior belief, we use the deviation from the mean divided by 100. For change in belief, we use the difference between the posterior and the prior, divided by 100.

For our first-stage and reduced-form estimates, we use a stratified estimator for the analyses with binary and two-factor codings of the instrument.[14] We then regress outcomes on interactions between stratum indicators and treatment. In the case of the IV analysis, the second stage involves the interaction of stratum indicators and belief change. For that second-stage interaction, we use a binary indicator for whether the prior belief is below the true figure and the original continuous variable.

[12]See Section A.4.3 in the appendix for more details on our sampling and stratified randomization.

[13]In Section A.4.6 in the appendix, we also show results using a continuous coding of the instrument.

[14]In Section A.4.5 in the appendix, we show that the estimates retrieved from an OLS regression are substantively similar.

5.3 Results

Table A4 in the appendix summarizes the three covariates we use for blocking as well as `gender` and `education` by instrument status. The summary statistics for the blocking variables suggest the stratified randomization was successful.

Table A5 in the appendix shows the respondents' prior beliefs, posterior beliefs, and changes in beliefs by instrument status. It is noteworthy that the median of their guesses is the true figure, a finding consistent with the empirical regularity of the "wisdom of the crowd" documented across a variety of settings (Galton 1907; Hong and Page 2004). The descriptive statistics show the instrument works as intended: Those who are assigned to the instrument correct their beliefs to be closer to the true figure and the standard deviation of the posterior beliefs of those assigned to the instrument shrinks compared to those assigned to control, albeit not by a very large magnitude. The shrunk but still troublingly large standard deviation is evidence that the respondents assigned to the instrument do not expend high cognitive effort on the information provided, the post-treatment guessing task, or both.

The estimated effects of the instrument on belief change are presented in Panel A of Figure 7. Note that, because the change in belief is not rescaled, the coefficients reflect effects on changes in the belief (which itself is measured on a percentage scale), not changes in proportional terms. When there is no distinction in the heterogeneity in the instrument (i.e., the instrument is coded as `Binary`), the effects of the instrument on belief change for respondents with beliefs below and above the truth cancel each other, resulting in an estimate statistically indistinguishable from 0 at a significance level of 0.05. When the non-monotonicity of the instrument is taken into account, however, the effects are substantively significant. For example, in the model that uses a two-factor coding of the instrument (`2factor`), the upward-correction instrument raises the reported export dependence by 7.5 percentage points relative to the control group. When coded as a continuous variable, the first stage is also strong. Roughly, the instrument would correct the belief of a respondent with a prior of 21% upwards by

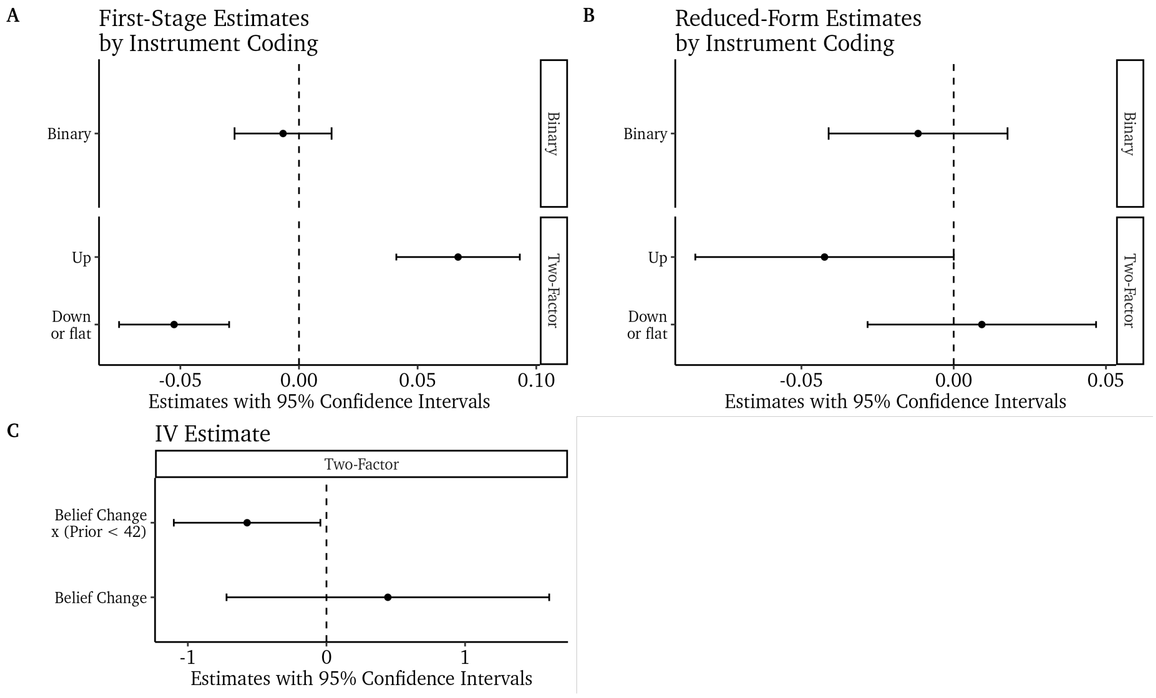

Figure 7. Estimates of the Taiwan experiment (N=1036). Confidence intervals are calculated with HC2 standard errors. Table A6 shows the results of unweighted regressions for the first stage and reduced form. For the IV estimation, Table A7 in the appendix shows the full results.

20% in absolute terms while decreasing the belief of someone with a prior of 63% by 20%. In Table A3 of the appendix, we display the average prior and posterior belief across treatment conditions.

Panel B in Figure 7 displays the reduced-form effects of the instrument using two different factorial codings. When coded as a binary variable, as is the common practice in many survey experiments that use information treatments, the estimates are noisy and statistically indistinguishable from 0. When the non-monotonicity is taken into account, however, we can see that exposure to an upward correction reduces the respondents' `EconDem` attitude by 4.3%, i.e., learning about Taiwan's greater trade dependence with China induces a 4.3% shift in prioritizing economic development over democratic institutions. This effect is consistent with the Chinese government's messaging strategy on these issues having some effect.

Panel C in Figure 7 shows results of the two-stage least square (2SLS) regressions using the `2factor` coding of the instrument. Again, we offer a binary coding of prior belief to ease interpretation. The effect of increasing belief in trade dependence by 10%, shifting individuals with a lower-than-truth prior (e.g., from 32% to 42%), is to increase support for development over democracy by about 5.7%. Again, this result suggests that Chinese government's messaging strategy could be effective and that Taiwanese respondents do not exhibit a backlash to dependence on the mainland. This suggests, contrary to a "rally-around-the-flag" effect that citizens' awareness of globalization, even from a dominant and increasingly assertive government, does not generate support for democratic institutions at home.

6 Conclusion

In this paper, we seek to draw political scientists' attention to a problem in the use of experiments to study how changes in informational beliefs affect their downstream attitudes and behavior. The goal in such studies is to vary the independent variable, in this case, knowledge and beliefs. When the treatment of theoretical interest is thus defined, the treatment effect of the provision of some information could mask a lack of movement on two levels: failure to absorb information and lack of belief change.

Put together, we provide three suggestions for political scientists using informational survey experiments for their research, summarized in Figure 8. First, if the researcher is interested in finding the most effective, policy-relevant informational intervention, using the ITT as the estimand is justified. We suggest an adaptive design rather than a static one in such cases. Second, there is the question of whether the goal of the analysis is to study the effect of information reception or belief change as the independent variable of interest. In the former case, researchers should include treatment-relevant manipulation checks in their surveys. They can then use placebo tests to examine whether their experiment manipulates the theoretically expected variable.

