When Guidance Changes: Government Stances and Public Beliefs*

Charlie Rafkin†
Advik Shreekumar‡
Pierre-Luc Vautrey§

October 20, 2020

Abstract

Governments often make early recommendations about issues that remain uncertain. Do governments' early positions affect how much people believe the latest recommendations? We investigate this question using an incentivized online experiment with 1,900 US respondents in early April 2020. We present all participants with the latest CDC projection about coronavirus death counts. We randomize exposure to information that highlights how President Trump previously downplayed the coronavirus threat. When the President's inconsistency is salient, participants are less likely to revise their prior beliefs about death counts from the projection. They also report lower trust in the government. These results align with a simple model of signal extraction from government communication, and have implications for the design of changing guidelines in other settings.

*This paper previously circulated as "When Guidance Changes: Government Inconsistency and Public Beliefs" and "When Guidance Changes: Early Government Stances and Downstream Credibility." It also includes material that previously circulated in a draft titled "Public Inference from Crisis Policy: Evidence from the COVID-19." For helpful feedback, we thank Daron Acemoglu, Hunt Allcott, Abhijit Banerjee, Jonathan Cohen, Rebekah Dix, Esther Duflo, Ahmet Gulek, Lisa Ho, Andrei Shleifer, Evan Soltas, and particularly Amy Finkelstein and Frank Schilbach. We thank the MIT Shultz Fund for financial support. This work has been supported (in part) by Grant #2003-22443 from the Russell Sage Foundation. Any opinions expressed are those of the principal investigator(s) alone and should not be construed as representing the opinions of the Foundation. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1122374. This study was approved by MIT's Committee on the Use of Humans as Experimental Subjects under exempt protocol #E2040 and was pre-registered at the AEA RCT Registry under ID AEARCTR-0005639.

†Department of Economics, MIT: crafkin@mit.edu.
‡Department of Economics, MIT: adviks@mit.edu.
§Department of Economics, MIT: vautrey@mit.edu.

1 Introduction

> "Right now, everyone is so confused by all the conflicting messages that, each time the guidance evolves, fewer and fewer people might follow it."
> — Richard Besser, former director of the Centers for Disease Control and Prevention (in Duhigg, 2020).

Governments carefully manage the release of official guidance to inform the public. They sometimes take a strong initial stance on issues that remain uncertain. When a government's private information subsequently evolves, it may wish to change its guidance to convey the latest available information. However, inconsistent messaging may suggest that the government's private information is inaccurate, undermining its credibility. How much do earlier statements that turn out to be ill-advised affect the government's credibility and the public's reception of official information?

This question takes on particular importance in the context of the COVID-19 (COVID) pandemic in the United States. The pandemic drew considerable attention to official guidance, with President Donald Trump's March 2020 coronavirus briefings attracting about 8.5 million cable viewers on average (Grynbaum, 2020). In such situations, public health officials emphasize the importance of "consistent messaging" and "transparent risk communications,"[1] even though these can be at odds when guidance changes: in a high-profile example from early April, the U.S. Centers for Disease Control revised its stance and began recommending mask-wearing for the general public. Gollust et al. (2020) speculate that such conflicting messages may have inhibited the COVID crisis response. This concern that inconsistent messaging undermines credibility is also present in other settings well beyond public health, for instance when the government guides the public on topics like national security or economic forecasts.

We study the effects of early stances and subsequent inconsistency by examining how shifts in government messaging affect the public's receptiveness to the most recent official information as well as attitudes toward the government. We conduct an incentivized online experiment on April 3–4, 2020, a period when official guidance was still evolving. Our intervention is light-touch, consisting of short, truthful, and well-publicized statements about federal messages or projections. In one group, the "Consistent" control, participants receive information about a recent statement the Trump administration released. In the other group, the "Inconsistent" treatment, participants receive the same recent statement, but also receive an

[1] Quotations from: Canadian Pandemic Influenza Preparedness Task Group and Henry (2019)

older Trump administration statement that downplayed the crisis and contrasts with the later statement. Statements are clearly dated. Both treatment groups then receive information about a contemporaneous projection by the CDC that up to 240,000 Americans would die of COVID. We elicit incentivized prior and posterior beliefs about the number of people who would die from COVID in the next six months, as well as perceptions of the government's response. We focus on three main outcomes: (i) the propensity to take an option to revise one's prior beliefs, (ii) the magnitude of the belief update, i.e. the absolute difference between prior and posterior, and (iii) an index of perceived government credibility.

We find substantial effects on beliefs about the severity of the crisis and perceptions of the government's credibility. First, we find that the Inconsistent treatment reduces belief updating from the CDC projection that all participants received. When exposed to the earlier downplaying messages, participants' prior beliefs become stickier. In our primary specification, 45.0% of the Consistent group chooses to update their beliefs after receiving the official projection; the Inconsistent treatment makes our participants 4.0 pp (standard error: 2.3) less likely to update at all. Moreover, the treatment reduces the magnitude of participants' average belief updates by 50.6 log points (SE: 24.9), or 39.7% (SE: 15.0). Second, we find that the Inconsistent treatment makes participants less likely to trust the government. In particular, inconsistent messaging reduces a standardized index of beliefs in the government's credibility by 0.037 SD (SE: 0.016).[2]

Our results align with a simple model where an agent jointly updates about the state of the world and the quality of the government's private information. In particular, the result of reduced government credibility suggests that the reduction in belief updating does not merely arise from confusion. Rather, inconsistent messaging reduces the government's credibility, so people place less faith in the new information. We provide suggestive evidence for this mechanism by studying heterogeneous effects of the Inconsistent sequence by levels of prior beliefs about the coronavirus's severity, showing that (i) inconsistency particularly reduces belief updating and trust among participants with moderate and pessimistic prior beliefs, and (ii) it reduces perceived government credibility the most among people with moderate prior beliefs about the coronavirus's severity. The latter result in particular suggests that the inconsistent nature of the messages plays a role.

[2]Some participants may already have encountered the information we provide. This suggests that our results are a lower bound on the actual effects of the Trump administration's having publicized inconsistent statements.

However, our experiment is underpowered to rely heavily on granular tests for treatment effect heterogeneity. We therefore lack the ability to definitively identify inconsistency as the mechanism behind the loss of credibility and reduced belief updating. Specifically, we cannot reject an alternative mechanism whereby downplaying a crisis early on hurts the government's credibility simply because it goes against most people's priors. Future research can investigate the relative importance of these two mechanisms.[3] Regardless of the mechanism, this paper shows that early downplaying by the President leads to reduced public belief updating.

Our main contribution is to provide empirical evidence for the concern that early government stances on uncertain issues can undermine faith in official public guidance. Governments frequently face a tradeoff between conveying early information and possibly losing credibility as a result of subsequent changes in guidance. For example, the public is exposed to conflicting health guidance in a variety of contexts, from cancer screening to nutrition guidelines. Carpenter et al. (2016) review a literature in public health communications that finds that conflicting messages reduce self-reported adherence to guidelines and reduce experts' credibility.[4] In other settings, governments often obtain new information about potential crises (e.g., the risk of terrorism or economic collapse), and they must weigh when to update the public. Even so, research in economics that examines the effects of changing policies on inference and government credibility — guided by formal models of belief updating and using precise and incentivized outcome measures — is scarce. Moreover, our treatment emphasizes the inconsistency of the President's statements only. Yet it also induces reduced belief updating about information that is clearly labeled as being from the CDC. We thus provide evidence of spillovers in credibility across institutions.

Our paper adds to several literatures. Most narrowly, this paper contributes to contemporaneous research on how communication shapes public perceptions about COVID (Allcott et al., 2020; Ajzenman et al., 2020; Barrios and Hochberg, 2020; Bursztyn et al., 2020a; Grossman et al., 2020; Painter and Qiu, 2020). Several studies analyze how people update their beliefs from information about COVID, often with a focus

[3]For instance, future experiments building on our design could add a treatment arm where an initial position is provided in isolation. Comparing this arm with the Inconsistent arm can test whether inconsistency in itself plays a role.

[4]For instance, Nagler (2014) finds a correlation between exposure to conflicting nutrition information and self-reported confusion about best practices. Several health communications studies, including Tan et al. (2017) (in the context of e-cigarettes) and Clark et al. (2019) (in the context of diet), randomize providing conflicting vs. consistent news articles. Our paper adds to this literature by using a randomized intervention with incentivized measures of prior and posterior beliefs to test a formal model of belief updating. Unlike these papers, our experimental design also tests how inconsistent messaging affects the propensity to believe a projection that all participants received; these papers typically just show that the valence of news coverage affects attitudes toward the health behavior.

on predictions about the economy (Coibion et al., 2020; Gutierrez et al., 2020; Fetzer et al., Forthcoming; Binder, 2020).[5] Unlike these papers, we focus on inconsistencies in public information by employing a design that holds constant a common projection, and varies exposure to contradicting messages.[6]

More broadly, our work relates to a large literature that studies how to optimally provide information to improve public outcomes.[7] In contrast to the many papers that provide information and study the effects on outcomes, we are not principally concerned with an evaluation of the messages' content. Instead, we emphasize the implications of early government recommendations about topics that are uncertain. Other papers studying the effects of official recommendations include Oster (2018), Einav et al. (Forthcoming), and Oster (2020). These papers do not examine whether inconsistent guidance affects the credibility of the advisor (and therefore reduces the public's belief updates). Besides our focus on inconsistency, we depart from other research by conducting a randomized evaluation of the effects of government information on quantitative belief updates; we show that downplaying a (real) crisis can erode public confidence either because of inconsistency or because it will be perceived as non-credible. Furthermore, we show that this erosion of confidence can spill over from the President to other institutions (here, the CDC).

In addition, we contribute to a theoretical and empirical literature in political economy that studies people's inferences from information disclosed by the government, particularly during political crises or regime changes (Gehlbach and Sonin, 2014; Hollyer et al., 2015; Iachan and Nenov, 2015; Gehlbach et al., 2016).[8] A separate body of work in political economy considers the interaction between regime strength and public information disseminated through various channels, including the internet, public media, or state information (Ananyev et al., 2019; Chen and Yang, 2019; Cantoni et al., 2019). Relative to these literatures, our paper asks different questions about the inferences made from crisis communications. We also leverage a randomized intervention and precise measures of beliefs. Relatedly, we provide an empirical assessment of the conventional wisdom among communications strategists that politicians benefit from staying "on

[5] More generally, we relate to a broader literature in macroeconomics that considers the implications of vacillating macroeconomic policy for public inferences (Morris and Shin, 2002; Blinder, 2007; Morris and Shin, 2007; Chahrour, 2014). As we do not focus on the signals' macroeconomic effects, our modeling and empirical focus have a substantially different flavor.

[6] In heterogeneity analysis, we find consistency has a smaller effect on people who support Trump, which adds to research on partisanship and coronavirus policies (Allcott et al., 2020; Barrios and Hochberg, 2020; Bursztyn et al., 2020a).

[7] See Haaland et al. (2020) for a review of information provision experiments in various settings.

[8] Many of the related theoretical models emphasize what can be learned from propaganda, particularly in non-democratic contexts (Edmond, 2013; Huang, 2015; Shadmehr and Bernhardt, 2015; Little, 2017). We consider official government messages about a public-health crisis, rather than propaganda.

message" (Sigelman and Sigelman, 1986; Perloff, 1998; Carville and Begala, 2002; Benoit et al., 2011; Burton et al., 2015; Doherty et al., 2016). Similar to Tomz and Van Houweling (2012) and Tomz and Van Houweling (2016), we conduct an experiment that presents changing positions from a politician; unlike these papers, we emphasize beliefs about the world and the overall credibility of the government, rather than electoral outcomes.

Before proceeding with the analysis, we note two caveats. First, because COVID is salient to participants, it serves as an ideal setting to experimentally study the question of evolving public guidance. On the other hand, we acknowledge that it may be difficult to generalize from the COVID crisis to other contexts. Second, we interpret statements linked to President Trump as an appropriate proxy for the federal government's communications. These statements are a natural starting point because of the attention paid to Trump's March and April COVID policy. Nevertheless, to the extent that agents update differently about Trump's statements versus other politicians', our experiment may not generalize to other federal communications. At the very least, the experiment shows that Presidential communication may affect trust in the federal government as a whole and in other institutions.

2 Framework

To highlight the framework that guides our experiment's design, we informally discuss how an agent might conduct inference from the government's communications. In Appendix A, we provide a model of Bayesian updating that formalizes these intuitions.

Suppose the world can be in three possible states: no crisis, moderate crisis, or severe crisis. There are two periods and the state of the world is sticky, so the first period state persists into the second period with probability greater than one half. In each period the government observes a private signal about the state of the world and can publicize it. An agent holds prior beliefs about the current state of the world as well as about the accuracy of the government's private signals. Given these prior beliefs and a sequence of official messages, she updates her beliefs about the state of the world and the informativeness of official signals.

Consider inferences under the following sequences of official signals, which map onto two sequences of information the government could release about COVID. First, the government could issue an early optimistic signal ("there is no crisis") and then a more pessimistic signal ("there is a moderate crisis") about the state of the world, which we call an "Inconsistent" sequence of signals. Alternatively, the government could say nothing initially and only later issue the "moderate crisis" signal about the state of the world, which we call a "Consistent" sequence.[9] Observing an Inconsistent sequence generally leads the agent to believe that the government's signals are less informative than if she observed a Consistent sequence. The exception is if the agent is very optimistic about the initial state and believes the state of the world was likely to change quickly, in which case an Inconsistent sequence more plausibly mirrors reality. Thus, the agent generally updates less about the state of the world with inconsistent messaging, because her prior beliefs about the state are tied to her beliefs about the signal's informativeness. To summarize:

Implication 1 Inconsistent signals reduce belief updates about the state relative to Consistent signals.

Implication 2 Inconsistent signals reduce posterior government credibility relative to Consistent signals.

Implications 1 and 2 hold unless people have very optimistic priors about the crisis and believe that the state of the world is very likely to change between the two signals.

Next, consider how Consistent and Inconsistent sequences are received by people who have different prior beliefs about the crisis in the first period. Define three categories of prior beliefs about the crisis: optimists, moderates, and pessimists, whose priors respectively favor the "no crisis," "moderate crisis," and "severe crisis" states.

Since the state of the world is sticky across periods, Inconsistent signals appear contradictory and suggest that the government is ill-informed. However, the difference in updating between the Inconsistent sequence to the Consistent sequence depends on the *marginal effect* of adding a no-crisis signal in the first period to a moderate crisis signal in the second period. This initial no-crisis signal aligns with optimists' prior beliefs and rescues some of the government's credibility in this group. Meanwhile, for pessimists, the marginal impact of the no-crisis signal is attenuated by the presence of the moderate crisis signal, which already damages the government's credibility. Among moderates, the no-crisis signal is the first piece of evidence of an ill-informed government. Hence the Inconsistent sequence is the most damaging for this

[9]We emphasize these sequences because they match the design of our experiment, which varies information about past inconsistency. Because many subjects in our sample have prior beliefs that are more pessimistic than the CDC's contemporaneous projection, we allow for two levels of crisis severity.

group. In short, compared to the Consistent sequence, the Inconsistent sequence adds a plausible message to an implausible one for optimists; adds an implausible message to an already implausible one for pessimists; and adds an implausible message to a plausible one for moderates. We thus obtain these auxiliary implications:

Implication 1a Consider the ratio of belief updates about the state from Inconsistent signals to belief updates about the state from Consistent signals. This ratio is lower for moderates and pessimists than for optimists.

Implication 1b This ratio is lower for moderates than pessimists.

Implication 2a Inconsistent signals reduce the posterior credibility of the government more for moderates and pessimists than for optimists.

Implication 2b Inconsistent signals reduce posterior credibility of the government more for moderates than pessimists.

Implications 1a, 1b, 2a, and 2b give an ordering of the effect of the Inconsistent sequence across both belief updates about the state and the credibility of the government, depending on prior beliefs. They yield a non-monotonicity: the Inconsistent sequence has the **largest** effect on moderates, the next largest effect on pessimists, and the **smallest** effect on optimists. Implications 1a and 2a follow from disagreement with the initial optimistic signal. Meanwhile, Implications 1b and 2b arise from the contradictory effect of the initial signal as part of the entire sequence.

Note that, holding signals constant, we expect update size to change mechanically as prior beliefs change, due to the varying distance between priors and signals. Taking the ratio in Implications 1a and 1b introduces this mechanical effect in both the numerator and denominator, so that *changes* in the ratio reflect other channels (i.e., credibility).

Finally, we note that other comparisons of messages could induce similar responses. In particular, consider a government that sends either a single no-crisis message or a single moderate crisis message. Relative to the moderate crisis message, the no-crisis message would reduce the government's credibility most among moderates and second most among pessimists. It would increase the government's credibility among optimists. Put differently, Implications 1a–2b also follow from downplaying a crisis in any period and sending no other messages. We choose to develop and test these implications in the language of inconsistent messaging to more closely reflect our experimental intervention. Our experiment always includes a "moderate crisis" report in the later period. We frame this report as contrasting with an earlier "no crisis" report for participants who receive the Inconsistent sequence. As a result, it is difficult to ascribe our empirical findings to a setting with a single message.[10]

3 Experiment

To test these predictions, we ran a pre-registered experiment on April 3–4, 2020 using an online survey provider with a nationally representative sample of 1,900 participants.[11,12,13] Most states had imposed emergency policies by this time. According to the COVID Tracking Project (2020), there were about 9,000 total deaths from COVID in the United States by April 4. We ran our experiments on Lucid.io, a nationally representative online survey platform which has been used in many academic studies (Wood and Porter, 2019; Bursztyn et al., 2020b).

3.1 Treatment

All participants received the following official projection:

> *On March 31 government officials projected that between 100,000 and 240,000 people could die from COVID-19 in the United States.*

Participants in the Consistent (control) group received two additional statements before this projection:

1. *The novel coronavirus has affected American life. On March 29, President Trump announced social distancing measures would last until at least May 1st.*

[10] This is also suggested by our empirical finding that our most optimistic respondents also have reduced trust when they see the no-crisis message.

[11] AEA pre-registration ID: AEARCTR-0005639.

[12] The sample size of 1,900 excludes participants who fail two attention checks or who did not begin the experiment by the pre-registered end date; see Appendix C for details.

[13] Our study was approved as exempt by MIT's Committee on the Use of Human Subjects (COUHES). We surveyed participants and provided them with widely circulating and accurate information from credible sources. Prior to obtaining consent from participants, we described the study as being COVID related, and instructed participants that they could stop participating at any time. We also provide helplines and other resources. While we randomize which information participants were exposed to before eliciting outcomes, we provided all participants with the full set of information (along with sources) in the experiment's debrief.

2. *The novel coronavirus has played a large role in the news. On March 31, President Trump said the coronavirus is a "great national trial unlike any we have ever faced before."*

Participants in the Inconsistent (treatment) group instead received the following statements before the CDC projection:

1. *President Trump originally said that he wanted to re-open the country by April 12. Then, on March 29, he announced social distancing measures would last until at least May 1st.*

2. *President Trump repeatedly suggested that the novel coronavirus was no worse than the flu. Then, on March 31, President Trump said the coronavirus is a "great national trial unlike any we have ever faced before."*

In short, both groups view a statement about a recent position expressed by the government followed by the CDC projection. The only experimental variation between the Consistent and Inconsistent groups is that the Inconsistent group sees a statement about a previous position that is inconsistent with the recent position. In the Consistent group, we replace the early statement with a generic sentence that has no information content. We use this design, rather than simply omitting the first sentence in the Consistent group, to avoid a potential concern that the Inconsistent group read more sentences and therefore was more fatigued.

Within each treatment group, participants received both statements by the end of the experiment. We randomized the order of the statements. We presented one of the statements directly before the information about the CDC projection, and then elicited posterior beliefs. We presented the second statement after eliciting posterior beliefs and before eliciting auxiliary outcomes. This design permits a test for heterogeneity by specific statement on the main outcome of belief updates (see Appendix B).

In context of national discussion at the time, we model Trump's initial statements as "no crisis" signals, and his later statements together with the CDC projection as "moderate crisis" signals. Although Trump's later statements contrast strongly with his initial statements, the treatment messaging provides a more moderate forecast than a previously circulating worst-case projection of 2.2 million deaths from COVID.[14]

[14]See, for example, Fink (2020) from March 16, 2020 or Stimson (2020) from March 27, 2020.

3.2 Outcomes and Heterogeneity

1. **Beliefs about COVID's severity: number of deaths.** Before treatment, we ask all participants to report their prior beliefs about the number of deaths from COVID over the next 6 months. When eliciting posterior beliefs, we explicitly remind respondents of their prior and ask them whether they wish to update. If participants do not choose to update, they are then asked to enter posterior beliefs that are identical to their prior beliefs (which we present to them). As a result, participants who do not update are making an active choice not to do so. If participants do choose to update, they assess whether their prior was too high or too low, and report a precise number for their posterior belief consistent with this assessment. These design choices are intended to isolate intentional belief updates from changes in reports due to trembling hand, uncertainty, or limited memory. We incentivize beliefs via a binarized quadratic scoring rule (Hossain and Okui, 2013) to be implemented 6 months after the experiment, discussed in Appendix D.[15]

   We principally focus on two measures:

   - **Update propensity**: the fraction of people who actively choose to update.
   - **Update magnitude**: the log absolute change in predicted number of deaths (our pre-registered main outcome), $\ln(|\text{posterior} - \text{prior}| + 1)$.[16] If participants choose not to update, we assign them posterior beliefs that are identical to prior beliefs.

   Because our design asks people whether they want to update their beliefs, the propensity to update is a simple measure that captures the tendency to use the information provided. On the other hand, the update magnitude captures additional variation in the intensive margin of belief updates, and is therefore better powered.

   We focus on the *absolute* belief update magnitude (i.e., unsigned/non-directional), rather than the *net* update magnitude (i.e., signed/directional). In our setting where signals are only partially informative, point estimates of prior beliefs are not sufficient to determine which direction each subject

[15] We implement incentives with probability 1/1,000, and discuss concerns about the credibility of low incentives paid in the future in Appendix D. See the review by Schotter and Trevino (2014) for background on belief elicitation.

[16] We pre-registered the log transform because piloting revealed that the beliefs about deaths were subject to a small amount of outliers that nevertheless induce substantial noise in the untransformed outcome. We present robustness to the untransformed variable and the inverse hyperbolic sine transform in Appendix B, and winsorization before taking logs in Appendix C.

should update beliefs as Bayesians without strong additional assumptions on prior beliefs.[17] By contrast, the absolute belief updates capture *how much* subjects use the information, without the need to benchmark *how* they use it.

2. **Perceptions and attitudes about the government.** We elicit, on a scale from 0 to 10, prior and posterior attitudes toward the government response to COVID. We ask the same questions before and after treatment, and aggregate both prior and posterior attitudes into separate indices which we call the prior and posterior "government credibility index." We normalize each index to be in units of standard deviations relative to the distribution of the Consistent group.[18]

3. **Heterogeneity by prior beliefs.** We compare results for participants with prior beliefs above, within, and below the range of the government projection of 100,000 to 240,000 deaths. Participants with prior beliefs that the number of deaths in six months will exceed 240,000 (with priors above the official projection) are "pessimists." Those with priors that deaths will be less than 100,000 (below the projection) are "optimists." Those with priors between 100,000–240,000 deaths are "moderates."

4. **Knowledge of the federal COVID response.** As a benchmark for the magnitude of our results, we provide a comparison based on whether participants self-report having knowledge about COVID policies. Before treatment, we asked participants on a scale of 0–10 about the extent to which they know about the federal COVID response.[19] We then split the sample into two groups: people with above- vs. below-median knowledge.[20]

5. **Additional outcomes.** We collected several additional outcomes. We present these, per the guidance in Banerjee et al. (2020), in Appendix B.

[17] For example, suppose participants' beliefs about COVID severity come from averaging several sources, like cable news speculation and government reports. If the Inconsistent treatment reduces the weight they place on the government report, participants can update beliefs either up or down. For example, if the government report lies below cable news, the Inconsistent treatment will cause them to update up. By contrast, if government reports lie above cable news, the Inconsistent treatment reduces their posteriors relative to priors.

[18] Appendix C provides the questions used to construct this variable. We standardize each sub-scale before aggregating, as in Kling et al. (2007). We normalize prior and posterior indices separately, relative to the Consistent group.

[19] See Appendix C for the question.

[20] 51.0% of the sample reported 4 or less.

3.3 Econometric Specifications

Our main specification is:

$$y_j = \beta n_j + \boldsymbol{X}_j \delta + \varepsilon_j, \tag{1}$$

where n is an indicator variable that is 1 if participant j received the Inconsistent treatment, and β is the coefficient of interest. In our primary specification, $\boldsymbol{X}_j$ consists of a set of stratum fixed effects, which we include since randomization was performed within strata. We present robustness to including additional demographic controls.

Inference. We obtain p-values from t-tests with heteroskedasticity robust standard errors for hypothesis tests of coefficients. We employ Wald tests when using seemingly unrelated regression or considering nonlinear hypotheses. In robustness tests, we present p-values from permutation tests (Young, 2019).

3.4 Sample, Balance, and Attrition

Sample and balance. Our sample broadly matches the general population (Table 1, Columns 1–3). Relative to the general population, our participants are slightly younger, more white, have completed more education, and earn lower amounts. Moreover, we document experimental balance across the treatment arms (Column 4, p-value of joint F-test across all baseline covariates: 0.593).

Attrition. The sample only includes participants who complete the experiment. 70.9% of participants finish the survey, and we find no evidence of differential attrition by treatment (Appendix Table D.1).

3.5 Additional Details

Appendix D discusses details about the experiment's incentives, discusses concerns about experimenter demand, explains how we stratified the sample, and provides details on minor deviations from the pre-registration that we made due to data collection issues. See the authors' websites for survey instruments.[21]

[21] See: `http://charlierafkin.com/papers/rsv_covid_changing_survey.pdf` or `https://adviksh.com/files/rsv_covid_changing_instructions.pdf`

Table 1: Experimental Sample and Balance

	(1) Proportion in US adult population	(2) Mean in Consistent arm (N = 933)	(3) Mean in Inconsistent arm (N = 967)	(4) p-value for difference Consistent - Inconsistent
Demographics				
Age Group				
18-29	0.212	0.214	0.195	0.335
30-39	0.172	0.178	0.179	1.000
40-49	0.160	0.159	0.172	0.482
50-59	0.168	0.163	0.185	0.225
60-69	0.149	0.181	0.192	0.570
70+	0.139	0.105	0.077	0.037
Female	0.513	0.547	0.517	0.213
Political Party				
Democrat	—	0.348	0.332	0.481
Republican	—	0.340	0.347	0.760
Other	—	0.312	0.321	0.721
Census Region				
Midwest	0.209	0.213	0.190	0.233
Northeast	0.172	0.211	0.214	0.921
South	0.238	0.379	0.410	0.195
West	0.381	0.196	0.186	0.620
Race				
White	0.722	0.768	0.764	0.868
Black	0.127	0.083	0.098	0.266
Asian	0.056	0.042	0.055	0.225
Native American	0.009	0.017	0.013	0.637
Other race	0.052	0.057	0.043	0.218
Prefer not to answer	—	0.033	0.026	0.415
Hispanic	0.183	0.100	0.087	0.378
Household Income				
$14,999 or less	0.106	0.152	0.120	0.047
$15,000-$24,999	0.090	0.094	0.108	0.379
$25,000-$34,999	0.089	0.119	0.111	0.619
$35,000-$49,000	0.124	0.138	0.147	0.639
$50,000-$74,999	0.174	0.164	0.178	0.458
$75,000-$99,999	0.126	0.107	0.121	0.382
$100,000-$149,999	0.150	0.099	0.086	0.377
$150,000-$199,999	0.066	0.031	0.037	0.541
$200,000 or more	0.076	0.028	0.030	0.889
Prefer not to answer	—	0.068	0.063	0.765
Education				
Some high school or less, or other	0.118	0.042	0.034	0.450
High school graduate	0.275	0.227	0.195	0.101
Associate's degree, some college, or vocational training	0.307	0.305	0.322	0.478
Bachelor's degree	0.190	0.262	0.285	0.264
Graduate or professional degree	0.111	0.164	0.163	1.000
Baseline Measures				
Attention to COVID news	—	7.75 (0.07)	7.76 (0.07)	0.92
ln(1 + expected number of COVID deaths in next 6 months)	—	11.13 (0.07)	11.18 (0.07)	0.63
Government action is not strong enough	—	5.56 (0.11)	5.48 (0.1)	0.58
Government action is appropriate	—	5.37 (0.1)	5.48 (0.1)	0.44
Government is overreacting	—	2.75 (0.09)	2.62 (0.09)	0.32
Government has, and acts on, private information	—	5.87 (0.09)	5.7 (0.09)	0.17
			Omnibus F test	p-value
			Unadjusted:	0.593
			With stratum FE:	0.784

Note: This table provides demographics for respondents who complete our survey, as well as baseline beliefs about COVID and government action. Column 1 presents statistics for the general US population from the 2018 American Communities Survey (ACS). Dashed entries are fields not reported in the ACS. Columns 2 and 3 present corresponding statistics by treatment arm. We report proportions for categorical measures, and the mean and standard error for continuous measures. Attention to news about COVID and beliefs about government action were elicited on an 11-point Likert scale (0 = not at all / strongly disagree; 10 = very closely / strongly agree). The top rows of Column 4 show Pearson χ^2-tests of differences in proportions across individual subcategory for categorical measures. The middle rows show t-tests for differences in means for continuous measures. The bottom rows of Column 4 present omnibus F-tests for balanced treatment assignment. The first test is a joint F-test across all differences (regressing treatment on indicators for subcategories as well as continuous measures); the second is a joint F-test across all differences that also includes the stratum fixed effects from our main regressions (Equation (1)).

4 Results

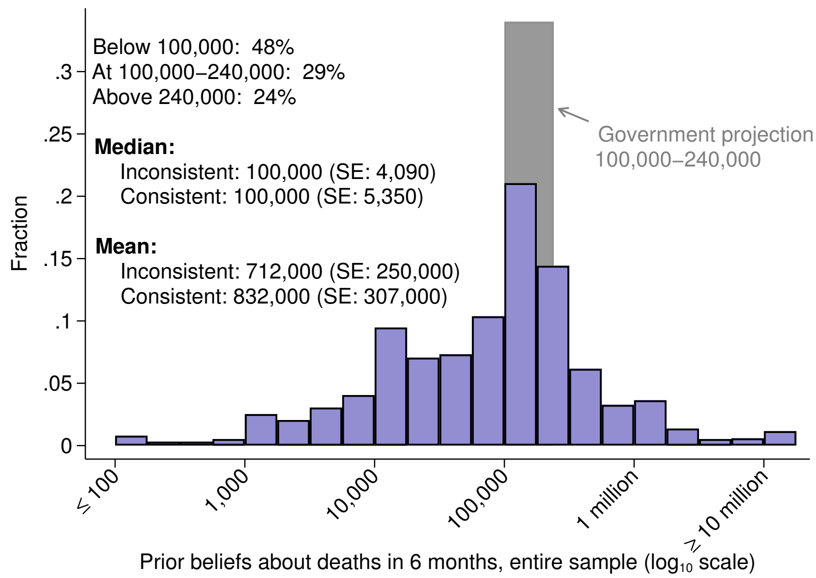
4.1 Main Result: Reduced Belief Updating

The distribution of prior beliefs elicited before the experimental manipulation shows a wide dispersion (Figure 1, Panel A). More than 99% of participants' prior beliefs about deaths in 6 months fall within the range of 0–10 million; the median prior belief is 100,000, and about a quarter of the sample has prior beliefs larger than 240,000, the upper bound of the government projection. The distribution is right-skewed and a small fraction of participants report beliefs above 10 million; we noticed this feature during piloting and therefore pre-registered our main outcomes in logs. The figure also documents that prior beliefs are balanced between the Inconsistent and Consistent groups, which is reassuring since we randomly assign treatment only after eliciting priors.