Alternatively, or in addition, researchers may be interested in the effect of changes in some informational beliefs. In this case, they should assess whether "nuisance beliefs," i.e., beliefs that are not of direct interest to the researchers but can be manipulated by the experimental conditions, are downstream with respect to the belief of theoretical interest. This is the case when they change only after changes in the belief of interest; they are not if the instrument manipulates them simultaneously with, or prior to, the belief of interest. In the former scenario, researchers can use an IV analysis to retrieve the effect of changes in beliefs; in the latter, in addition to an IV analysis, they should conduct a sensitivity analysis to check if the IV results hold even with certain violations of the exclusion restriction.

To distill our paper into one takeaway, we advocate that researchers design their studies in a way that allows them to evaluate whether they are manipulating the independent variable of theoretical interest. In the case of informational survey experiments, this means that, unless the goal is finding the most effective, policy-relevant intervention, researchers should incorporate treatment-relevant manipulation checks into their analysis and perform placebo tests in addition to the usual ITT estimation.

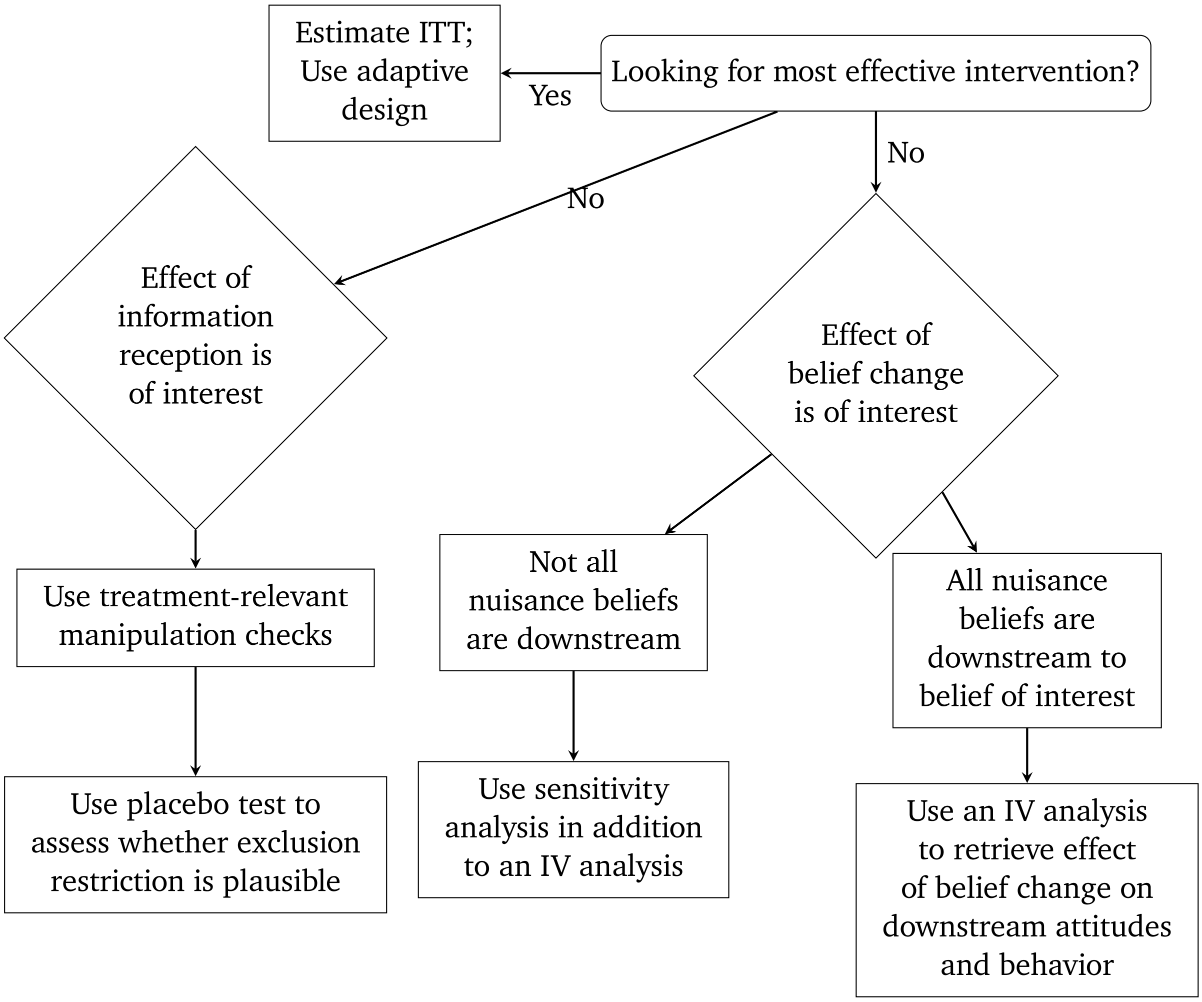

Figure 8. Decision diagram for experiments with information treatments

References

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2018. "Analyzing Causal Mechanisms in Survey Experiments." *Political Analysis* 26 (4): 357–378.

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–455.

Berinsky, Adam J., Michele F. Margolis, and Michael W. Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science* 58 (3): 739–753.

Boynton, Marcella H., David B. Portnoy, and Blair T. Johnson. 2013. "Exploring the Ethics and Psychological Impact of Deception in Psychological Research." *IRB: Ethics & Human Research* 35 (2): 7–13.

Brutger, Ryan, Joshua D. Kertzer, Jonathan Renshon, Dustin Tingley, and Chagai M. Weiss. 2022a. "Abstraction and Detail in Experimental Design." *American Journal of Political Science.*

Brutger, Ryan, Joshua D Kertzer, Jonathan Renshon, and Chagai M Weiss. 2022b. *Abstraction in experimental Design: testing the tradeoffs.* Cambridge University Press.

Butler, Daniel M, Craig Volden, Adam M Dynes, and Boris Shor. 2017. "Ideology, Learning, and Policy Diffusion: Experimental Evidence." *American Journal of Political Science* 61 (1): 37–49.

Chang, Yu-Tzung, and Huan-Kai Tseng. 2023. "Wanglu yuqing fenxi yu diaocha yanjiu fangfa xin tiaozhan 網路輿情分析與調查研究方法新挑戰 [New Challenges in Online Public Opinion Analysis and Survey Research Methods]." In *Zhengzhixue de xiankuang yu zhanwang [The State and Outlook of Political Science],* 73–104. In Mandarin Chinese. Wu-Nan Book Inc.

Christensen, Love. 2023. "Optimal Persuasion under Confirmation Bias: Theory and Evidence From a Registered Report." *Journal of Experimental Political Science* 10 (1): 4–20.

Cinelli, Carlos, and Chad Hazlett. 2019. "Making Sense of Sensitivity: Extending Omitted Variable Bias." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82 (1): 39–67.

Coppock, Alexander. 2022. *Persuasion in Parallel: How Information Changes Minds about Politics.* Chicago Studies in American Politics. Chicago; London: The University of Chicago Press.

Dafoe, Allan, Baobao Zhang, and Devin Caughey. 2018. "Information Equivalence in Survey Experiments." *Political Analysis* 26 (4): 399–416.

Druckman, James N. 2022. *Experimental Thinking: A Primer on Social Science Experiments.* Cambridge University Press.

Druckman, James N., and Mary C. McGrath. 2019. "The Evidence for Motivated Reasoning in Climate Change Preference Formation." *Nature Climate Change* 9 (2): 111–119.

Eggers, Andy, Guadalupe Tuñón, and Allan Dafoe. 2023. "Placebo Tests for Causal Inference." *American Journal of Political Science.*

Foot, Rosemary. 2020. "China's Rise and US Hegemony: Renegotiating Hegemonic Order in East Asia?" *International Politics* 57 (2): 150–165.

Galton, Francis. 1907. "Vox Populi." *Nature* 75 (7): 450–451.

Gul, Faruk, and Wolfgang Pesendorfer. 2008. "The Case for Mindless Economics." *The Foundations of Positive and Normative Economics: A Handbook* 1:3–42.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart. 2023. "Designing Information Provision Experiments." *Journal of Economic Literature* 61 (1): 3–40.

Hainmueller, Jens, and Dominik Hangartner. 2013. "Who Gets a Swiss Passport? A Natural Experiment in Immigrant Discrimination." *American Political Science Review* 107 (1): 159–187.

Hong, Lu, and Scott E. Page. 2004. "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers." *Proceedings of the National Academy of Sciences* 101 (46): 16385–16389.

Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* 1st edition. New York: Cambridge University Press.

Kane, John V., and Jason Barabas. 2019. "No Harm in Checking: Using Factual Manipulation Checks to Assess Attentiveness in Experiments." *American Journal of Political Science* 63 (1): 234–249.

Kennedy, Brian, Alec Tyson, and Cary Funk. 2022. *Americans' Trust in Scientists, Other Groups Declines.* https://www.pewresearch.org/science/2022/02/15/americans-trust-in-scientists-other-groups-declines/. Accessed: July 27, 2023.

Krishnarajan, Suthan, Jonathan Doucette, and David Andersen. 2022. "Early-Adulthood Economic Experiences and the Formation of Democratic Support." *British Journal of Political Science,* 1–20.

Lin, Syaru Shirley. 2016. *Taiwan's China Dilemma: Contested Identities and Multiple Interests in Taiwan's Cross-Strait Economic Policy.* Stanford, California: Stanford University Press.

Little, Andrew T., and Thomas B. Pepinsky. 2021. "Learning from Biased Research Designs." Publisher: The University of Chicago Press, *The Journal of Politics* 83 (2): 602–616.

Little, Andrew T., Keith E. Schnakenberg, and Ian R. Turner. 2022. "Motivated Reasoning and Democratic Accountability." *American Political Science Review* 116 (2): 751–767.

Lobo, Daniel, and Ryan Brutger. 2023. "Fairness According to Whom?: Divergent Perceptions of Fairness Among White and Black Americans and its Effect on Trade Attitudes." *Working Paper.*

Magistretti, Giacomo, and Marco Tabellini. 2022. "Economic Integration and the Transmission of Democracy."

Mattingly, Daniel, Trevor Incerti, Changwook Ju, Colin Moreshead, Seiki Tanaka, and Hikaru Yamagishi. 2022. "Chinese Propaganda Persuades a Global Audience That the 'China Model' is Superior: Evidence from a 19-Country Experiment." *Working Paper.*

Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. 2018. "How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It." *American Journal of Political Science* 62 (3): 760–775.

Mutz, Diana C. 2021. "Improving Experimental Treatments in Political Science." In *Advances in Experimental Political Science,* edited by James N. Druckman and Donald P. Green, 219–238. Cambridge University Press.

Mutz, Diana C., and Eunji Kim. 2017. "The Impact of In-group Favoritism on Trade Preferences." *International Organization* 71 (4): 827–850.

Mutz, Diana C., and Amber Hye-Yon Lee. 2020. "How Much is One American Worth? How Competition Affects Trade Preferences." *American Political Science Review* 114 (4): 1179–1194.

Mutz, Diana, Edward D Mansfield, and Eunji Kim. 2021. "The racialization of international trade." *Political Psychology* 42 (4): 555–573.

Naoi, Megumi. 2020. "Survey Experiments in International Political Economy: What We (Don't) Know about the Backlash Against Globalization." *Annual Review of Political Science* 23:333–356.

Nicholson, Stephen P. 2012. "Polarizing Cues." *American Journal of Political Science* 56 (1): 52–66.

Offer-Westort, Molly, Alexander Coppock, and Donald P. Green. 2021. "Adaptive Experimental Design: Prospects and Applications in Political Science." *American Journal of Political Science* 65 (4): 826–844.

Press, Daryl G., Scott D. Sagan, and Benjamin A. Valentino. 2013. "Atomic Aversion: Experimental Evidence on Taboos, Traditions, and the Non-Use of Nuclear Weapons." *American Political Science Review* 107 (1): 188–206.

Rho, Sungmin, and Michael Tomz. 2017. "Why Don't Trade Preferences Reflect Economic Self-Interest?" *International Organization* 71 (S1): S85–S108.

Rothwell, Jonathan. 2020. *Assessing the Economic Gains of Eradicating Illiteracy Nationally and Regionally in the United States.* https://www.barbarabush.org/reports. Accessed: July 10, 2023.

Strezhnev, Anton, Judith G. Kelley, and Beth A. Simmons. 2021. "Testing for Negative Spillovers: Is Promoting Human Rights Really Part of the "Problem"?" *International Organization* 75 (1): 71–102.

Taber, Charles S., and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50 (3): 755–769.

Tomz, Michael R., and Jessica L. P. Weeks. 2013. "Public Opinion and the Democratic Peace." *The American Political Science Review* 107 (4): 849–865.

Villar, Sofía S., Jack Bowden, and James Wason. 2015. "Multi-armed Bandit Models for the Optimal Design of Clinical Trials: Benefits and Challenges." *Statistical Science* 30 (2): 199–215.

Yu, Arthur Zeyang. 2023. “A Binary IV Model for Persuasion: Profiling Persuasion Types among Compliers.” *Working Paper.*

A Appendix for *Information Exposure and Belief Manipulation in Survey Experiments*

A.1 Checking the Use of Checks

We first run a search for articles that include the word "experiment" in either the title, abstract, or keyword list. We then exclude articles that use conjoint, discrete choice, or field experiments. We include experiments that manipulate a piece of information between treatment arms to change respondents' beliefs about some aspect of the world, real or hypothetical. We thus exclude articles that use textual variation to arouse different psychological states in the respondents that cannot be fully characterized by changes in factual beliefs. We also exclude studies that vary non-textual visual stimuli, such as the skin tone of a hypothetical candidate.

We include the resulting papers as studies that include survey experiments with information treatments. We then search in the main articles and the appendices for one of the following word stems: "check," "manipu," and "atten" to examine if the papers mention they include manipulation or attention checks in their main studies (not just in their pilot studies).

Table A1. A Review of Papers That Use Informational Survey Experiments.