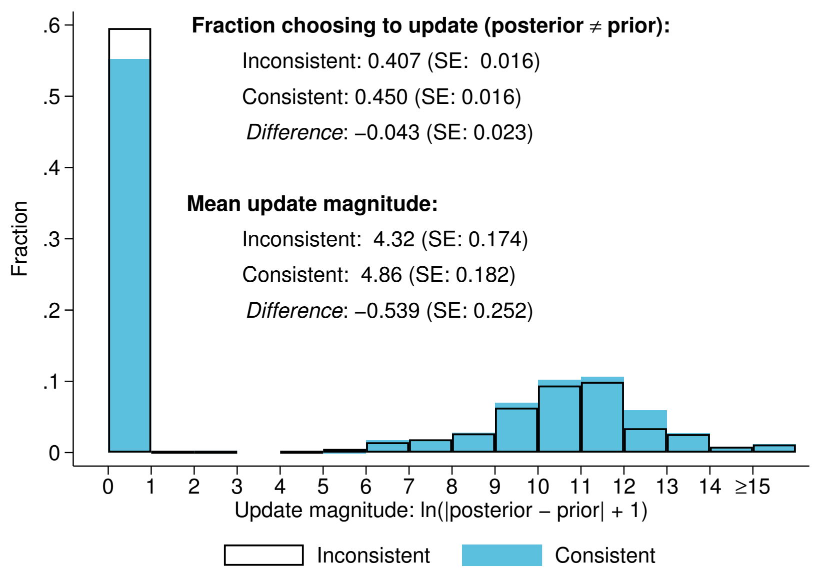
We present our first evidence in support of Implication 1 by inspecting the distribution of belief updates from the CDC information using the main outcome of $\ln(|\text{posterior} - \text{prior}| + 1)$ (Figure 1B). Less than half of participants in both the Inconsistent and Consistent groups actively choose to update their beliefs. However, the Inconsistent group is 4.3 pp less likely to update (SE: 2.3) than the Consistent group. Since 45.0% of the Consistent group updates, this difference in means constitutes about 9.6% of the control group's mean. As noted, a feature of our design is that participants must first actively choose whether to update beliefs before reporting their posteriors. Thus, this treatment effect is unlikely to arise from elicitation error. In addition, the histogram shows a smaller fraction of Inconsistent updates in the right tail of the distribution relative to the Consistent group; conditional on updating, the Inconsistent group updates less. Altogether, the difference between the Inconsistent and Consistent groups' mean unconditional update magnitudes is -53.9 log points (SE: 25.2), showing that the Inconsistent treatment substantially reduces the inferences made from the government projection.

Figure 1: Distribution of Prior Beliefs and Belief Updates

(A) Distribution of Prior Beliefs about Deaths in 6 Months, Full Sample



(B) Distribution of Belief Updates (Unsigned), Inconsistent vs. Consistent



Note: This figure presents estimates of the distribution of prior beliefs about the number of deaths in 6 months (Panel A) and the update magnitude $\ln(|\text{posterior} - \text{prior}| + 1)$ (Panel B). The gray shaded region in Panel A indicates the range of the projection of 100,000–240,000 deaths shown to all participants. Panel B shows the fraction who choose to update and mean update magnitude by treatment group, as well as the difference between groups. The mean update magnitudes are unconditional, i.e. they include people who do not update. Sample size of Consistent group: 933. Sample size of Inconsistent group: 967.

Formal tests, including stratum fixed effects, support this visual evidence (Table 2). In the full sample (Row 1), the Inconsistent treatment reduces participants' propensity to choose to update beliefs by 4.0 pp (SE: 2.3, $p = 0.075$). The treatment effect on the update propensity is 8.9% of the Consistent group's update propensity of 45.0%. Moreover, participants reduce the magnitude of their belief updates by 50.6 log points (SE: 24.9, $p = 0.043$). [22] Converting log points to percent changes, we find that the Inconsistent treatment reduces belief updates by about 39.7% (SE: 15.0). [23] [24]

Table 2: Effect of Inconsistent Treatment on Belief Updating

Sample:	(1) Update propensity: $1(\text{chooses to update})$	(2) Update magnitude: $\ln(\lvert\text{posterior} - \text{prior}\rvert + 1)$	(3) N
1. All	-0.040*	-0.506**	1,900
	(0.023)	(0.249)	
2. Priors below projection (optimists)	-0.019	-0.364	909
	(0.033)	(0.333)	
3. Priors at projection	-0.045	-0.419	544
	(0.041)	(0.448)	
4. Priors above projection (pessimists)	-0.072	-0.849	447
	(0.047)	(0.612)	
Inconsistent mean	0.407	4.32	967
Consistent mean	0.450	4.86	933

Note: This table presents treatment effect coefficients $\hat{\beta}$ from the Inconsistent treatment (Equation (1)) for the indicated outcome. All columns include stratum fixed effects. The sample in Row 2 comprises participants whose prior beliefs about deaths in 6 months (elicited before treatment) were between 0–99,999. The sample in Row 3 comprises participants whose prior beliefs were between 100,000–240,000 (i.e., at the level of the government projection). The sample in Row 4 comprises participants whose prior beliefs were 240,001 or more deaths. Stars present p-values: $*$ indicates $p < 0.1$; $**$ indicates $p < 0.05$; $***$ indicates $p < 0.01$. Treatment effects in Figure 1B differ slightly from those in this table because the table includes stratum fixed effects.

We provide a benchmark that suggests these treatment effects are large in magnitude. Intuitively, people who know about the federal COVID response should be less likely to update from our light-touch treatment. Indeed, people reporting above-median knowledge about the federal response are 4.2 pp less likely to update than people with below-median knowledge (SE: 2.3). The Inconsistent treatment effect (4.0) is

[22] If we limit the sample to study the update magnitude among the 814 participants who do update (the intensive margin), we find that the Inconsistent treatment reduces belief updates by 16.0 log points (SE: 15.5, p-value = 0.302).

[23] Note that $\exp(-0.506) - 1 \approx -0.397$; the log approximation to percent changes is imperfect for the magnitude of treatment effect we find. We obtain standard errors via the delta method.

[24] The magnitude of treatment effects on beliefs rises if we drop participants who fail either of two attention checks, rather than both (Appendix Table C.1). For example, the treatment effect on the update propensity rises to -4.7 pp (SE: 2.4, p-value: 0.049).

about equal to this difference. We find a similar result if we benchmark the update magnitude: people with above-median knowledge have update magnitudes that are 48.4 log points lower than people with below-median knowledge (SE: 25.2). This comparison illustrates that the Inconsistent treatment effect is sizeable, since baseline knowledge about the federal response is likely to influence belief updating.[25]

4.2 Mechanism: Government Credibility

We now present auxiliary evidence that supports reduced government credibility as the mechanism behind reduced belief updating.[26]

First, continuing to inspect Table 2, we find a larger reduction in updating among pessimists (who predict more deaths than the official projection) relative to optimists. For example, people with pessimistic priors tend to reduce their update magnitudes by 84.9% (SE: 61.2) and update propensity by 7.2 pp (SE: 4.7), both larger in magnitude than the corresponding point estimates for optimists (Row 4 vs. Row 2). We caveat that our estimates are imprecise; the two-sided p-values for a difference in Row 4 vs. Row 2 treatment effects are 0.478 (update magnitude) and 0.341 (update propensity).[27]

This heterogeneity serves as evidence against an alternative model, in which a combination of good and bad signals reduces belief updates without affecting government credibility. Holding credibility fixed, participants would form posterior beliefs purely by comparing signals to their priors. Participants whose prior beliefs lie between the initial "no crisis" and the recent "moderate crisis" signals should update less, since the signals partially cancel. On the other hand, subjects with prior beliefs that are more pessimistic than *both* signals would update *more* from the Inconsistent sequence, since it provides two downward signals compared to just one from the Consistent sequence. Instead, we find that very pessimistic subjects update *less* from the Inconsistent sequence (i.e., the point estimate is negative). Moreover, we even find that pessimists' reduction in updating is suggestively *larger* than for optimistic subjects.

Next, we provide evidence consistent with Implication 1a, which states that the ratio of belief updates

[25] This measure is elicited on a scale of 0–10, and the standard deviation is 2.9. Thus, the difference between above- and below-median knowledge is meaningful.

[26] Here, we mean lost *government* credibility *due to inconsistent signals*, rather than lost credibility due to the signals' conflicting with prior beliefs (non-credible signals).

[27] Even though Implications 1a and 1b are one-sided assertions that imply one-sided tests, we present results from two-sided tests, as is standard in experimental economics.

from the Inconsistent to the Consistent messages is largest for prior optimists. Intuitively, this ratio accounts for how surprising the new information is relative to priors. We consider the following specification without fixed effects:

$$y_j^d = \beta_n^d n_j + \beta_c^d + \varepsilon_j^d, \tag{2}$$

estimated within each group d of the prior beliefs distribution (where d indexes whether prior beliefs were below, at, or above the government projection).[28] We then examine the ratio of parameters $\frac{\hat{\beta}_n^d + \hat{\beta}_c^d}{\hat{\beta}_c^d}$ and inspect how this ratio varies with prior beliefs d. The numerator is the average belief update for the Inconsistent group, and the denominator is the average belief update for the Consistent group.

First, we obtain the ratio in the full sample (Table 3, Row 1); the ratio is less than 1 for both the update propensity (Column 1) and update magnitude (Column 2), which corroborates the earlier evidence that the Inconsistent treatment reduces belief updating. Note that inspecting $\frac{\hat{\beta}_n^d + \hat{\beta}_c^d}{\hat{\beta}_c^d}$ by groups of prior beliefs also addresses concerns that our main specification does not normalize belief updates. As discussed, we do not have enough information to determine the direction and magnitude in which a Bayesian updates. Reassuringly, the ratio of Inconsistent absolute belief updates to Consistent absolute belief updates is less than 1 within all the groups. Therefore, our central result holds when we account for prior beliefs: the Inconsistent treatment reduces the magnitude of belief updating.[29]

In support of Implication 1a, we find that the ratio of parameters is largest for optimists across both the update propensity and update magnitude (Table 3). We present the ratio for people with prior beliefs below (Row 2), at (Row 3), and above (Row 4) the government projection. Consistent with Implication 1a, the ratio of belief updates is largest for optimists, although formal tests are not powered to reject that the ratio among optimists is equal to moderates and pessimists. We find stronger evidence for Implication 1a if we limit to a test that optimists and pessimists have the same ratio (p-value: 0.214 for update propensity, 0.392 for update magnitude). We find no evidence for Implication 1b, since pessimists exhibit a larger reduction in the update ratio than moderates; however, our tests are imprecise, and tests of equality between the updating for moderates and optimists or pessimists have $p > 0.5$.

Given this suggestive heterogeneity that is consistent with the government credibility mechanism, we

[28] We omit fixed effects for simplicity in obtaining the average belief update for the control group.
[29] We find similar results if we form groups of prior beliefs by *deciles* (Appendix Figure B.2).

Table 3: Ratio of Inconsistent to Consistent Belief Updates: $\frac{\hat{\beta}_n^d + \hat{\beta}_c^d}{\hat{\beta}_c^d}$

Sample:	(1) Update propensity: $1(\text{chooses to update})$	(2) Update magnitude: $\ln(\lvert\text{posterior} - \text{prior}\rvert + 1)$	(3) N
1. All	0.905***	0.889***	1,900
	(0.048)	(0.049)	
2. Priors below projection (optimists)	0.955***	0.917***	909
	(0.063)	(0.063)	
3. Priors at projection	0.898***	0.915***	544
	(0.111)	(0.114)	
4. Priors above projection (pessimists)	0.814***	0.818***	447
	(0.094)	(0.097)	
5. Tests:			
Row 2 vs. Row 3	0.652	0.987	
Row 3 vs. Row 4	0.567	0.519	
Row 2 vs. Row 4	0.214	0.392	

Note: This table presents the ratio $\frac{\hat{\beta}_n^d + \hat{\beta}_c^d}{\hat{\beta}_c^d}$ from Equation (2). The sample in Row 2 comprises participants whose prior beliefs about deaths in 6 months (elicited before treatment) were between 0–99,999. The sample in Row 3 comprises participants whose prior beliefs were between 100,000–240,000 (i.e., at the level of the government projection). The sample in Row 4 comprises participants whose prior beliefs were 240,001 or more deaths. Stars present p-values: $*$ indicates $p < 0.1$; $**$ indicates $p < 0.05$; $***$ indicates $p < 0.01$. We obtain standard errors using the delta method and p-values in Row 5 using seemingly unrelated regression.

proceed to examine Implication 2 directly. We present the treatment effect on the changes in the posterior government credibility index (elicited after treatment), controlling for the prior government credibility index (elicited before treatment).

This approach yields our second main result: the Inconsistent treatment reduces self-reported beliefs in the government's credibility (Table 4). The Inconsistent treatment reduces government credibility by 0.037 SD (SE: 0.016, $p = 0.016$), which provides empirical support for Implication 2 (Column 1). We decompose the index into sub-scales in subsequent columns, and we find the largest effects on beliefs that the government is taking appropriate action (Column 3). We also present an auxiliary measure of interest — intention to vote for Trump — in which we find no effect (Column 6). Given that support for Trump may be sticky and affected by the administration's non-COVID policies, it is not surprising that we do not find an effect on this measure.

To interpret our results, we note that a 1 SD increase in the prior propensity to vote for Trump is associated with a 0.259 SD increase (SE: 0.011) in the prior government credibility index. As a result, the Inconsistent treatment has an effect on government trust that is about 14% as large as a 1 SD reduction in the propensity to vote for Trump. Given that faith in government is sticky, we interpret this effect as sizeable. Moreover, we note that trust in institutions has fallen in recent years and may be affected by economic conditions (Stevenson and Wolfers, 2011); as a result, it is not surprising that it is difficult to move in a light-touch experiment.

Table 4: Effect of Inconsistent Treatment on Government Credibility

	(1) Index	(2) Sub-scale	(3)	(4)	(5)	(6) Related outcome	(7)
Sample:	Government credibility index (columns 2–5)	Government's action is not strong enough	Government's action is appropriate	Government is over-reacting	Government has private info	Support President Trump	N
1. All	-0.037**	-0.005	-0.119***	0.038	0.004	-0.010	1,900
	(0.016)	(0.035)	(0.026)	(0.031)	(0.031)	(0.016)	
2. Priors below projection (optimists)	-0.022	-0.039	-0.096***	0.027	-0.005	0.004	909
	(0.022)	(0.047)	(0.036)	(0.047)	(0.047)	(0.025)	
3. Priors at projection	-0.075**	0.027	-0.208***	0.073	-0.010	-0.002	544
	(0.030)	(0.069)	(0.050)	(0.052)	(0.052)	(0.029)	
4. Priors above projection (pessimists)	-0.011	0.040	-0.046	0.042	0.082	-0.035	447
	(0.033)	(0.078)	(0.059)	(0.069)	(0.068)	(0.027)	
5. Tests:							
Row 2 vs. Row 3	0.150						
Row 3 vs. Row 4	0.143						
Row 2 vs. Row 4	0.775						

Note: This table presents treatment effect coefficients $\hat{\beta}$ from the Inconsistent treatment (Equation (1)) for the indicated outcome. All columns include strata fixed effects. The sample in Row 2 comprises participants whose prior beliefs about deaths in 6 months (elicited before treatment) were between 0–99,999 deaths. The sample in Row 3 comprises participants whose prior beliefs were between 100,000–240,000 (i.e., at the level of the government projection). The sample in Row 4 comprises participants whose prior beliefs were 240,001 or more deaths. The outcome in columns 2–6 is a standardized scale (in units of standard deviations) of post-treatment agreement with the given statement, controlling for prior (pre-treatment) agreement with the statement. We form the index in Column 1 by averaging Columns 2–5 (resigning outcomes such that a positive sign indicates government credibility). Column 6 presents posterior intention to vote for Donald Trump in 2020, controlling for prior intention to vote for Donald Trump in 2020. Stars present p-values: $*$ indicates $p < 0.1$; $**$ indicates $p < 0.05$; $***$ indicates $p < 0.01$. We obtain p-values in Row 5 using seemingly unrelated regression. Sample size of Consistent group: 933. Sample size of Inconsistent group: 967.