Paper	MC[1]	TRMC[2]	SMC[3]	Pass[4]	Journal
Arriola and Grossman (2021)	0	0	0		JoP
Aytaç (2021)	1	0	1		APSR
Bakker, Lelkes, and Malka (2020)	0	0	0		JoP
Bayram and Graham (2022)	0	0	0		JoP
Bisgaard (2019)	0	0	0		AJPS
Boas, Hidalgo, and Toral (2021)[15]	0	0	0		JoP
Boudreau, Elmendorf, and MacKenzie (2019)	0	0	0		AJPS
Bush and Zetterberg (2021)	1	1	0	0.29	AJPS
Bøttkjær and Justesen (2021)	0	0	0		JoP
Campbell et al. (2019)	0	0	0		JoP
Campbell and Spilker (2022)	0	0	0		JoP
Cebul, Dafoe, and Monteiro (2021)[16]	1	1	0		JoP
Chapman and Chaudoin (2020)	0	0	0		JoP
Chow and Han (2023)	1	0	0		JoP
Chu and Recchia (2022)	0	0	0		JoP
Clayton, O'Brien, and Piscopo (2019)	1	1	0	0.93	AJPS
Condon and Wichowsky (2020)	1	0	0		JoP
Croco, Hanmer, and McDonald (2021)	0	0	0		JoP
Culpepper, Jung, and Lee (2023)	1	0	0		AJPS
Dias and Lelkes (2022)	0	0	0		AJPS
Druckman et al. (2022)	1	0	0		JoP
Duell et al. (2023)	0	0	0		JoP
Eck et al. (2021)	1	0	0		JoP
Fang and Li (2020)	0	0	0		JoP
Findor et al. (2023)	0	0	0		APSR
Gaikwad and Nellis (2021)	0	0	0		AJPS
Gandhi and Ong (2019)	0	0	0		AJPS
Gerber, Patashnik, and Tucker (2022)	0	0	0		JoP
Gerver, Lown, and Duell (2023)	0	0	0		JoP
Gottlieb (2022)	0	0	0		AJPS

Continued on next page

[15]The field experiment in this paper uses a manipulation check but the survey experiment does not.
[16]Data on the manipulation checks are not available in the public data set.

Table A1 – continued from previous page

Paper	MC	TRMC	SMC	Pass	Journal
Herzog, Baron, and Gibbons (2022)	0	0	0		JoP
Hill and Huber (2019)	1	0	0		AJPS
Jones and Brewer (2019)	0	0	0		JoP
Kam and Deichert (2020)	0	0	0		JoP
Karpowitz et al. (2021)	1	0	0		JoP
Kenwick and Maxey (2022)	1	1	0	0.54	JoP
Kim et al. (2023)	1	0	0		AJPS
Klar and McCoy (2021)	0	0	0		AJPS
Kobayashi et al. (n.d.)	0	0	0		AJPS
Krupnikov and Levine (2019)	0	0	0		JoP
Larsen (2021)	0	0	0		JoP
Lupu and Wallace (2019)[17]	1	1	0		AJPS
Madsen et al. (2022)	0	0	0		APSR
Malhotra, Monin, and Tomz (2019)	0	0	0		APSR
Manekin and Mitts (2022)	1	0	1		APSR
Martin and Raffler (2021)	0	0	0		AJPS
Mattes and Weeks (2019)	1	1	0	0.41	AJPS
Mullin and Hansen (2022)	0	0	0		AJPS
Mutz and Lee (2020)	1	0	1		APSR
Myrick (2020)	1	1	0	0.47	JoP
Nelson and Gibson (2019)	0	0	0		JoP
Pereira et al. (2022)	0	0	0		JoP
Porter and Wood (2022)	0	0	0		JoP
Powers and Altman (2023)	1	1	1	0.93	AJPS
Powers and Renshon (2021)	1	1	0	0.87	AJPS
Robison (2022)	0	0	0		JoP
Sances (2021)	0	0	0		AJPS
Stephens-Dougan (2023)	0	0	0		APSR
Thachil (2020)	1	0	1		JoP
Todd et al. (2021)	0	0	0		JoP
Tomz and Weeks (2020a)	0	0	0		JoP

Continued on next page

[17]Data on the manipulation checks are not available in the public data set.

Table A1 – continued from previous page

Paper	MC	TRMC	SMC	Pass	Journal
Tomz and Weeks (2020b)	1	0	0		APSR
Velez, Porter, and Wood (2023)	0	0	0		JoP
Westwood, Messing, and Lelkes (2020)	0	0	0		JoP
Xu, Kostka, and Cao (2022)	1	0	0		JoP
Yair, Sulitzeanu-Kenan, and Dotan (2020)	1	0	0		JoP
Zhu and Shi (2019)	0	0	0		JoP

[1] Manipulation checks

[2] Treatment-relevant manipulation checks

[3] Subjective manipulation checks

[4] For treatment-relevant manipulation checks

A.2 Measurement Error for Information Reception

We prove Proposition 1 in this section.

Proof. First, note that we have one-sided noncompliance by design: those assigned to the control condition cannot take up the treatment information. Let $\tilde{D}_i$ be our measure of D_i for those assigned to the treatment. Then

$$\mathbb{E}(\hat{\theta}) = \frac{\mathbb{E}(Y_i \mid Z_i = 1) - \mathbb{E}(Y_i \mid Z_i = 0)}{\underbrace{\mathbb{E}(\tilde{D}_i \mid Z_i = 1)}_{\tau}} \tag{4}$$

$$\tau = \mathbb{E}(\tilde{D}_i = 1, D_i = 0 \mid Z_i = 1) + \mathbb{E}(\tilde{D}_i = 1, D_i = 1 \mid Z_i = 1) \tag{5}$$

$$= \underbrace{\mathbb{E}(\tilde{D}_i = 1, D_i = 0 \mid Z_i = 1)}_{\epsilon_1} + \mathbb{E}(D_i = 1 \mid Z_i = 1) - \underbrace{\mathbb{E}(\tilde{D}_i = 0, D_i = 1 \mid Z_i = 1)}_{\epsilon_2} \tag{6}$$

ϵ_1 is the mismeasurement of those who are assigned to treatment but fail to take it up and ϵ_2 the mismeasurement of those who are assigned to treatment and also take it

Table A2. Results of replication and re-analysis of the Nuclear Weapons and Elite Messaging studies

	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	−0.20***	−0.20***	−0.20***	−0.06	−0.06	0.22
	(0.06)	(0.06)	(0.06)	(0.09)	(0.09)	(0.16)
Nuclear Weapons ITT	0.47***		0.08			
	(0.09)		(0.09)			
Nuclear Weapons Info Repcetion		0.81***				
		(0.15)				
Elite Messaging ITT				0.43***		−0.01
				(0.13)		(0.21)
Elite Messaging Info Repcetion					0.77***	
					(0.24)	
N	535	524	365	278	278	108
R^2	0.052	0.103	0.002	0.036	0.019	0.000

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

up. When $\epsilon_1 = \epsilon_2$, we have

$$\mathbb{E}(\hat{\theta}) = \theta \tag{7}$$

□

A.3 More Information on Re-analysis of Three Experiments

Columns (1) and (3) in Table A2 replicate the original analysis of the Nuclear Weapons and Elite Messaging experiments in Brutger et al. (2022). These are ITT estimates. Columns (2) and (4) show the results of the IV estimation we visualize in Figure 3.

Figure A1 shows the results of placebo tests for Brutger et al.'s (2022) replications of Press, Sagan, and Valentino (2013) and Nicholson (2012). The results show that there is no evidence that differences in outcomes are driven by informational mechanisms outside of that theorized by the original authors. These results correspond to Columns (5) and (6) in Table A2.