One may be concerned that reduced credibility is simply a direct effect of changes in beliefs. However, recall that the Inconsistent treatment *reduces* belief updates. Therefore, we might expect that the Inconsistent treatment also *reduces* changes in perceptions that the government is taking the appropriate action. Instead, our findings suggest that the Inconsistent treatment directly affects the government's credibility.

We find evidence generally in support of Implications 2a and 2b, which state that the credibility reduction is smallest among optimists and largest among moderates (Table 4, Rows 2–4). First, we find that optimists exhibit a smaller reduction in credibility than moderates, which adheres with Implication 2a. While optimists have a larger reduction in credibility than pessimists (Row 2 vs. Row 4), the point estimates are very similar and we cannot reject equality ($p = 0.775$). Most strikingly, we find that the reduction in credibility is concentrated among moderates (Row 2 vs. Row 3, $p = 0.150$ and Row 3 vs. Row 4, $p = 0.143$), exactly as Implication 2b predicts. While we are underpowered to lean heavily on tests for heterogeneity by prior beliefs, we show in Appendix A that this evidence is difficult to reconcile with other candidate mechanisms.

4.3 Additional Analyses

This section presents additional robustness checks and analyses discussed in Appendix B.

Robustness. We first document that our main outcomes are robust to the choice of controls or adding additional fixed effects. Second, we present robustness to using the variable $|\text{posteriors} - \text{priors}|$, without applying the log transform. A third concern is that, due to the large number of zeroes, the log transform may behave poorly. We show similar results if we use the inverse hyperbolic sine transform (Burbidge et al., 1988), or if we simply show the effects on $1(|\text{posterior} - \text{prior}| \leq c)$ for various c.

Continuous interaction. To test for treatment effect heterogeneity, we also try specifications with continuous interaction terms.

Heterogeneity by statement. We study the treatment effect heterogeneity by statement presented in the Inconsistent treatment and find modest evidence of heterogeneity.

Test of additional implication. Another implication of our framework is that people who have strong prior beliefs that the government's signal is either reliable or unreliable will have a smaller treatment effect from Inconsistent. We document that the treatment effects are concentrated among people who do not support Trump. We show that these participants do not have strong beliefs that the information is reliable, so this finding is consistent with the additional implication. This auxiliary result also adds to a growing literature on partisanship and COVID (Allcott et al., 2020; Barrios and Hochberg, 2020; Bursztyn et al., 2020a).

Additional outcomes. We study the treatment effect on other outcomes in our pre-registration. We give particular attention to effects on self-reported intent to social distance. In one of two specifications, the Inconsistent treatment increases self-reported intention to social distance in the future. The question does not ask whether participants will follow expert guidelines or official recommendations, so participants were not likely to interpret it as a question about adherence to official policies. As the Inconsistent treatment reduces faith in the government response, our interpretation is that private measures like intention to social distance are perceived as substitutes to a reliable public response. We present an exploratory IV strategy that lends suggestive evidence in favor of this interpretation.[30]

5 Conclusion

We use an experiment to provide empirical evidence for the concern that strong early positions from the federal government can reduce people's belief updates when presented with new guidance later on. We highlight a tradeoff between providing *consistent* and *transparent* messaging about the COVID crisis. Given that scientific knowledge about COVID has advanced quickly, taking a strong early stance has the possibility to cause confusion and reduce the government's credibility. Our work also has implications for the design of public-health guidance outside the COVID crisis: inconsistent recommendations about, for example, cancer screening or diet have the potential to reduce people's inferences in the future. However, while we find some heterogeneity that is suggestive of the inconsistency channel, we caveat that the heterogeneity is not statistically significant, and some of the heterogeneity runs counter to this channel. Moreover, we are underpowered to conduct joint hypothesis tests of all the implications we developed, which would provide the sharpest evidence in favor of this mechanism. Future work could further develop the theoretical framework to study when the planner optimally changes official guidance. It could also leverage experiments to more cleanly distinguish between the effects of inconsistent signals that change over time vs. non-credible signals that contradict people's prior beliefs.

[30]We relegate the social distancing analysis to the appendix because this exploratory outcome is self-reported (and may be especially subject to experimenter demand issues).

References

Ajzenman, Nicolás, Tiago Cavalcanti, and Daniel Da Mata, "More than Words: Leaders' Speech and Risky Behavior During a Pandemic," Technical Report August 2020.

Allcott, Hunt, Levi Boxell, Jacob Conway, Matthew Gentzkow, Michael Thaler, and David Yang, "Polarization and Public Health: Partisan Differences in Social Distancing during the Coronavirus Pandemic," *Journal of Public Economics*, 2020, *191* (104254).

Ananyev, Maxim, Dimitrios Xefteris, Galina Zudenkova, and Maria Petrova, "Information and Communication Technologies, Protests, and Censorship," Technical Report 2019.

Armantier, Olivier and Nicolas Treich, "Eliciting Beliefs: Proper Scoring Rules, Incentives, Stakes and Hedging," *European Economic Review*, August 2013, *62*, 17–40.

Banerjee, Abhijit, Esther Duflo, Amy Finkelstein, Lawrence F. Katz, Benjamin A. Olken, and Anja Sautmann, "In Praise of Moderation: Suggestions for the Scope and Use of Pre-Analysis Plans for RCTs in Economics," National Bureau of Economic Research Working Paper 26993, Cambridge, MA April 2020.

Barrios, John M. and Yael Hochberg, "Risk Perception through the Lens of Politics in the Time of the COVID-19 Pandemic," National Bureau of Economic Research Working Paper 27008, Cambridge, MA April 2020.

Benoit, William L., Mark J. Glanz, Anji L. Phillips, Leslie A. Rill, Corey B. Davis, Jayne R. Henson, and Leigh Anne Sudbrock, "Staying "On Message": Consistency in Content of Presidential Primary Campaign Messages Across Media," *American Behavior Scientist*, 2011, *55* (4), 457–468.

Binder, Carola, "Coronavirus Fears and Macroeconomic Expectations," *Review of Economics and Statistics*, 2020, *102* (4), 1–10.

Blinder, Alan S., "Monetary Policy by Committee: Why and How?," *European Journal of Political Economy*, March 2007, *23* (1), 106–123.

Burbidge, John B., Lonnie Magee, and A. Leslie Robb, "Alternative Transformations to Handle Extreme Values of the Dependent Variable," *Journal of the American Statistical Association*, 1988, *83* (401), 123–127.

Bursztyn, Leonardo, Aakaash Rao, Christopher Roth, and David Yanagizawa-Drott, "Misinformation During a Pandemic," National Bureau of Economic Research Working Paper 27417, Cambridge, MA September 2020.

_ , Ingar Haaland, Aakaash Rao, and Christopher Roth, "I Have Nothing Against Them, But. . .," National Bureau of Economic Research Working Paper 27288, Cambridge, MA May 2020.

Burton, Michael John, William J. Miller, and Daniel M. Shea, *Campaign Craft: The Strategies, Tactics, and Art of Political Campaign Management*, fifth ed., Santa Barbara, California: ABC-CLIO, 2015.

Camerer, Colin F., "The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List," December 2011.

Canadian Pandemic Influenza Preparedness Task Group and B. Henry, "Canadian Pandemic Influenza Preparedness: Public Health Measures Strategy," *Canadian Communicable Disease Report*, June 2019, *45* (6), 159–163.

Cantoni, Davide, David Y Yang, Noam Yuchtman, and Y Jane Zhang, "Protests as Strategic Games: Experimental Evidence from Hong Kong's Antiauthoritarian Movement," *The Quarterly Journal of Economics*, May 2019, *134* (2), 1021–1077.

Carpenter, Delesha M., Lorie L. Geryk, Annie T. Chen, Rebekah H. Nagler, Nathan F. Dieckmann, and Paul K. J. Han, "Conflicting Health Information: A Critical Research Need," *Health Expectations*, December 2016, *19* (6), 1173–1182.

Carville, James and Paul Begala, *Buck Up, Suck Up . . . and Come Back When You Foul Up: 12 Winning Secrets from the War Room*, New York, NY: Simon & Schuster, 2002.

Chahrour, Ryan, "Public Communication and Information Acquisition," *American Economic Journal: Macroeconomics*, July 2014, *6* (3), 73–101.

Charness, Gary, Uri Gneezy, and Brianna Halladay, "Experimental Methods: Pay One or Pay All," *Journal of Economic Behavior & Organization*, November 2016, *131*, 141–150.

Chen, Yuyu and David Y. Yang, "The Impact of Media Censorship: 1984 or Brave New World?," *American Economic Review*, June 2019, *109* (6), 2294–2332.

Clark, Danielle, Rebekah H Nagler, and Jeff Niederdeppe, "Confusion and Nutritional Backlash from News Media Exposure to Contradictory Information about Carbohydrates and Dietary Fats," *Public Health Nutrition*, December 2019, *22* (18), 3336–3348.

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber, "Does Policy Communication During Covid Work?," National Bureau of Economic Research Working Paper 27384, Cambridge, MA June 2020.

COVID Tracking Project, https://covidtracking.com/ 2020.

Crawford, Vincent P. and Joel Sobel, "Strategic Information Transmission," *Econometrica*, November 1982, *50* (6), 1431–1451.

Cullen, Zoe and Ricardo Perez-Truglia, "The Salary Taboo: Privacy Norms and the Diffusion of Information," National Bureau of Economic Research Working Paper 25145, Cambridge, MA February 2020.

de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth, "Measuring and Bounding Experimenter Demand," *American Economic Review*, November 2018, *108* (11), 3266–3302.

_ , Lise Vesterlund, and Alistair J. Wilson, "Experimenter Demand Effects," in Arthur Schram and Aljaž Ule, eds., *Handbook of Research Methods and Applications in Experimental Economics*, Edward Elgar Publishing, 2019, pp. 384–400.

Doherty, David, Conor M. Dowling, and Michael G. Miller, "When Is Changing Policy Positions Costly for Politicians? Experimental Evidence," *Political Behavior*, June 2016, *38* (2), 455–484.

Duhigg, Charles, "Seattle's Leaders Let Scientists Take the Lead. New York's Did Not," *The New Yorker*, April 2020.

Edmond, Chris, "Information Manipulation, Coordination, and Regime Change," *Review of Economic Studies*, October 2013, *80* (4), 1422–1458.

Einav, Liran, Amy Finkelstein, Tamar Oostrom, Abigail Ostriker, and Heidi Williams, "Screening and Selection: The Case of Mammograms," *American Economic Review*, Forthcoming.

Fetzer, Thiemo, Lukas Hensel, Johannes Hermle, and Christopher Roth, "Coronavirus Perceptions And Economic Anxiety," *Review of Economics and Statistics*, Forthcoming.

Fink, Sheri, "White House Takes New Line After Dire Report on Death Toll," *The New York Times*, Mar 2020.

Gächter, Simon and Elke Renner, "The Effects of (Incentivized) Belief Elicitation in Public Goods Experiments," *Experimental Economics*, September 2010, *13* (3), 364–377.

Gehlbach, Scott and Konstantin Sonin, "Government Control of the Media," *Journal of Public Economics*, October 2014, *118*, 163–171.

_ , _ , and Milan W. Svolik, "Formal Models of Nondemocratic Politics," *Annual Review of Political Science*, May 2016, *19* (1), 565–584.

Gentzkow, Matthew and Jesse M. Shapiro, "Media Bias and Reputation," *Journal of Political Economy*, 2006, *114* (2), 280–316.

Ghanem, Dalia, Sarojini Hirshleifer, and Karen Ortiz-Becerra, "Testing Attrition Bias in Field Experiments," Technical Report March 2020.

Gollust, Sarah E., Rebekah H. Nagler, and Erika Franklin Fowler, "The Emergence of COVID-19 in the U.S.: A Public Health and Political Communication Crisis," *Journal of Health Politics, Policy and Law*, May 2020, p. 8641506.

Grossman, Guy, Soojong Kim, Jonah Rexer, and Harsha Thirumurthy, "Political Partisanship Influences Behavioral Responses to Governors' Recommendations for COVID-19 Prevention in the United States," *Proceedings of the National Academy of Sciences*, September 2020, *117* (39), 24144–24153.

Grynbaum, Michael M., "Trump's Briefings Are a Ratings Hit. Should Networks Cover Them Live?," *The New York Times*, March 2020.

Gutierrez, Emilio, Adrian Rubli, and Tiago Tavares, "Information and Behavioral Responses during a Pandemic: Evidence from Delays in COVID-19 Death Reports," *Covid Economics, Vetted and Real-Time Papers*, July 2020, *37*, 100–140.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart, "Designing Information Provision Experiments," Technical Report June 2020.

Henningsen, Arne and Jeff D. Hamann, "Systemfit: A Package for Estimating Systems of Simultaneous Equations in R," *Journal of Statistical Software*, 2008, *23* (4).

Hollard, Guillaume, Sébastien Massoni, and Jean-Christophe Vergnaud, "In Search of Good Probability Assessors: An Experimental Comparison of Elicitation Rules for Confidence Judgments," *Theory and Decision*, March 2016, *80* (3), 363–387.

Hollyer, James R., B. Peter Rosendorff, and James Raymond Vreeland, "Transparency, Protest, and Autocratic Instability," *American Political Science Review*, November 2015, *109* (4), 764–784.

Hossain, Tanjim and Ryo Okui, "The Binarized Scoring Rule," *The Review of Economic Studies*, July 2013, *80* (3), 984–1001.

Huang, Haifeng, "Propaganda as Signaling," *Comparative Politics*, July 2015, *47* (4), 419–444.

Iachan, Felipe S. and Plamen T. Nenov, "Information Quality and Crises in Regime-Change Games," *Journal of Economic Theory*, July 2015, *158*, 739–768.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 2007, *75* (1), 83–119.

Little, Andrew T., "Propaganda and Credulity," *Games and Economic Behavior*, March 2017, *102*, 224–232.

Morris, Stephen and Hyun Song Shin, "Social Value of Public Information," *The American Economic Review*, 2002, *92* (5), 1521–1534.

— **and** — , "Optimal Communication," *Journal of the European Economic Association*, 2007, *5* (2-3), 594–602.

Nagler, Rebekah H., "Adverse Outcomes Associated With Media Exposure to Contradictory Nutrition Messages," *Journal of Health Communication*, January 2014, *19* (1), 24–40.

Oster, Emily, "Expert Behavior Change in Response to Best Practice Changes: Evidence from Obstetrics," Technical Report 2018.

— , "Health Recommendations and Selection in Health Behaviors," *American Economic Review: Insights*, June 2020, *2* (2), 143–160.

Painter, Marcus and Tian Qiu, "Political Beliefs Affect Compliance with COVID-19 Social Distancing Orders," Technical Report 2020.

Perloff, Richard M., *Political Communication: Politics, Press, and Public in America*, Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., 1998.

Schlag, Karl H., James Tremewan, and Joël J. van der Weele, "A Penny for Your Thoughts: A Survey of Methods for Eliciting Beliefs," *Experimental Economics*, 2015, *18* (3), 457–490.