Table A3 shows the results we visualize in Figures 4 and 5. Column (1) shows the results of the placebo test for the Outgroup Favoritism experiment in Brutger et al. (2022), run on the subsample of participants who failed to recall the treatment

Table A3. Results of replication and re-analysis of the Ingroup Favoritism study

	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	0.03	−0.17***	0.39***	0.14***	0.32***	0.28***
	(0.05)	(0.03)	(0.05)	(0.04)	(0.04)	(0.06)
US Gains More	0.26***	0.50***				
	(0.08)	(0.05)				
Altered Indicator			−0.41***	−0.23***		
			(0.07)	(0.05)		
Factor: Other Country Loses					−0.86***	−0.79***
					(0.07)	(0.11)
Factor: US Gains Less					−0.19***	0.10
					(0.06)	(0.08)
N	668	1507	668	1507	1507	668
R^2	0.015	0.049	0.038	0.011	0.120	0.129

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

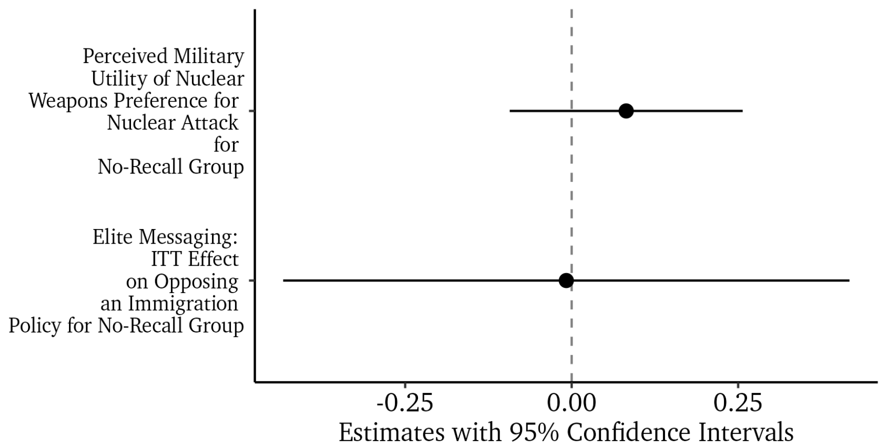

Figure A1. Difference in means for respondents with an incorrect recall of the provided information. 95% confidence intervals calculated with HC2 standard errors.

as intended by the researchers. Column (2) shows results of the corresponding ITT estimation. Column (3) shows the results of another placebo test after we altered the treatment indicator to be more aligned with the proposed theory. Column (4) shows the corresponding ITT estimate. Columns (5) and (6) show the results of a placebo test with a factorial assignment indicator and the corresponding ITT estimation.

A.4 The Taiwan Survey

A.4.1 Relations between the People's Republic of China and the Republic of China

Since the end of the Chinese Civil War, the People's Republic of China (PRC) has contested the sovereignty of Taiwan's government and opposed its moves toward formal independence. In 1996, the Republic of China held its first direct presidential election, contested by the "Chinese Nationalist Party," Kuomintang (KMT) and the Democratic Progressive Party (DPP). In subsequent elections, while both parties are committed to the status quo, the KMT favors unification and closer ties with the PRC while the DPP leans toward independence. At this point, a stable majority of the voting population supports maintaining the status quo over moving toward independence or unification, but over 60% identifies as "Taiwanese" as opposed to either "Chinese" or "both Taiwanese and Chinese," according to the National Chengchi University Election Study Center (2023).

A.4.2 Questionnaire Design

We vary the format of the questions we use to elicit the respondents' prior and posterior beliefs. Figures A2 and A3 show the questions we use for the prior and posterior, respectively. The question for eliciting the prior reads as follows:

> In 2021 (ROC Year 110), Taiwan's exports to the United States will account for about 15% of the total. What do you think is the percentage of Taiwan's export value that went to mainland China (including Hong Kong) in 2021? Please the slider to indicate a guess that you think is closest to the true value.

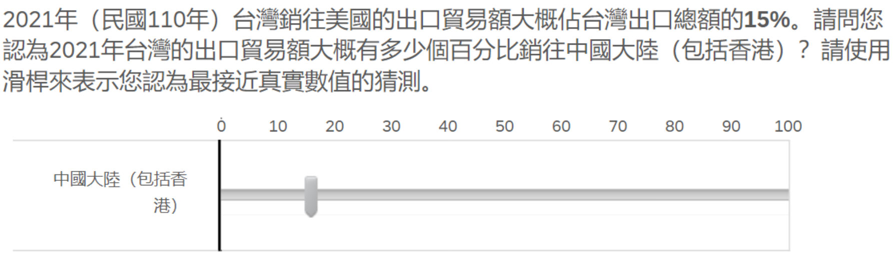

Figure A2. Elicitation of the prior

- Mainland China (including Hong Kong) [slider]

(default is 15.)

The question for eliciting the posterior reads as follows:

What percentage of Taiwan's exports in 2021 (ROC Year 110) do you think went to the following countries or regions? Please use the slider below to answer this question. The total should equal 100%.

- Mainland China (including Hong Kong)
- USA
- Japan
- South Korea
- European Union
- All other economies

The information provided to the treatment group reads as follows:

According to the Ministry of Finance, about 42%, or four-tenths, of Taiwanese exports in 2021 (ROC Year 110) went to mainland China (including Hong Kong). This is [A] and 180% higher than Taiwan's export to the US.

where [A] is filled out using the following rules:

If Prior $<$ 39 ("Up"): $\frac{(42-\text{Prior})}{\text{Prior}} * 100\%$ higher than your previous guess

Else if Prior $>$ 45 ("Down"): $\frac{(42-\text{Prior})}{\text{Prior}} * 100\%$ lower than your previous guess

Else ("Flat"): about the same as your previous guess

Figure A3. Elicitation of the posterior

A.4.3 More Information on Randomization and Sampling

We use stratified randomization for potential gains in efficiency. For partisan lean, we divide respondents into three categories: KMT-leaning, DPP-leaning, and independent. For political knowledge, we divide respondents into two strata based on whether they correctly answer get at least three questions. For the third covariate, we subtract their perception of the alignment between the US and Taiwan from their perception of that between mainland China and Taiwan and divide the respondents based on whether this figure is positive. This procedure implies twelve strata. We implement this with the branching function in Qualtrics.

Because of the inherent limitations of online survey platforms, we under-sample from people aged 55 and above and over-sample from other age groups relative to the population. In terms of educational level, we over-sample from college graduates. The NTUWS team uses weighted sampling to obtain a sample from the respondent pool. It then sends up to five waves of text messages to the selected users to control the differences in non-response rates among different strata. As we show in the descriptive statistics in Table A4 in the appendix, there is some discrepancy between our target quotas for the two age groups and the actual percentages in our sample. For the other covariates, the resulting sample largely meets our targets.

A.4.4 Descriptive Statistics

Table A4 shows summary statistics for the three covariates we use for blocking as well as `gender` and `education` by instrument status. : the proportions of those assigned to control in each category of each variable are nearly identical to the percentages of each category in the sample. `gender` and `education` are also balanced between the instrument and control group, even though we did not block on these two covariates. The covariate statistics broken down by instrument status suggest that the blocking covariates that have some predictive power for whether individuals underestimate ("up" for "corrected upwards") or overestimate ("down" for "corrected downwards") are `partisanship` and `alignment`. KMT supporters and independents on average overestimate the figure while DDP supporters do not exhibit such a deviation on average. Second, those who view Taiwan as more aligned with China relative to the US on average overshoot in their estimate and are more likely to overshoot than those who view Taiwan as more aligned with the US are likely to undershoot. Overall, however, even these two covariates are not strong predictors of the direction of deviation in individual belief relative to the truth.