Schotter, Andrew and Isabel Trevino, "Belief Elicitation in the Laboratory," *Annual Review of Economics*, August 2014, *6* (1), 103–128.

Shadmehr, Mehdi and Dan Bernhardt, "State Censorship," *American Economic Journal: Microeconomics*, May 2015, *7* (2), 280–307.

Sigelman, Lee and Carol K. Sigelman, "Shattered Expectations: Public Responses to "Out-of-Character" Presidential Actions," *Political Behavior*, 1986, *8* (3), 262–286.

Sonnemans, Joep and Theo Offerman, "Is the Quadratic Scoring Rule Really Incentive Compatible?," December 2001, p. 19.

Stevenson, Betsey and Justin Wolfers, "Trust in Public Institutions over the Business Cycle," *American Economic Review Papers and Proceedings*, 2011, *101* (3), 281–287.

Stimson, Brie, "How long will coronavirus last in the US?," *Fox News*, Mar 2020.

Tan, Andy S.L., Chul joo Lee, Rebekah H. Nagler, and Cabral A. Bigman, "To Vape or Not to Vape? Effects of Exposure to Conflicting News Headlines on Beliefs about Harms and Benefits of Electronic Cigarette Use: Results from a Randomized Controlled Experiment," *Preventive Medicine*, December 2017, *105*, 97–103.

Tomz, Michael and Robert P. Van Houweling, "Candidate Repositioning," Technical Report 2012.

_ **and** _ , "Political Repositioning: A Conjoint Analysis," Technical Report April 2016.

Trautmann, Stefan T. and Gijs van de Kuilen, "Belief Elicitation: A Horse Race among Truth Serums," *The Economic Journal*, December 2015, *125* (589), 2116–2135.

Wood, Thomas and Ethan Porter, "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence," *Political Behavior*, March 2019, *41* (1), 135–163.

Young, Alwyn, "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results," *The Quarterly Journal of Economics*, May 2019, *134* (2), 557–598.

Appendix Materials

A Theory Appendix

This Appendix develops the theoretical framework. We begin with a simple model with two states of the world to develop intuition. Then, we develop a model with three states of the world that yields the implications discussed in Section 2.

A.1 Two-State Model Setup

States of the world. The state of the world $\omega_t \in \{0,1\}$ evolves stochastically over two periods: $t \in \{0,1\}$. The high state $\omega_t = 1$ indicates a crisis. The low state $\omega_t = 0$ indicates normalcy. States persist over time with probability α and randomly transition to the other state with probability $1-\alpha$, where $\alpha \in (\frac{1}{2}, 1)$.

Prior beliefs about states of the world. An agent holds a prior beliefs density μ_0 about ω_0, so that $\mu_0 = \mathbb{P}(\omega_0 = 1)$. The state transition process induces prior beliefs μ_1 about ω_1, such that:

$$\mu_1 = \alpha\mu_0 + (1-\alpha)(1-\mu_0)$$

Government communications. At the beginning of each period the government receives a private signal $s_t \in \{0,1\}$ about the true state. The government's signals are either always accurate (for all t, $s_t = \omega_t$, denoted $A = 1$), or always pure noise (for all t, $s_t \sim \text{Uniform}(0,1)$, denoted $A = 0$). Agents do not observe A, but have the prior belief $\mathbb{P}(A = 1) = q$. We assume $0 < q < 1$.[31] The government's report $r_t \in \{\varnothing, 0, 1\}$ either contains no information ($r_t = \varnothing$) or truthfully publicizes the government's signal ($r_t = s_t$).

Agent's inference. In period 1, upon observing $\mathbf{r} = \{r_0, r_1\}$, agents make inferences using Bayesian reasoning. The mere presence or absence of a report is uninformative. Agents simply use reports as noisy signals, taking into account their prior beliefs over the states of the world and government's information accuracy. We denote by $\mu_1^{\mathbf{r}}$ the posterior belief about the state ω_1 upon observing reports $\mathbf{r}$. We denote by $q^{\mathbf{r}}$ the posterior belief about the government's signal's informativeness A.

A.2 Inconsistent Government Reports

We use this model to study the impact of an inconsistent sequence of reports from the government, as opposed to consistent reports. To fix ideas and match our empirical setting, we focus on comparing beliefs upon observing two sequences of reports:

$$\text{Inconsistent} : \mathbf{n} = \{r_0 = 0, r_1 = 1\}$$

$$\text{Consistent} : \mathbf{c} = \{r_0 = \varnothing, r_1 = 1\}.$$

That is, holding the report about period 1 constant ($r_1 = 1$), we study the effect of having publicized an inconsistent signal in the prior period ($r_0 = 0$) vs. saying nothing ($r_1 = \varnothing$).[32]

Belief updates. Given that $r_1 = 1$, for $\mathbf{r} \in \{\mathbf{n}, \mathbf{c}\}$, we can write the posterior beliefs as:

$$\mu_1^{\mathbf{r}} = 1 \cdot q^{\mathbf{r}} + \mu_1 \cdot (1 - q^{\mathbf{r}}). \tag{A.1}$$

See Appendix A.4 for proofs. This equation states that $\mu_1^{\mathbf{r}}$ is a convex combination of 1 and μ_1 where the weights depend on the posterior belief that government reports (and thus in particular r_1) are accurate. In this simple model, r_0 only affects beliefs updating about the crisis in period 1 through a credibility channel.[33]

Proposition 1. ***Effect of inconsistent government reports on government credibility and belief updating.***

[31] Degenerate cases $q = 0$ and $q = 1$ are less interesting: people ignore public signals entirely or fully update to certain beliefs about the state of the world.

[32] We define the Consistent signal as communicating $r_0 = \varnothing$ (rather than, e.g., $r_0 = 1$) because this is precisely what we implement in the experiment, so this permits a tight connection between the theoretical predictions and the experimental design.

[33] This is a consequence of assuming that government signals are either pure noise or fully accurate. Relaxing this assumption would lead to an additional role for r_0 whereby a previous optimistic report carries over some reassurance over the present. Our empirical results suggest that in our application, the credibility effects that we model dominates, as subjects with pessimistic prior beliefs tend to be made more pessimistic upon observing a previous optimistic report.

1. *If $\mu_0 > 1-\alpha$, the Inconsistent signal reduces belief updating about the crisis:*

$$\mu_1^{\mathbf{n}} - \mu_1 < \mu_1^{\mathbf{c}} - \mu_1.$$

2. *The ratio of belief updating from the Inconsistent sequence to belief updating from the Consistent signal $\frac{\mu_1^{\mathbf{n}}-\mu_1}{\mu_1^{\mathbf{c}}-\mu_1}$ is **decreasing** in μ_0.*

3. *If $\mu_0 > 1-\alpha$, the Inconsistent signal reduces posterior credibility of the government signal:*

$$q^{\mathbf{n}} < q^{\mathbf{c}}.$$

4. *The credibility loss from the Inconsistent signal $-(q^{\mathbf{n}} - q^{\mathbf{c}})$ is **increasing** in μ_0.*

Proposition 1 states that for individuals with $\mu_0 > 1-\alpha$, observing a sequence of inconsistent signals reduces beliefs about severity of the crisis and reduces the posterior beliefs in the accuracy of the government's signal. Intuitively, inconsistent government reports are more likely to be the outcome of noisy communication since the state of the world is sticky. Thus, the model formalizes the intuition that governments that communicate inconsistent information do not appear credible. The condition $\mu_0 > 1-\alpha$ captures that very optimistic individuals (low μ_0) find the inconsistent sequence more credible, as it suggests the state has truly changed and the government got that change right.

Appendix Figure A.1 illustrates Proposition 1. We choose representative values of $\alpha = 0.9$ and $q = 0.75$ for exposition. The x-axis of each panel plots prior probabilities μ_0; a higher μ_0 indicates more pessimism.[34] Panel A plots posterior beliefs after observing Consistent and Inconsistent signals ($\mu_1^{\mathbf{c}}$ and $\mu_1^{\mathbf{n}}$) in the blue curve and the orange curves, respectively. Panel A illustrates Proposition 1, item 1: the blue curve lies above the orange curve for all prior beliefs $\mu_0 > \alpha$. Panel B shows that the ratio of belief updates is monotonically decreasing in prior beliefs (Proposition 1, item 2). Panel C presents the mechanism, highlighted in Proposition 1, items 3 and 4: the agent exposed to Consistent signals exhibits higher posterior credibility in the government for $\mu_0 > 1-\alpha$, and the difference between the agents with Consistent vs. Inconsistent signals rises monotonically with μ_0.

A.3 Three-State Model

In this section, we extend our model to allow for three states of the world, thereby allowing for nuance about how bad the crisis is. This version of the model is less intuitive but closer to our empirical application. We discuss the implications of this three-state model in Section 2.

The three-states model has several advantages. First, this model allows for the agent to be more pessimistic than the signal sent by the government in period 1. In our empirical context, many participants had prior beliefs that were more pessimistic than the government's information. Second, the three-state model has a formal implication that arises from government *inconsistency* and not from government *optimism* alone.

States of the world. $\omega_t \in \{0, 1, 2\}$ evolves stochastically over two periods: $t \in \{0, 1\}$. The high state $\omega_t = 2$ indicates a severe crisis, the low state $\omega_t = 0$ indicates normalcy, and the middle state $\omega_t = 1$ models a moderate crisis. States persist over time with probability α and randomly transition to either other state with probability $\frac{1-\alpha}{2}$,[35] where $\alpha \in (\frac{1}{2}, 1)$ captures the state persistence between the two periods.

Prior beliefs about states of the world. An agent holds a prior beliefs density μ_0 about ω_0, so that $\mu_0(i) = \mathbb{P}(\omega_0 = i)$ for $i = 0, 1, 2$, with $\mu_0(0) + \mu_0(1) + \mu_0(2) = 1$. The state transition process induces a prior beliefs distribution μ_1 over ω_1, such that, for $i \in \{0, 1, 2\}$,

$$\mu_1(i) = \alpha\mu_0(i) + \frac{1-\alpha}{2}(1 - \mu_0(i)).$$

[34] Note that pessimism about period 0 translates to pessimism about period 1 via the state evolution equation.

[35] Symmetric transitions to other states is a simplifying assumption to lighten the model exposition. Predictions are essentially unchanged if we allow for transition probabilities to differ between pairs of states.

Government communications. At the beginning of each period the government receives a private signal $s_t \in \{0, 1, 2\}$ about the true state. The government's signals are either always accurate (for all t, $s_t = \omega_t$, denoted $A = 1$), or always pure noise (for all t, $s_t \sim \text{Uniform}(0, 1, 2)$, denoted $A = 0$). Agents do not observe A, and they have the prior belief $\mathbb{P}(A = 1) = q$. The government's report $r_t \in \{\emptyset, 0, 1, 2\}$ either contains no information ($r_t = \emptyset$) or truthfully publicizes the government's signal ($r_t = s_t$).

Agent's inference. In period 1, upon observing r_0 and r_1, agents make inferences using Bayesian reasoning. The mere presence or absence of a report is uninformative. Agents simply use reports as noisy signals, taking into account their prior beliefs over the states of the world and government's information accuracy. We denote by $\mu_1^{r_0,r_1}$ (resp. q^{r_0,r_1}) the posterior distribution over the state ω_1 (resp. over government's signals informativeness A) upon observing reports r_0, r_1.

We use this model to study the impact of a changing sequence of reports from the government, as opposed to consistent reports. To fix ideas and match our empirical setting, we focus on comparing beliefs upon observing two sequences of reports:

$$\text{Inconsistent} : \{r_0 = 0, r_1 = 1\}$$

$$\text{Consistent} : \{r_0 = \emptyset, r_1 = 1\}$$

That is, holding the latest report constant on a "moderate crisis" report, we study the effect of having publicized an inconsistent signal in the last period, here an optimistic "normalcy" signal.

Belief updates. Given that $r_1 = 1$, we can write the posterior beliefs as:

$$\mu_1^{r_0,1}(1) = q^{r_0,1} + \mu_1(1)(1 - q^{r_0,1}) \tag{A.2}$$

$$\mu_1^{r_0,1}(i) = \mu_1(i)(1 - q^{r_0,1}), \text{ for } i = 0, 2 \tag{A.3}$$

That is, $\mu_1^{r_0,1}(i)$ is a convex combination of $\mathbb{1}(r_1 = i)$ and $\mu_1(i)$ where the weights depend on the posterior belief that government reports (and thus in particular r_1) are accurate.

First, we obtain a result similar to Item 1 of Proposition 1, which posits that for most agents except the very optimistic ones, inconsistent information reduces belief updating.

Proposition 2. ***Effect of inconsistent government reports on government credibility and belief updating with 3 states.*** *A necessary and sufficient condition on prior beliefs μ_0 for the Inconsistent signal to hurt the credibility of the government and reduce belief updating about the crisis is the following:*

$$\alpha\mu_0(1) + \frac{1}{2}\mu_0(2) \geq \frac{1}{3} \tag{A.4}$$

If and only if A.4 holds, then:

$$\begin{aligned} q^{\mathbf{n}} &< q^{\mathbf{c}} \\ \mu_1^{\mathbf{n}}(1) - \mu_1(1) &< \mu_1^{\mathbf{c}}(1) - \mu_1(1) \\ \mu_1^{\mathbf{n}}(i) - \mu_1(i) &> \mu_1^{\mathbf{c}}(i) - \mu_1(i) \text{ for } i = 0, 2 \end{aligned}$$

In this framework, by reduced belief updating, we mean that posterior beliefs have less mass on the state of the world corresponding to the signal (here $\omega_1 = 1$) and more mass on the other states of the world.

Proposition 2 shows that for sufficiently pessimistic individuals, Inconsistent signals hurt government credibility and reduce belief updating. Proposition 2 implies Implications 1 and 2 in Section 2. For high values of α, Inequality A.4 will be verified for most prior beliefs distributions.

Note that this prediction is in stark contrast with another intuitive model where all signals are used on an equal footing and informative about the current state. In such a model, very pessimistic individuals would update more from the Inconsistent sequence, which from their perspective contains two signals that the world is not as bad as they thought.

Next, we derive a comparative statics result that is analogous to Item 4 of Proposition 1. Since there are three states, comparative statics with respect to prior beliefs μ_0 cannot be derived on one dimension. We develop comparative statics along one-dimensional subspaces of the belief distribution, by holding mass on one state of the world fixed.

Proposition 3. ***Comparative statics of Inconsistent vs. Consistent*** *Consider a non-degenerate prior distribution μ_0, i.e. such that $\mu_0(i) > 0$ for $i = 0, 1, 2$.*

1. *Holding constant $\mu_0(2)$, the credibility loss from the Inconsistent signal $-(q^n - q^c)$ is **increasing** in $\mu_0(1)$.*
2. *Holding constant $\mu_0(1)$, the credibility loss from the Inconsistent signal $-(q^n - q^c)$ is **increasing** in $\mu_0(2)$.*
3. *Holding constant $\mu_0(0)$, the credibility loss from the Inconsistent signal $-(q^n - q^c)$ is **increasing** in $\mu_0(1)$.*

Items 1 and 2 imply that credibility loss from the Inconsistent sequence relative to the Consistent sequence is larger for more pessimistic beliefs if this pessimism is fueled by reduction in $\mu_0(0)$, the belief that there is no crisis. Put another way, the most optimistic people have the smallest reduction in credibility loss. On the other hand, Item 3 shows that the credibility loss is largest for individuals whose prior belief distribution puts more mass on a moderate crisis, which is the second message of the government. Intuitively, very pessimistic individuals already lose some confidence in the government when seeing the Consistent sequence with a moderate crisis report. By contrast, moderates find the Consistent signal to be very credible. Their trust falls relatively more when they see an Inconsistent sequence.[36] The ratio of belief updates follows the same comparative statics because $\frac{\mu_1^{0,1}(i) - \mu_1(i)}{\mu_1^{\varnothing,1}(i) - \mu_1(i)} = \frac{q^{0,1}}{q^{\varnothing,1}}$ for $i \in \{0, 1, 2\}$, from Equations (A.2) and (A.3). As a result, Proposition 3 yields Implications 1a, 1b, 2a, and 2b in Section 2.

A.3.1 Distinguishing Inconsistency

In our empirical application, the treatment presents an early optimistic report $r_0 = 0$. One may be concerned that our empirical analysis simply identifies the effect of providing an early optimistic signal ($r_0 = 0$ vs. $r_0 = \varnothing$), since both Inconsistent and Consistent groups receive $r_1 = 1$. As shown in Gentzkow and Shapiro (2006), people have a preferences for like-minded news, so disagreement in prior beliefs with the early optimistic positions of Trump might in itself drive the results. While we cannot reject that part of our result might reflect this effect, we now show that an empirical prediction of the model (Item 3 of Proposition 3) cannot be generated if $r_1 = \varnothing$ or $r_1 = 0$.

Put another way, our experiment compares $\{\varnothing, r_1\}$ to $\{0, r_1\}$ for $r_1 = 1$. We now show that comparing these sequences does not yield the same patterns if $r_1 \in \{\varnothing, 0\}$. Thus, we show that the theory yields a testable prediction about heterogeneity if r_0 *contradicts* r_1, rather than if r_1 and r_0 do not contradict. In practice, we lack power fully to distinguish the heterogeneity, but the theory admits a clear test.[37]

To formalize this point, we derive two propositions analogous to Proposition 3 that give comparative statics of the effect of showing $\{r_0 = 0\}$ relative to $\{r_0 = \varnothing\}$, if followed by either $r_1 = \varnothing$ (Proposition 4) or $r_1 = 0$ (Proposition 5).

Proposition 4. ***Comparative statics of*** $\{r_0 = 0, r_1 = \varnothing\}$ ***vs.*** $\{r_0 = \varnothing, r_1 = \varnothing\}$ *Consider a non-degenerate prior distribution μ_0, i.e. such that $\mu_0(i) > 0$ for $i = 0, 1, 2$.*

1. *Holding constant $\mu_0(2)$, the credibility loss $-(q^{0,\varnothing} - q^{\varnothing,\varnothing})$ is **increasing** in $\mu_0(1)$.*
2. *Holding constant $\mu_0(1)$, the credibility loss $-(q^{0,\varnothing} - q^{\varnothing,\varnothing})$ is **increasing** in $\mu_0(2)$.*
3. *Holding constant $\mu_0(0)$, the credibility loss $-(q^{0,\varnothing} - q^{\varnothing,\varnothing})$ is **constant** in $\mu_0(1)$.*

Proposition 5. ***Comparative statics of*** $\{r_0 = 0, r_1 = 0\}$ ***vs.*** $\{r_0 = \varnothing, r_1 = 0\}$ *Consider a non-degenerate prior distribution μ_0, i.e. such that $\mu_0(i) > 0$ for $i = 0, 1, 2$.*

1. *Holding constant $\mu_0(2)$, the credibility loss $-(q^{0,0} - q^{\varnothing,0})$ **varies ambiguously** with respect to $\mu_0(1)$.*
2. *Holding constant $\mu_0(1)$, the credibility loss $-(q^{0,0} - q^{\varnothing,0})$ **varies ambiguously** with respect to $\mu_0(2)$.*

[36] We thank an anonymous referee for pointing out this intuition which led us to include this proposition.

[37] We do not need to distinguish comparative statics in the case where $r_1 = 2$ for both signals. That case would also correspond to an Inconsistent sequence.

3. *Holding constant $\mu_0(0)$, the credibility loss $-(q^{0,0} - q^{\varnothing,0})$ is **constant** in $\mu_0(1)$.*

Comparing Item 3 from Propositions 3 to Item 3 of Propositions 4 and 5, we obtain a prediction allowing us to test whether we are merely identifying the effects of showing Early Optimistic vs. None (i.e, $\{0, \varnothing\}$ vs. $\{\varnothing, \varnothing\}$) or Optimistic vs. Late Optimistic (i.e., $\{0, 0\}$ vs. $\{\varnothing, 0\}$), rather than Inconsistent vs. Consistent signals.[38] Relative to the empirics, only Inconsistent signals yield the non-monotonic treatment effects we observe in Table 4, where participants *at* the government information (moderates) have the largest reduction in posterior government credibility.

While the above discussion shows comparative statics that crucially depend on the later signal being a moderate signal ($r_1 = 1$), we note an important caveat. The comparative statics that we find would be qualitatively unchanged if we were instead to compare an isolated downplaying report $\{r = 0\}$ to an isolated moderate report $\{r = 1\}$.[39] Our analysis focuses on inconsistency (comparing different sequences of two signals) rather than downplaying (comparing one isolated signal) because we find it a more natural way to model our intervention, which does contain two signals. Even so, we cannot rule out the possibility that respondents pool the signals together or only pay attention to the downplaying signal when shown the Inconsistent sequence.[40]

We highlight the difference between Proposition 3 and Propositions 4 and 5 in Appendix Figure A.2. Comparing Inconsistent to Consistent sequences, the reductions in belief update exhibit an inverted-U shape (black line): moderates have the largest reduction in belief updates, followed by pessimists and then optimists. On the other hand, comparing Early Optimistic vs. None ($r_1 = \varnothing$) or Optimistic vs. Late Optimistic ($r_1 = 0$) does not yield this inverted-U shape (green and pink lines). In particular, for both alternates of r_1, the pessimists and moderates exhibit identical reductions in belief updates. For Inconsistent vs. Consistent signals, the black line is decreasing between the center and right of the figure; for the other signals, it is flat. On the other hand, following the caveat above, we do find a similar inverted-U shape in a one-signal model comparing $\{r = 1\}$ to $\{r = 0\}$ (blue line).

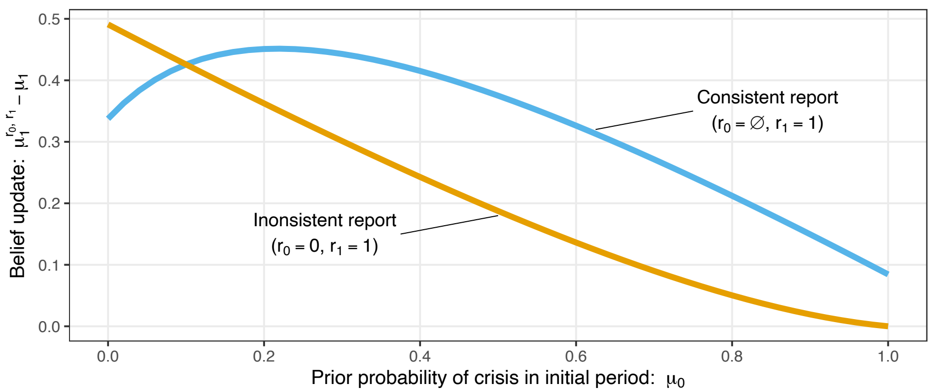
[38]In addition, one can intuitively see that Item 3 of Propositions 4 and 5 would be even more starkly different from Proposition 3, had we introduced more structure in the beliefs space by making state 0 more likely for moderates. The credibility loss would then be **decreasing** in $\mu_0(1)$. This would be a reasonable assumption but we abstract away from it for simplicity and tractability.

[39]We can also embed this set-up in our model as comparing the sequence $\{\varnothing, 0\}$ to $\{\varnothing, 1\}$.

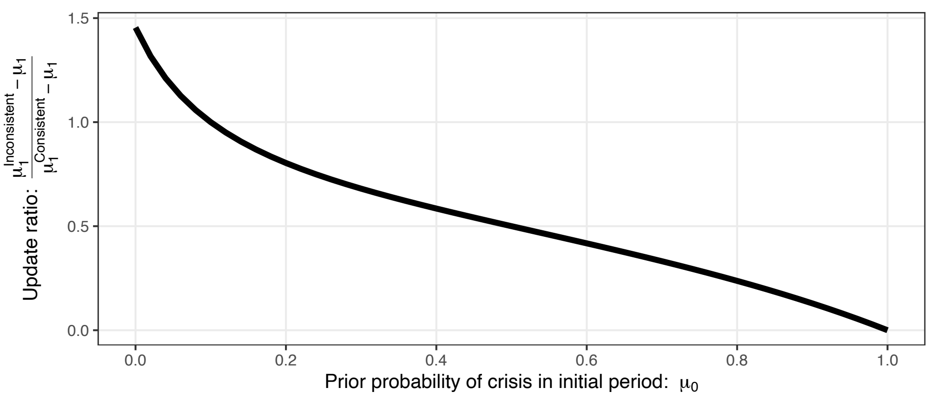
[40]We do note one qualitative difference between the curves is that optimists have an increased trust in the government in the model with one signal. We do not find evidence of this empirically. Even so, we note two caveats. First, we are underpowered to reject an increased trust in government among prior optimists. Second, we may not be focusing on a sample of people who are truly optimistic relative to the signals we provide.

Figure A.1: Posterior Beliefs About Future Crisis After Observing Current Policy Response (Two-State Model)

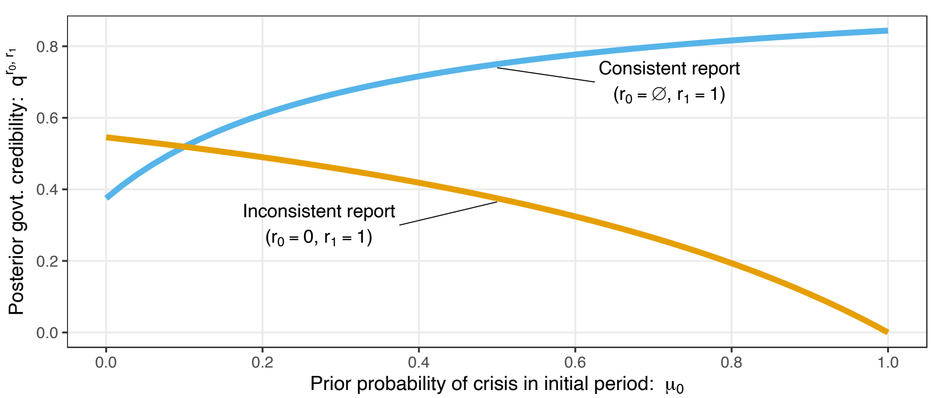
(A) Belief Update



(B) Ratio of Belief Updates

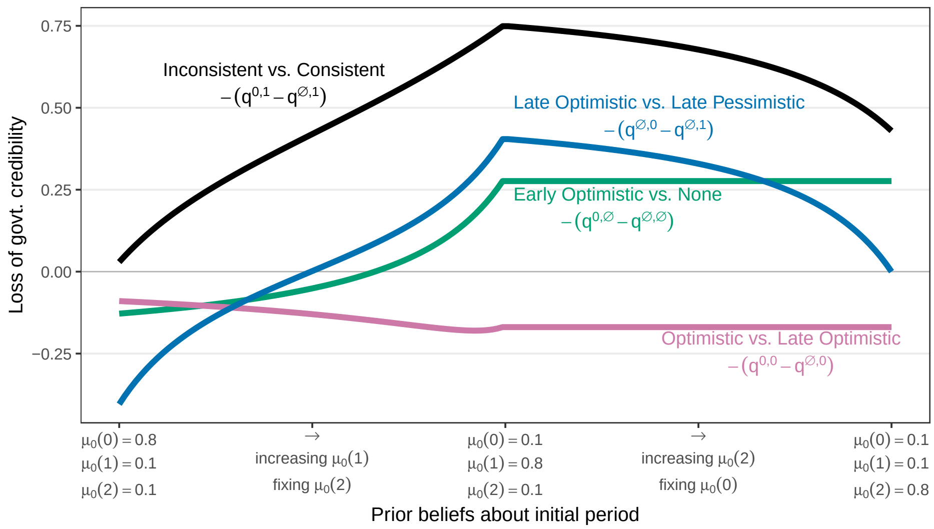


(C) Posterior Government Credibility



Note: The three panels of this figure illustrate Proposition 1, holding fixed the state persistence parameter ($\alpha = 0.9$) and prior government credibility ($q = 0.75$) for ease of visualization. All three panels plot the prior belief about crisis, μ_0, on the x-axis. Panel A demonstrates the relationship between belief updates and priors beliefs. The blue curve plots the update after observing a Consistent report, and the orange curve plots the update after observing an Inconsistent report. When beliefs are sufficiently pessimistic (i.e. $\mu_0 > \alpha$), the update after observing the Inconsistent Signal is smaller than the update after observing a Consistent Signal. Panel B demonstrates that the ratio of these updates is monotonically decreasing in the prior belief. Panel C demonstrates the model's mechanism, plotting the government's credibility after Inconsistent and Consistent reports with the orange and blue curves, respectively. The difference in credibility $q^{\text{Inconsistent}} - q^{\text{Consistent}}$ is decreasing in the prior belief, and for sufficiently pessimistic priors ($\mu_0 > \alpha$) the government is less credible after an Inconsistent report than a Consistent one.

Figure A.2: Loss in Government Credibility After Various Reports (Three-State Model)



Note: This figure illustrates the key comparative statics in Propositions 2, 3 and 4. It holds fixed the state persistence parameter ($\alpha = 0.9$) and prior government credibility ($q = 0.75$) for ease of visualization. The x-axis plots prior beliefs about the initial state, shifting probability mass from the no crisis state to the moderate crisis state, and then from the moderate crisis state to the severe crisis state. This is one way to capture a monotonic increase in the expected severity of crisis; that is, the transition from optimist to moderate, and moderate to pessimist. The black curve plots the loss in government credibility from sending an Inconsistent report ($r_0 = 0, r_1 = 1$) versus a Consistent report ($r_0 = \varnothing, r_1 = 1$). The blue curve plots the loss in government credibility from sending a Late Optimistic Report ($r_0 = \varnothing$, $r_1 = 0$) versus a Late Pessimistic Report Late Optimistic Report ($r_0 = \varnothing$, $r_1 = 1$). The green curve plots the loss in government credibility from sending an initial, isolated Optimistic report ($r_0 = 0, r_1 = \varnothing$) versus sending no report ($r_0 = r_1 = \varnothing$). The pink curve plots the loss in government credibility from sending an Early Optimistic report ($r_0 = r_1 = 0$) versus a Late Optimistic report ($r_0 = \varnothing, r_1 = 0$).

A.3.2 Discussion and Caveats

Our theoretical framework has the advantage of delivering sharp testable predictions that we can take to the data. On the other hand, we acknowledge that it imposes several restrictive assumptions. First, we do not incorporate strategic motives on the part of the government. For example, the government may have a motive to report biased results. Alternatively, the government may withhold results until after elections. To generate predictions about belief updates with strategic concerns, we would need assumptions about agents' awareness of and responsiveness to such concerns. Assuming full rationality, we could describe Trump's communication as a cheap talk game. The classical result from Crawford and Sobel (1982) would then predict a multiplicity of equilibria, including a babbling equilibrium where no information is conveyed, and therefore does not easily generate testable predictions. Additionally, people may not engage in such complex equilibrium thinking when interpreting Presidential communication. Our approach might be described as modeling agents as 1-level thinkers, who do not take into account the government's motive to persuade them.

Second, extensions could enrich the informational environment to permit signals from the media. In our empirical exercise, the existence of the media should push against our ability to find a treatment effect of providing information, since the information we provide was widely circulated by the media already.

Finally, a richer set-up could involve heterogeneity in α or other parameters. We have abstracted away from such heterogeneity since we do not measure them in our empirical application.