Table A5 shows prior beliefs, posterior beliefs, and changes in beliefs by instrument status.

A.4.5 Regression Results

Figure A4 and Table A6 show the results of the first stage and reduced form obtained with an unweighted regression estimator for binary and two-factor codings of the instrument, respectively. The point and uncertainty estimates are very close to those retrieved from the stratified estimator because our strata are largely balanced.

Columns (1) and (2) in Table A7 show the full results of the IV regressions for the

Table A4. Pre-treatment covariates by instrument status

	Control	Down	Flat	Up	Total
	(N=511)	(N=232)	(N=80)	(N=213)	(N=1036)
Partisanship					
KMT	134 (26.2%)	65 (28.0%)	15 (18.8%)	55 (25.8%)	269 (26.0%)
DPP	175 (34.2%)	70 (30.2%)	43 (53.8%)	73 (34.3%)	361 (34.8%)
Neither	202 (39.5%)	97 (41.8%)	22 (27.5%)	85 (39.9%)	406 (39.2%)
Polknow					
<3	58 (11.4%)	29 (12.5%)	8 (10.0%)	28 (13.1%)	123 (11.9%)
≥ 3	453 (88.6%)	203 (87.5%)	72 (90.0%)	185 (86.9%)	913 (88.1%)
Alignment					
China	249 (48.7%)	127 (54.7%)	41 (51.3%)	94 (44.1%)	511 (49.3%)
US	262 (51.3%)	105 (45.3%)	39 (48.8%)	119 (55.9%)	525 (50.7%)
Gender					
Female	250 (48.9%)	110 (47.4%)	36 (45.0%)	106 (49.8%)	502 (48.5%)
Male	261 (51.1%)	122 (52.6%)	44 (55.0%)	107 (50.2%)	534 (51.5%)
Education					
$\geq$ College	351 (68.7%)	163 (70.3%)	55 (68.8%)	147 (69.0%)	716 (69.1%)
$<$ College	160 (31.3%)	69 (29.7%)	25 (31.3%)	66 (31.0%)	320 (30.9%)

Table A5. Prior, posterior, and belief change by instrument status

	Control	Down	Flat	Up	Total
	(N=511)	(N=232)	(N=80)	(N=213)	(N=1036)
Prior belief					
Mean (SD)	43.5 (20.5)	62.4 (10.3)	41.4 (1.78)	21.1 (9.98)	43.0 (20.9)
Median	42.0	62.0	41.0	20.0	42.0
Posterior belief					
Mean (SD)	39.1 (19.3)	40.4 (15.5)	40.3 (11.3)	35.5 (14.7)	38.8 (17.2)
Median	40.0	42.0	41.0	40.0	40.0
Change in belief					
Mean (SD)	-4.41 (17.0)	-22.0 (17.5)	-1.10 (11.5)	14.4 (15.6)	-4.22 (20.4)
Median	0	-22.5	0	12.0	-1.00

two-factor coding in Figure 7 in the main text. We include all indicators for the strata we use for randomization as well as their interactions with the instrument and the endogenous variable (belief change).

Table A6. Results of first-stage and reduced-form regressions

	Binary	Two-factor	Binary	Two-factor
(Intercept)	−0.04***	−0.04***	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Instrument	0.00		−0.01	
	(0.01)		(0.02)	
Prior	−0.64***	−0.51***	0.00	−0.06
	(0.03)	(0.03)	(0.04)	(0.05)
Instrument: Down or flat		−0.05***		0.01
		(0.01)		(0.02)
Instrument: Up		0.07***		−0.05**
		(0.01)		(0.02)
N	1036	1036	1036	1036
R^2	0.427	0.458	0.000	0.005

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

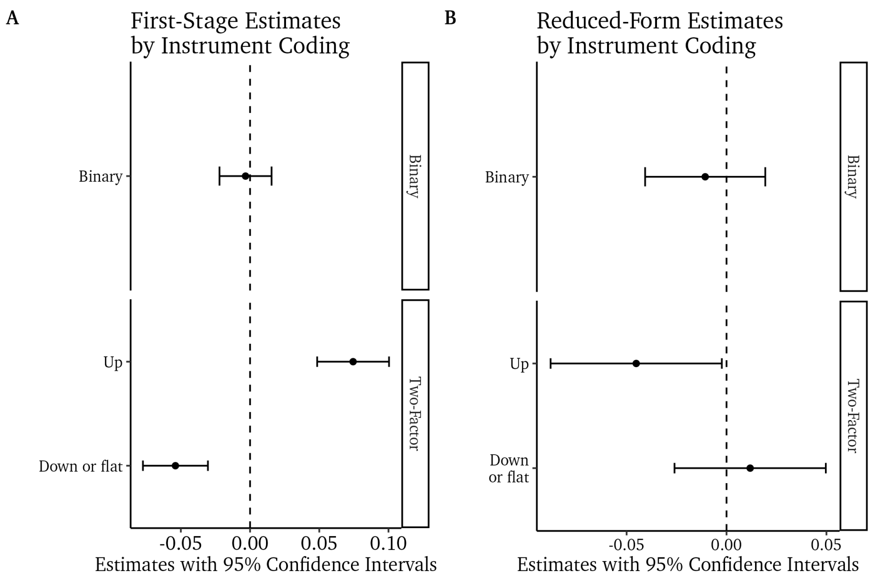

Figure A4. Estimates of the first stage and reduced form using an unweighted OLS estimator (i.e., blocking covariates are not used in the estimation). Confidence intervals are calculated using HC2 standard errors

Table A7. Results of IV estimation with a two-factor coding

	(1)	(2)	(3)	(4)
(Intercept)	0.07	0.10*	0.08	0.10**
	(0.04)	(0.05)	(0.06)	(0.05)
Belief Change (BC)	-0.23	0.44	-0.47	0.40
	(0.54)	(0.59)	(0.95)	(1.00)
Prior	-0.23	-0.07	-0.07	-0.10
	(0.14)	(0.12)	(0.12)	(0.12)
Prior < 42		-0.04		-0.04
		(0.03)		(0.03)
BC × Prior < 42		-0.57**		-0.53**
		(0.27)		(0.22)
N	1036	1036	1036	1036
R^2	-0.004	0.035	0.015	0.011

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

A.4.6 Alternative Specifications

We also show the results of the IV model using a continuous coding of the instrument. The continuous coding ($\texttt{Cont}_i$) is a fine-grained measure that captures all the variation in the construction of our instrument and is calculated as follows:

$$\texttt{Cont}_i = -\texttt{Binary}_i \times (\texttt{Prior}_i - 42)/100 \quad (8)$$

where 42 is the true figure. The coding is non-monotonic in that it accounts for the opposite signs of the expected effects of the instrument on those with priors below versus above the given figure. The last two columns in Table A7 show the results of the regressions. As before, we include all indicators for the randomization strata and their interactions with the instrument and the endogenous variable. The estimates show that an upward correction affects attitude. Loosely, a 10% increase in terms of level in one's belief for respondents with a lower-than-truth prior causes the respondents to be about 5.3% more favorable toward economic development relative to democratic elections. These estimates are relatively imprecise and should be interpreted with caution when used to inform substantive theories.

Papers Cited in Appendix

Arriola, Leonardo R., and Allison N. Grossman. 2021. "Ethnic Marginalization and (Non)Compliance in Public Health Emergencies." *The Journal of Politics* 83 (3): 807–820.

Aytaç, SELIM ERDEM. 2021. "Effectiveness of Incumbent's Strategic Communication during Economic Crisis under Electoral Authoritarianism: Evidence from Turkey." *American Political Science Review* 115 (4): 1517–1523.

Bakker, Bert N., Yphtach Lelkes, and Ariel Malka. 2020. "Understanding Partisan Cue Receptivity: Tests of Predictions from the Bounded Rationality and Expressive Utility Perspectives." *The Journal of Politics* 82 (3): 1061–1077.

Bayram, A. Burcu, and Erin R. Graham. 2022. "Knowing How to Give: International Organization Funding Knowledge and Public Support for Aid Delivery Channels." *The Journal of Politics* 84 (4): 1885–1898.

Bisgaard, Martin. 2019. "How Getting the Facts Right Can Fuel Partisan-Motivated Reasoning." *American Journal of Political Science* 63 (4): 824–839.

Boas, Taylor C., F. Daniel Hidalgo, and Guillermo Toral. 2021. "Competence versus Priorities: Negative Electoral Responses to Education Quality in Brazil." *The Journal of Politics* 83 (4): 1417–1431.

Bøttkjær, Louise, and Mogens K. Justesen. 2021. "Why Do Voters Support Corrupt Politicians? Experimental Evidence from South Africa." *The Journal of Politics* 83 (2): 788–793.

Boudreau, Cheryl, Christopher S. Elmendorf, and Scott A. MacKenzie. 2019. "Racial or Spatial Voting? The Effects of Candidate Ethnicity and Ethnic Group Endorsements in Local Elections." *American Journal of Political Science* 63 (1): 5–20.

Brutger, Ryan, Joshua D. Kertzer, Jonathan Renshon, Dustin Tingley, and Chagai M. Weiss. 2022. "Abstraction and Detail in Experimental Design." *American Journal of Political Science.*

Bush, Sarah Sunn, and Pär Zetterberg. 2021. "Gender Quotas and International Reputation." *American Journal of Political Science* 65 (2): 326–341.

Campbell, Rosie, Philip Cowley, Nick Vivyan, and Markus Wagner. 2019. "Why Friends and Neighbors? Explaining the Electoral Appeal of Local Roots." *The Journal of Politics* 81 (3): 937–951.

Campbell, Susanna P., and Gabriele Spilker. 2022. "Aiding War or Peace? The Insiders' View on Aid to Postconflict Transitions." *The Journal of Politics* 84 (3): 1370–1383.

Cebul, Matthew D., Allan Dafoe, and Nuno P. Monteiro. 2021. "Coercion and the Credibility of Assurances." *The Journal of Politics* 83 (3): 975–991.

Chapman, Terrence L., and Stephen Chaudoin. 2020. "Public Reactions to International Legal Institutions: The International Criminal Court in a Developing Democracy." *The Journal of Politics* 82 (4): 1305–1320.

Chow, Wilfred, and Enze Han. 2023. "Descriptive Legitimacy and International Organizations: Evidence from United Nations High Commissioner for Refugees." *Journal of Politics* 85 (2): 357–371.

Chu, Jonathan A., and Stefano Recchia. 2022. "Does Public Opinion Affect the Preferences of Foreign Policy Leaders? Experimental Evidence from the UK Parliament." *The Journal of Politics* 84 (3): 1874–1877.

Clayton, Amanda, Diana Z. O'Brien, and Jennifer M. Piscopo. 2019. "All Male Panels? Representation and Democratic Legitimacy." *American Journal of Political Science* 63 (1): 113–129.

Condon, Meghan, and Amber Wichowsky. 2020. "Inequality in the Social Mind: Social Comparison and Support for Redistribution." *The Journal of Politics* 82 (1): 149–161.

Croco, Sarah E., Michael J. Hanmer, and Jared A. McDonald. 2021. "At What Cost? Reexamining Audience Costs in Realistic Settings." *The Journal of Politics* 83 (1): 8–22.

Culpepper, Pepper D., Jae-Hee Jung, and Taeku Lee. 2023. "Banklash: How Media Coverage of Bank Scandals Moves Mass Preferences on Financial Regulation." *American Journal of Political Science* n/a (n/a).

Dias, Nicholas, and Yphtach Lelkes. 2022. "The Nature of Affective Polarization: Disentangling Policy Disagreement from Partisan Identity." *American Journal of Political Science* 66 (3): 775–790.

Druckman, James N., Samara Klar, Yanna Krupnikov, Matthew Levendusky, and John Barry Ryan. 2022. "(Mis)estimating Affective Polarization." *The Journal of Politics* 84 (2): 1106–1117.

Duell, Dominik, Lea Kaftan, Sven-Oliver Proksch, Jonathan Slapin, and Christopher Wratil. 2023. "Communicating the Rift: Voter Perceptions of Intraparty Dissent in Parliaments." *The Journal of Politics* 85 (1): 78–91.

Eck, Kristine, Sophia Hatz, Charles Crabtree, and Atsushi Tago. 2021. "Evade and Deceive? Citizen Responses to Surveillance." *The Journal of Politics* 83 (4): 1545–1558.

Fang, Songying, and Xiaojun Li. 2020. "Historical Ownership and Territorial Disputes." *The Journal of Politics* 82 (1): 345–360.

Findor, Andrej, Matej Hruška, Roman Hlatky, Tomáš Hrustič, and Zuzana Bošeľová. 2023. "Equality, Reciprocity, or Need? Bolstering Welfare Policy Support for Marginalized Groups with Distributive Fairness." *American Political Science Review* 117 (3): 805–821.

Gaikwad, Nikhar, and Gareth Nellis. 2021. "Do Politicians Discriminate Against Internal Migrants? Evidence from Nationwide Field Experiments in India." *American Journal of Political Science* 65 (4): 790–806.

Gandhi, Jennifer, and Elvin Ong. 2019. "Committed or Conditional Democrats? Opposition Dynamics in Electoral Autocracies." *American Journal of Political Science* 63 (4): 948–963.

Gerber, Alan S., Eric M. Patashnik, and Patrick D. Tucker. 2022. "How Voters Use Contextual Information to Reward and Punish: Credit Claiming, Legislative Performance, and Democratic Accountability." *The Journal of Politics* 84 (3): 1839–1843.

Gerver, Mollie, Patrick Lown, and Dominik Duell. 2023. "Proportional Immigration Enforcement." *The Journal of Politics* 85 (3): 949–968.

Gottlieb, Jessica. 2022. "How Economic Informality Constrains Demand for Programmatic Policy." *American Journal of Political Science* 00 (00): 1–18.

Herzog, Stephen, Jonathon Baron, and Rebecca Davis Gibbons. 2022. "Antinormative Messaging, Group Cues, and the Nuclear Ban Treaty." *The Journal of Politics* 84 (1): 591–596.

Hill, Seth J., and Gregory A. Huber. 2019. "On the Meaning of Survey Reports of Roll-Call "Votes"." *American Journal of Political Science* 63 (3): 611–625.

Jones, Philip Edward, and Paul R. Brewer. 2019. "Gender Identity as a Political Cue: Voter Responses to Transgender Candidates." *The Journal of Politics* 81 (2): 697–701.