A.4 Proofs

A.4.1 Derivation of Equation (A.1)

We apply the Law of Total Probability, conditioning on events $A = 1$ and $A = 0$:

$$\begin{aligned}\mu_1^{\mathbf{r}} =& \mathbb{P}(\omega_1 = 1|r_0, r_1)\\ =& \mathbb{P}(\omega_1 = 1|s_0, s_1)\\ =& \mathbb{P}(\omega_1 = 1|A = 1, s_0, s_1)\mathbb{P}(A = 1|s_0, s_1) + \mathbb{P}(\omega_1 = 1|A = 0, s_0, s_1)(1 - \mathbb{P}(A = 1|s_0, s_1))\\ =& \mathbb{P}(\omega_1 = 1|A = 1, s_0, s_1)\mathbb{P}(A = 1|s_0, s_1) + \mu_1(1 - \mathbb{P}(A = 1|s_0, s_1))\end{aligned}$$

We focus on reports with $r_1 = s_1 = 1$, and given that $\mathbb{P}(\omega_1 = 1|A = 1, s_0, s_1 = 1) = 1$, we have:

$$\mu_1^{\mathbf{r}} = q^{\mathbf{r}} + \mu_1(1 - q^{\mathbf{r}})$$

A.4.2 Proof of Proposition 1

Part 1: expressing posterior beliefs about government accuracy. We begin by expressing posterior beliefs about accuracy of the government signals in terms of prior beliefs μ_0 and q. First, by assumption agents learn nothing from the government withholding a report, so that $\mathbb{P}(A = 1|r_i = \varnothing, \dots) = \mathbb{P}(A = 1|\dots)$. Second, the conditional distribution of non-null reports on high quality signals is just the unconditional distribution of the corresponding states. In particular: $\mathbb{P}(r_0, r_1|A = 1, r_0 \neq \varnothing, r_1 \neq \varnothing) = \mathbb{P}(s_0 = 0, s_1 = 1|A = 1) = \mathbb{P}(\omega_0 = 0, \omega_1 = 1)$. Finally, the conditional distribution of non-null reports on low quality signals $\mathbb{P}(r_0, r_1|A = 0, r_0 \neq \varnothing, r_1 \neq \varnothing) = \mathbb{P}(s_0 = 0, s_1 = 1|A = 0) = \frac{1}{4}$.

Then, applying Bayes' rule we have:

$$\begin{aligned}q^{\mathbf{c}} &= \mathbb{P}(A = 1|r_1 = 1)\\ &= \mathbb{P}(A = 1|s_1 = 1)\\ &= \frac{\mathbb{P}(s_1 = 1|A = 1)\mathbb{P}(A = 1)}{\mathbb{P}(s_1 = 1)}\\ &= \frac{\mathbb{P}(\omega_1 = 1|A = 1)\mathbb{P}(A = 1)}{\mathbb{P}(s_1 = 1)}\\ &= \frac{[(1-\alpha)(1-\mu_0) + \mu_0\alpha]q}{[(1-\alpha)(1-\mu_0) + \mu_0\alpha]q + \frac{1}{2}(1-q)}\\ q^{\mathbf{c}} &= \frac{(1-\alpha)(1-\mu_0)q + \mu_0\alpha q}{(1-\alpha)(1-\mu_0)q + \frac{1}{2}(1-q) + \mu_0\alpha}\end{aligned} \tag{A.5}$$

$$\begin{aligned}q^{\mathbf{n}} &= \mathbb{P}(A = 1|r_0 = 0, r_1 = 1)\\ &= \mathbb{P}(A = 1|s_0 = 0, s_1 = 1)\\ &= \frac{\mathbb{P}(s_0 = 0, s_1 = 1|A = 1)\mathbb{P}(A = 1)}{\mathbb{P}(s_0 = 0, s_1 = 1)}\\ &= \frac{\mathbb{P}(\omega_0 = 0, \omega_1 = 1)\mathbb{P}(A = 1)}{\mathbb{P}(s_0 = 0, s_1 = 1)}\\ q^{\mathbf{n}} &= \frac{(1-\mu_0)(1-\alpha)q}{(1-\mu_0)(1-\alpha)q + \frac{1}{4}(1-q)}\end{aligned} \tag{A.6}$$

Part 2: Monotonicity of $q^{\mathbf{n}} - q^{\mathbf{c}}$ with respect to μ_0. We partially differentiate Equations (A.5) and (A.6) with respect to μ_0:

$$\frac{\partial q^{\mathbf{c}}}{\partial \mu_0} = \frac{\frac{1}{4}q(1-q)(2\alpha-1)}{(q(\alpha(4\mu_0-2) - 2\mu_0 + 1) + 1)^2} > 0 \qquad \frac{\partial q^{\mathbf{n}}}{\partial \mu_0} = -\frac{\frac{1}{2}q(1-q)(1-\alpha)}{(q(4\alpha(\mu_0-1) - 4\mu_0 + 3) + 1)^2} < 0.$$

Hence, $q^{\mathbf{c}}$ is increasing in μ_0 and $q^{\mathbf{n}}$ is decreasing in μ_0. This proves Item 4. of Proposition 1.

Part 3: Proof of Items 1 and 3. Solving for μ_0 towards equality of the RHS of Equations (A.5) and (A.6), under the assumption that $q \in (0, 1)$, we obtain a **unique** solution, proving item 3:

$$q^{\mathbf{c}} = q^{\mathbf{n}} \Leftrightarrow \mu_0 = 1 - \alpha.$$

Now, applying the monotonicity of $q^{\mathbf{c}} - q^{\mathbf{n}}$ in μ_0, we get:

$$\forall \mu_0 < 1-\alpha, q^{\mathbf{c}} - q^{\mathbf{n}} < 0 \qquad \text{and} \qquad \forall \mu_0 > 1-\alpha, q^{\mathbf{c}} - q^{\mathbf{n}} > 0$$

Next, denote $\Delta\hat{\mu}_1 := \mu_1^{r_0=\varnothing, r_1=1} - \mu_1^{r_0=0, r_1=1}$, and analogously $\Delta\hat{q}$. We now prove that $\Delta\hat{\mu}_1 > 0$ for $\mu_0 > 1-\alpha$. Using Equation A.1 twice and subtracting, we obtain the following:

$$\Delta\hat{\mu}_1 = \Delta\hat{q}(1-\mu_1)$$

Since

$$\forall \mu_0 > 1-\alpha, \Delta\hat{q} > 0,$$

the proof of Item 1 immediately follows with

$$\forall \mu_0 > 1-\alpha, \Delta\hat{\mu}_1 > 0.$$

Part 4: Proof of Item 2. Writing Equation A.1 for both sequence of signals, subtracting μ_1 from both and taking the ratio, we immediately obtain that:

$$\frac{\mu_1^{\mathbf{n}} - \mu_1}{\mu_1^{\mathbf{c}} - \mu_1} = \frac{q^{\mathbf{n}}}{q^{\mathbf{c}}}$$

We have seen that

$$\frac{\partial q^{\mathbf{c}}}{\partial \mu_0} > 0 \qquad\qquad \frac{\partial q^{\mathbf{n}}}{\partial \mu_0} < 0$$

The proof of item 2 immediately follows with

$$\frac{\partial \frac{\mu_1^{\mathbf{n}} - \mu_1}{\mu_1^{\mathbf{c}} - \mu_1}}{\partial \mu_0} < 0.$$

A.4.3 Proof of Proposition 2

First, using Equation A.2, we obtain that:

$$\left(\mu_1^{\mathbf{n}}(1) - \mu_1(1)\right) - \left(\mu_1^{\mathbf{c}}(1) - \mu_1(1)\right) = \left(1 - \mu_1(1)\right)\left(q^{\mathbf{n}} - q^{\mathbf{c}}\right)$$

Similarly using Equation A.3:

$$\left(\mu_1^{\mathbf{n}}(i) - \mu_1(i)\right) - \left(\mu_1^{\mathbf{c}}(i) - \mu_1(i)\right) = -\mu_1(i)\left(q^{\mathbf{n}} - q^{\mathbf{c}}\right).$$

This shows that the three inequalities we wish to obtain are equivalent. Therefore, we just need to prove the proposition for the first inequality. We thus turn to expressing posterior beliefs about accuracy of the government signals in terms of prior beliefs μ_0 and q. We follow a similar process as in deriving Equations (A.5) and (A.6). in this three state setup, the conditional distribution of non-null reports on low quality signals is now $\mathbb{P}(r_0, r_1 | A = 0, r_0 \neq \varnothing, r_1 \neq \varnothing) = \mathbb{P}(s_0 = 0, s_1 = 1 | A = 0) = \frac{1}{9}$. Then, applying Bayes' rule we have:

$$\begin{aligned}
q^{\mathbf{c}} &= \mathbb{P}(A = 1 | r_1 = 1) = \mathbb{P}(A = 1 | s_1 = 1) \\
&= \frac{\mathbb{P}(s_1 = 1 | A = 1)\mathbb{P}(A = 1)}{\mathbb{P}(s_1 = 1)} \\
&= \frac{\mu_1(1) q}{\frac{1}{3}(1-q) + \mu_1(1) q} \\
&= \frac{(1-\mu_0(1))\frac{1-\alpha}{2} q + \mu_0(1)\alpha q}{(1-\mu_0(1))\frac{1-\alpha}{2} q + \frac{1}{3}(1-q) + \mu_0(1)\alpha q}
\end{aligned}$$

$$\begin{aligned}
q^{\mathbf{n}} &= \frac{\mathbb{P}(s_0 = 0, s_1 = 1 | A = 1)\mathbb{P}(A = 1)}{\mathbb{P}(s_0 = 0, s_1 = 1)} \\
&= \frac{\mathbb{P}(\omega_0 = 0, \omega_1 = 1)\mathbb{P}(A = 1)}{\mathbb{P}(s_0 = 0, s_1 = 1)} \\
&= \frac{\mu_0(0)\frac{1-\alpha}{2}q}{\mu_0(0)\frac{1-\alpha}{2}q + \frac{1}{9}(1-q)}
\end{aligned}$$

We then solve for $q^{\mathbf{n}} < q^{\mathbf{c}}$ in terms of the prior beliefs:

$$\begin{aligned}
q^{\mathbf{n}} < q^{\mathbf{c}} \Leftrightarrow \frac{q^{\mathbf{n}}}{q^{\mathbf{c}}} < 1 &\Leftrightarrow \frac{q + \frac{\frac{2}{3}(1-q)}{2\mu_0(1)\alpha + (1-\mu_0(1))(1-\alpha)}}{q + \frac{\frac{1}{9}(1-q)}{\frac{1-\alpha}{2}\mu_0(0)}} < 1 \\
&\Leftrightarrow \mu_0(0) \leq \frac{1}{3} + \frac{3\alpha - 1}{1-\alpha}\mu_0(1) \text{ , assuming } q < 1 \\
&\Leftrightarrow 2\alpha\mu_0(1) + \mu_0(2) \geq \frac{2}{3}, \text{ using that } \mu_0(0) + \mu_0(1) + \mu_0(2) = 1.
\end{aligned}$$

A.4.4 Proof of Proposition 3

We start by writing down $\frac{q^n}{q^c}$ in a simple form, in terms of $\mu_0(0)$ and $\mu_0(1)$:

$$\frac{q^n}{q^c} = \frac{\frac{9q}{1-q} + \frac{3}{\mu_0(1)(2\alpha-1) + \frac{1-\alpha}{2}}}{\frac{9q}{1-q} + \frac{2}{\mu_0(0)(1-\alpha)}}$$

We can take partial derivatives with respect to $\mu_0(0)$ (resp. $\mu_0(1)$), holding constant $\mu_0(1)$ (resp. $\mu_0(0)$):

$$\begin{aligned}
\frac{\partial \frac{q^n}{q^c}}{\partial \mu_0(0)}|_{\mu_0(1)} &= \frac{\left(\frac{9q}{1-q} + \frac{3}{\mu_0(1)(2\alpha-1) + \frac{1-\alpha}{2}}\right)\frac{2(1-\alpha)}{(\mu_0(0)(1-\alpha))^2}}{\left(\frac{9q}{1-q} + \frac{2}{\mu_0(0)(1-\alpha)}\right)^2} > 0 \\
\frac{\partial \frac{q^n}{q^c}}{\partial \mu_0(1)}|_{\mu_0(0)} &= -\frac{\frac{3(2\alpha-1)}{(\mu_0(1)(2\alpha-1) + \frac{1-\alpha}{2})^2}}{\frac{9q}{1-q} + \frac{2}{\mu_0(0)(1-\alpha)}} < 0
\end{aligned}$$

$\frac{\partial \frac{q^n}{q^c}}{\partial \mu_0(0)}|_{\mu_0(1)} > 0$ shows that holding $\mu_0(1)$ constant, the credibility loss $-(q^n - q^c)$ from the Inconsistent signal increases when $\mu_0(0)$ decreases, i.e. when $\mu_0(2)$ increases. This is item (ii) of Proposition 3. Similarly, $\frac{\partial \frac{q^n}{q^c}}{\partial \mu_0(1)}|_{\mu_0(0)} < 0$ shows item (iii) of Proposition 3.

To show item (i), we rewrite $\frac{q^n}{q^c}$ in terms of $\mu_0(0)$ and $\mu_0(2)$:

$$\frac{q^n}{q^c} = \frac{\frac{9q}{1-q} + \frac{3}{(1-\mu_0(0)-\mu_0(2))(2\alpha-1) + \frac{1-\alpha}{2}}}{\frac{9q}{1-q} + \frac{2}{\mu_0(0)(1-\alpha)}}$$

Then, taking:

$$\frac{\partial \frac{q^n}{q^c}}{\partial \mu_0(0)}|_{\mu_0(2)} = \frac{\frac{3(2\alpha-1)}{\left((1-\mu_0(0)-\mu_0(2))(2\alpha-1) + \frac{1-\alpha}{2}\right)^2}\left(\frac{9q}{1-q} + \frac{2}{\mu_0(0)-\alpha}\right) + \frac{2}{(1-\alpha)\mu_0(0)^2}\left(\frac{9q}{1-q} + \frac{3}{(1-\mu_0(0-\mu_0(2))(2\alpha-1) + \frac{1-\alpha}{2}}\right)}{\left(\frac{9q}{1-q} + \frac{2}{\mu_0(0)(1-\alpha)}\right)^2} > 0$$

We obtain $\frac{\partial \frac{q^n}{q^c}}{\partial \mu_0(0)}|_{\mu_0(2)} > 0$, which shows item (i).