Kam, Cindy D., and Maggie Deichert. 2020. "Boycotting, Buycotting, and the Psychology of Political Consumerism." *The Journal of Politics* 82 (1): 72–88.

Karpowitz, Christopher F., Tyson King-Meadows, J. Quin Monson, and Jeremy C. Pope. 2021. "What Leads Racially Resentful Voters to Choose Black Candidates?" *The Journal of Politics* 83 (1): 103–121.

Kenwick, Michael R., and Sarah Maxey. 2022. "You and Whose Army? How Civilian Leaders Leverage the Military's Prestige to Shape Public Opinion." *The Journal of Politics* 84 (4): 1963–1978.

Kim, Sung Eun, Jong Hee Park, Inbok Rhee, and Joonseok Yang. 2023. "Target, Information, and Trade Preferences: Evidence from a Survey Experiment in East Asia." *American Journal of Political Science.*

Klar, Samara, and Alexandra McCoy. 2021. "Partisan-Motivated Evaluations of Sexual Misconduct and the Mitigating Role of the #MeToo Movement." *American Journal of Political Science* 65 (4): 777–789.

Kobayashi, Yoshiharu, Menevis Cilizoglu, Tobias Heinrich, and William Christiansen. n.d. "No Entry in a Pandemic: Public Support for Border Closures." *American Journal of Political Science* n/a (n/a).

Krupnikov, Yanna, and Adam Seth Levine. 2019. "Political Issues, Evidence, and Citizen Engagement: The Case of Unequal Access to Affordable Health Care." *The Journal of Politics* 81 (2): 385–398.

Larsen, Martin Vinæs. 2021. "How Do Voters Hold Politicians Accountable for Personal Welfare? Evidence of a Self-Serving Bias." *The Journal of Politics* 83 (2): 740–752.

Lupu, Yonatan, and Geoffrey P. R. Wallace. 2019. "Violence, Nonviolence, and the Effects of International Human Rights Law." *American Journal of Political Science* 63 (2): 411–426.

Madsen, Mikael Rask, Juan A. Mayoral, Anton Strezhnev, and Erik Voeten. 2022. "Sovereignty, Substance, and Public Support for European Courts' Human Rights Rulings." *American Political Science Review* 116 (2): 419–438.

Malhotra, Neil, Benoît Monin, and Michael Tomz. 2019. "Does Private Regulation Preempt Public Regulation?" *American Political Science Review* 113 (1): 19–37.

Manekin, Devorah, and Tamar Mitts. 2022. "Effective for Whom? Ethnic Identity and Nonviolent Resistance." *American Political Science Review* 116 (1): 161–180.

Martin, Lucy, and Pia J. Raffler. 2021. "Fault Lines: The Effects of Bureaucratic Power on Electoral Accountability." *American Journal of Political Science* 65 (1): 210–224.

Mattes, Michaela, and Jessica L. P. Weeks. 2019. "Hawks, Doves, and Peace: An Experimental Approach." *American Journal of Political Science* 63 (1): 53–66.

Mullin, Megan, and Katy Hansen. 2022. "Local News and the Electoral Incentive to Invest in Infrastructure." *American Political Science Review,* 1–6.

Mutz, Diana C., and Amber Hye-Yon Lee. 2020. "How Much is One American Worth? How Competition Affects Trade Preferences." *American Political Science Review* 114 (4): 1179–1194.

Myrick, Rachel. 2020. "Why So Secretive? Unpacking Public Attitudes toward Secrecy and Success in US Foreign Policy." *The Journal of Politics* 82 (3): 828–843.

National Chengchi University Election Study Center. 2023. *Taiwanese/Chinese Identity (1992/06~2022/06).* esc.nccu.edu.tw/PageDoc/Detail?fid=7800&id=6961. Accessed: 2023-03-27.

Nelson, Michael J., and James L. Gibson. 2019. "How Does Hyperpoliticized Rhetoric Affect the US Supreme Court's Legitimacy?" *The Journal of Politics* 81 (4): 1512–1516.

Nicholson, Stephen P. 2012. "Polarizing Cues." *American Journal of Political Science* 56 (1): 52–66.

Pereira, Frederico Batista, Natália S. Bueno, Felipe Nunes, and Nara Pavão. 2022. "Fake News, Fact Checking, and Partisanship: The Resilience of Rumors in the 2018 Brazilian Elections." *The Journal of Politics* 84 (4): 2188–2201.

Porter, Ethan, and Thomas J. Wood. 2022. "Political Misinformation and Factual Corrections on the Facebook News Feed: Experimental Evidence." *The Journal of Politics* 84 (3): 1812–1817.

Powers, Kathleen E., and Dan Altman. 2023. "The Psychology of Coercion Failure: How Reactance Explains Resistance to Threats." *American Journal of Political Science* 67 (1): 221–238.

Powers, Ryan, and Jonathan Renshon. 2021. "International Status Concerns and Domestic Support for Political Leaders." *American Journal of Political Science* n/a (n/a).

Press, Daryl G., Scott D. Sagan, and Benjamin A. Valentino. 2013. "Atomic Aversion: Experimental Evidence on Taboos, Traditions, and the Non-Use of Nuclear Weapons." *American Political Science Review* 107 (1): 188–206.

Robison, Joshua. 2022. "Partisan Influence in Suspicious Times." *The Journal of Politics* 84 (3): 1683–1696.

Sances, Michael W. 2021. "Presidential Approval and the Inherited Economy." *American Journal of Political Science* 65 (4): 938–953.

Stephens-Dougan, LAFLEUR. 2023. "White Americans' Reactions to Racial Disparities in COVID-19." *American Political Science Review* 117 (2): 773–780.

Thachil, Tariq. 2020. "Does Police Repression Spur Everyday Cooperation? Evidence from Urban India." *The Journal of Politics* 82 (4): 1474–1489.

Todd, Jason Douglas, Edmund J. Malesky, Anh Tran, and Quoc Anh Le. 2021. "Testing Legislator Responsiveness to Citizens and Firms in Single-Party Regimes: A Field Experiment in the Vietnamese National Assembly." *The Journal of Politics* 83 (4): 1573–1588.

Tomz, Michael R., and Jessica L. P. Weeks. 2020a. "Human Rights and Public Support for War." *The Journal of Politics* 82 (1): 182–194.

———. 2020b. "Public Opinion and Foreign Electoral Intervention." *American Political Science Review* 114 (3): 856–873.

Velez, Yamil R., Ethan Porter, and Thomas J. Wood. 2023. "Latino-Targeted Misinformation and the Power of Factual Corrections." *The Journal of Politics* 85 (2): 789–794.

Westwood, Sean Jeremy, Solomon Messing, and Yphtach Lelkes. 2020. "Projecting Confidence: How the Probabilistic Horse Race Confuses and Demobilizes the Public." *The Journal of Politics* 82 (4): 1530–1544.

Xu, Xu, Genia Kostka, and Xun Cao. 2022. "Information control and public support for social credit systems in China." *The Journal of Politics* 84 (4): 2230–2245.

Yair, Omer, Raanan Sulitzeanu-Kenan, and Yoav Dotan. 2020. "Can Institutions Make Voters Care about Corruption?" *The Journal of Politics* 82 (4): 1430–1442.

Zhu, Boliang, and Weiyi Shi. 2019. "Greasing the Wheels of Commerce? Corruption and Foreign Investment." *The Journal of Politics* 81 (4): 1311–1327.