A.4.5 Proof of Proposition 4

We start by deriving $q^{0,\varnothing}$:

$$\begin{aligned} q^{0,\varnothing} &= \frac{\mathbb{P}(s_0 = 0|A = 1)\mathbb{P}(A = 1)}{\mathbb{P}(s_0 = 0)} \\ &= \frac{\mathbb{P}(\omega_0 = 0)\mathbb{P}(A = 1)}{\mathbb{P}(s_0 = 0)} \\ &= \frac{\mu_0(0)q}{\mu_0(0)q + \frac{1}{3}(1-q)} \end{aligned}$$

Thus we have $\frac{q^{0,\varnothing}}{q} = \frac{\mu_0(0)}{\mu_0(0)q+\frac{1}{3}(1-q)}$. We immediately see that $\frac{q^{0,\varnothing}}{q}$ does not depend on $\mu_0(1)$ and $\mu_0(2)$ (other than via $\mu_0(0)$). This gives us item 3. of Proposition 4. Then, to obtain items 1. and 2., we simply derive:

$$\frac{\partial \frac{q^{0,\varnothing}}{q}}{\partial \mu_0(0)}|_{\mu_0(1)} = \frac{\partial \frac{q^{0,\varnothing}}{q}}{\partial \mu_0(0)}|_{\mu_0(2)} = \frac{\frac{1}{3}(1-q)}{\left(\mu_0(0)q + \frac{1}{3}(1-q)\right)^2} > 0$$

A.4.6 Proof of Proposition 5

Similarly, we derive $q^{0,0}$ and $q^{\varnothing,0}$:

$$\begin{aligned} q^{0,0} &= \frac{\mathbb{P}(s_0 = 0, s_1 = 0|A = 1)\mathbb{P}(A = 1)}{\mathbb{P}(s_0 = 0, s_1 = 0)} \\ &= \frac{\mathbb{P}(\omega_0 = 0, \omega_1 = 0)\mathbb{P}(A = 1)}{\mathbb{P}(s_0 = 0, s_1 = 0)} \\ &= \frac{\mu_0(0)\alpha q}{\mu_0(0)\alpha q + \frac{1}{9}(1-q)} \end{aligned}$$

$$\begin{aligned} q^{\varnothing,0} &= \frac{\mathbb{P}(s_1 = 0|A = 1)\mathbb{P}(A = 1)}{\mathbb{P}(s_1 = 0)} \\ &= \frac{\mathbb{P}(\omega_1 = 0)\mathbb{P}(A = 1)}{\mathbb{P}(s_1 = 0)} \\ &= \frac{\mu_0(0)(2\alpha - 1)q + \frac{1-\alpha}{2}q}{\mu_0(0)(2\alpha - 1)q + \frac{1-\alpha}{2}q + \frac{1}{3}(1-q)} \end{aligned}$$

Then taking the difference:

$$-(q^{0,0} - q^{\varnothing,0}) = \frac{\frac{1}{9}(1-q)}{\mu_0(0)\alpha q + \frac{1}{9}(1-q)} - \frac{\frac{1}{3}(1-q)}{\mu_0(0)(2\alpha - 1)q + \frac{1-\alpha}{2}q + \frac{1}{3}(1-q)}$$

We immediately see that $-(q^{0,0} - q^{\varnothing,0})$ does not depend on $\mu_0(1)$ and $\mu_0(2)$ (other than via $\mu_0(0)$). This gives us item 3. of Proposition 5. Then, to obtain items 1. and 2., we simply derive:

$$\frac{\partial - (q^{0,0} - q^{\varnothing,0})}{\partial \mu_0(0)}|_{\mu_0(1)} = \frac{\partial - (q^{0,0} - q^{\varnothing,0})}{\partial \mu_0(0)}|_{\mu_0(2)} = -\frac{\frac{1}{9}\alpha q(1-q)}{\left(\mu_0(0)\alpha q + \frac{1}{9}(1-q)\right)^2} + \frac{\frac{1}{3}(2\alpha - 1)q(1-q)}{\left(\mu_0(0)(2\alpha - 1)q + \frac{1-\alpha}{2}q + \frac{1}{3}(1-q)\right)^2}$$

which can be positive or negative depending on the parameters and the point value of $\mu_0(0)$.

B Robustness Appendix

B.1 Robustness

This section details several robustness exercises. First, we show robustness to adding or removing controls to our main specification. Second, we show that the main result of reduced belief updating from the Inconsistent treatment also obtains when we use another natural outcome variable, the absolute belief update (without applying log transform). Third, we present the results if we use the inverse hyperbolic sine transform instead of the log. Fourth, we present the results if we show effects on $1(|\text{posteriors} - \text{priors}| \leq c)$ for various c. Fifth, we show the ratio of belief updates by *decile* of the prior beliefs distribution. Sixth, we analyze heterogeneous treatment effects from each version of the statements.

B.1.1 Specification robustness

We focus on the outcomes of the update magnitude ($\ln(|\text{posterior} - \text{prior}| + 1)$) and update propensity ($1(\text{chooses to update})$). We also present p-values from permutation tests with 1,000 permutations (Young, 2019). Appendix Table B.1 presents the results; Columns 1–3 show the results for the update magnitude, and Columns 4–6 show the results for the update propensity. Column 1 and 4 include no fixed effects and therefore simply compare the means in the Inconsistent vs. Consistent group. Columns 2 and 5 include only the stratum fixed effects that we employ in our main specification, so they are identical to Table 2. Columns 3 and 6 additionally include demographic fixed effects for all demographic variables we collected: race/ethnicity, education, household income, and U.S. region.[41] Appendix Table B.1 shows that the point estimates are reasonably stable. Relative to our primary specification which includes stratum fixed effects, the update magnitude moves from 50.6 log points to 48.5 log points, and the update propensity moves from 4.02 pp to 3.85 pp once we add demographic fixed effects. All specifications, including with the additional controls, are significant at $\alpha = 0.1$, using either p-values from the t-test with robust standard errors or from the permutation test. Moreover, the permutation p-values are significant at $p < 0.05$ in all specifications when we study the update magnitude (which is better powered than the update propensity, since it incorporates changes in the intensive margin of updating). To summarize, the specifications with demographic fixed effects appear very similar to the specifications with stratum fixed effects.

B.1.2 Different measurement of belief updates

In Appendix Table B.2, we show a version of Table 2 that presents effects on absolute belief updates without applying the log transform. We pre-registered our primary measurements of the number of deaths using logs as a principled method of adjusting for outliers. About 2% of the sample has changes in beliefs of 1,000,000 deaths or more, while 57% of the sample has identical prior and posterior beliefs. Nevertheless, for completeness, we show the result and sensitivity using the outcome $|\text{posterior} - \text{prior}|$, without taking logarithms.

For clarity, we emphasize the average effect and omit heterogeneity by prior beliefs. In Column 1, we present the effect on the magnitude of belief update, without converting to logs (i.e., the outcome is $|\text{posterior} - \text{prior}|$). We find an extremely noisy estimate (the standard error exceeds 900,000) that is statistically indistinguishable from 0.

In Columns 2 and 3, we winsorize the changes outcome variable at the 95th and 90th percentiles for the Consistent group. Winsorizing is another strategy to handle outliers, but it has the potential to introduce bias by capping the outcome. As expected, we recover a negative coefficient of ($p = 0.114$ if we winsorize at the 95th percentile and $p = 0.051$ for the 90th percentile). In Columns 3 and 4, we drop people whose prior or posterior beliefs exceed the 95th or 90th percentiles of the Consistent group's distribution. The rationale for these specifications is that people with extremely high prior or posterior beliefs may simply be entering large numbers without thinking carefully about the question. Nevertheless, these columns select on a condition for intermediate variables that enter the dependent variable, so they are merely suggestive.

[41] As discussed in Appendix D, there are 18 strata based on political party, age group, and gender. We aggregate race/ethnicity into: White, Black, Hispanic, and Asian/Other/Missing. We aggregate household incomes into: missing/$30,000 or less; 30,000–100,000; 100,000 or more. We aggregate education groups into: missing/high school or less, some college, college or more. The region variable provided by Lucid.io is: North, South, East, West.

Even so, they both recover negative and significant point estimates; for example, dropping the sample that exceeds the 90th percentile, we obtain $\hat{\beta} = -7,407$ (SE: 3,034, $p = 0.015$). Together, the results from Appendix Table B.2 show that the main result of reduced belief updating holds if we use an outcome variable that is not transformed into logs, after making appropriate adjustments for outliers.

B.1.3 Other alternatives to the log transform

More than 60% of the sample does not update their prior beliefs. As a result, one may be concerned that applying the log transform $\ln(|\text{posterior} - \text{prior}| + 1)$ is not appropriate due to the large number of zeros.

We conduct several tests. First, we show the effects on a similar transform, the inverse hyperbolic sine (IHS), which is similar to the log transform but permits the data to be valued at 0 (Burbidge et al., 1988). Appendix Table B.3 presents the results, which are very similar to those in Table 2, Column 1. In addition, moving across columns of Appendix Table B.3, we still find the treatment effect is concentrated among pessimists.

Second, a standard concern about employing the log transform is which constant to add when forming update magnitude. To adhere to standard practice, we add the constant 1 (i.e., we study ln(|posterior − prior + 1)). We document that this choice of constant yields conservative results in our setting. We vary the constant we choose to form the outcome variable $\ln(|\text{posterior} - \text{prior}| + c)$, for $c \in \{1, .1, .01, .001\}$, in Appendix Table B.4. Column 1 is identical to our primary specification. This table documents that for smaller c, the effect on treatment rises; intuitively, these specifications place more weight on the smaller differences in update propensity between the treatments. Notably, across all specifications, we still find $p < 0.1$, and the magnitude of the treatment effect rises. As a result, our primary specification is conservative.

Third, in Figure B.1 we present the treatment effects on $1(|\text{posterior} - \text{prior}| \leq c)$ for various values of c. This test has the advantage of providing a transparent portrait of how the treatment affects belief updates. As with our update propensity measure $1(\text{chooses to update})$ this test has less power than our log outcome measure because it does not weight changes by the magnitude of the outcome. We find that the Inconsistent treatment increases the propensity for belief updates to be less than c, for a large number of c.

Altogether, we prefer using the log transform because it is familiar, easily interpretable, and what we pre-registered. Moreover, in all our main estimates, we find similar effects on the update propensity, which is an unambiguous variable. Finally, this subsection documents that reasonable alternatives to the log transform yield very similar results.

B.1.4 Continuous Interaction

Implications 1a, 1b, 2a, and 2b suggest that the Inconsistent treatment effects depend on prior beliefs. In the text, we simply compare treatment effects among people below, at, or above the government information. Alternatively, we could obtain estimates of treatment effect heterogeneity using the following formulation:

$$y_j = \beta n_j + \gamma_1 \ln(\text{prior deaths} + 1)_j + \gamma_2 \left(\ln(\text{prior deaths} + 1)_j \times n_j\right) + \boldsymbol{X}_j \delta + \varepsilon_j, \tag{B.1}$$

where $\ln(\text{prior deaths} + 1)$ is the participant's prior beliefs about the number of deaths from COVID, and the interaction coefficient γ_2 is the marginal effect of pessimism on the outcome.

We prefer formulation in the body of the paper for the following reasons:

1. The theory makes predictions about non-monotonic treatment effect heterogeneity (Implications 1b and 2b)

2. Showing treatment effect differences within subgroup does not require a parametric assumption that the additional effect of pessimism is additively linear in $\ln(\text{prior deaths} + 1)$.

On the other hand, the specification in Equation (B.1) may be better powered to conduct formal hypothesis tests (because it leverages the parametric assumption).[42]

For completeness, we present the results of Equation (B.1) (Appendix Table B.5). We find that increasing pessimism, as measured by log prior beliefs about the number of deaths, is associated with a reduced

[42]We thank a referee for the suggestion to conduct this test.

propensity to update (Column 1) and reduced update magnitude (Column 2), as predicted by Implication 1a and demonstrated non-parametrically in Table 2. Given the log specification, doubling the prior beliefs about deaths is associated with a 1.2 pp reduced propensity to update (SE: 0.01, p-value of interaction: 0.31). Similarly, doubling the prior beliefs about deaths is associated with an 11.6 log point reduction in update magnitudes (SE: 14.0, p-value 0.41), or about 11.0%. We also study the effect on the posterior government credibility index (Column 3); as in the body of paper, this specification additionally includes a control for the prior government credibility index. We find that the linear in logs specification (Equation B.1) yields a negative point estimate for the government credibility index, although the magnitude is small.

However, as in Table 4 we conclude that the small magnitude of the point estimate in Column 3 is driven by the tail of the distribution. In Columns 4–6, we limit the distribution to people whose prior beliefs about deaths are less than 1,000,000, dropping the 138 participants with the most pessimistic prior beliefs. The results for the update propensity and update magnitude are consistent with the results for the full sample (Columns 4 and 5). The result for the government credibility index grows in magnitude; doubling the prior beliefs about deaths is associated with a .011 SD reduction in the government credibility index (SE: 0.01, p-value: 0.24). Altogether, the parametric specification in Equation (B.1) yields point estimates that are suggestively consistent with Implications 1a and 2a, but we remain underpowered to provide formal evidence in favor of them.

B.1.5 Ratio by Decile of Prior Beliefs

The text presents the ratio of belief updates, splitting the sample into three groups (priors below, at, or above the government information). The advantage of our approach in the text is that it permits hypothesis tests of Implications 1a and 1b. On the other hand, the three groups presented in the text are coarse, so forming the ratio within finer groups may better account for the belief update in the control group. We therefore estimate a version of Equation (2) where d corresponds to the decile of prior beliefs. We find results that are consistent with the body of the paper (Appendix Figure B.2). The ratio is less than 1 for most deciles across both the update magnitude (Panel A) and update propensity (Panel B): even accounting for how people in the control group update their prior beliefs, the Inconsistent treatment reduces belief updating. Moreover, the ratio is largest for optimists and decreases with pessimism (as predicted by Implication 1a), but our estimates are imprecise when we only use deciles of prior beliefs.

B.1.6 Heterogeneity by Statement

This section studies treatment effect heterogeneity by the statement indicating an Inconsistent position. We randomize the order in which we present the statements, and we elicit the belief update before presenting the second statement. This design permits a natural test of whether the treatment effect for Inconsistent positions about the flu differs from the treatment effect for Inconsistent positions about social distancing as follows. We estimate:

$$y_j = \beta_{\text{flu}} 1(\text{Inconsistent and flu})_j + \beta_{\text{distancing}} 1(\text{Inconsistent and distancing})_j + \boldsymbol{X}_j \delta + \varepsilon_j, \qquad \text{(B.2)}$$

where $1(\text{Inconsistent and flu})$ and $1(\text{Inconsistent and distancing})$ denote that the participant was in the Inconsistent group and saw the flu or distancing statement before the posterior belief elicitation. We then conduct tests of $H_0 : \hat{\beta}_{\text{flu}} = \hat{\beta}_{\text{distancing}}$. In particular, the regressions in Equation (B.2) decompose the effect of Inconsistent treatment into the effect of seeing the flu statement first or the distancing statement first. The treatment effects $\hat{\beta}$ presented in the text will be a weighted average of $\hat{\beta}_{\text{flu}}$ and $\hat{\beta}_{\text{distancing}}$.[43]

Similarly, we can present the analogous treatment effect of presenting a Consistent position about the flu relative to an Inconsistent position about social distancing; our design does not guarantee these are symmetric. We employ a similar regression as Equation (B.2):

$$y_j = \gamma_{\text{flu}} 1(\text{Consistent and flu})_j + \gamma_{\text{distancing}} 1(\text{Consistent and distancing})_j + \boldsymbol{X}_j \delta + \varepsilon_j, \qquad \text{(B.3)}$$

[43] The coefficient $\hat{\beta}$ is a *weighted* average of $\hat{\beta}_{\text{flu}}$ and $\hat{\beta}_{\text{distancing}}$ because not precisely 50% of the final sample sees each statement first, due to attrition.

but $1(\text{Consistent and flu})$ and $1(\text{Consistent and distancing})$ denote that the participant was in the *Consistent* group and saw the consistent flu or distancing statement before the posterior belief elicitation. In this case, the coefficients from Equation (B.3) will generally have the opposite sign as the coefficients from Equation (B.2). The weighted average of $\hat{\gamma}_{\text{flu}}$ and $\hat{\gamma}_{\text{distancing}}$ will be identical to $-\hat{\beta}$ from Equation (1).

This exercise has two goals. First, if both statements yield treatment effects that are consistent with the theory, that suggests that *inconsistency*, rather than some other feature of the statement, drives the results.[44] Second, we can examine if one of the statements has a larger effect than the other.

Both statements generate belief updates that are consistent with the theory, although one specification shows a significantly stronger effect for the social distancing statement than the flu statement (Appendix Table B.6). Columns 1 and 2 present the treatment effects on the update magnitude $\ln(|\text{posterior} - \text{prior}| + 1)$ and Columns 3 and 4 present the treatment effect on the update propensity $1(\text{chooses to update})$. In all columns, we find that the point estimate of any individual Inconsistent or Consistent statement directionally aligns with the main results in Table 2. Column 1 shows that the treatment effects of both Inconsistent statements are negative; a test for the difference between the treatment effects gives $p = 0.417$. Column 2 shows that the Consistent statement about distancing gives a larger belief update than the Consistent statement about the flu ($p = 0.054$), but the belief updates are positive for both statements. We find no evidence of treatment effect heterogeneity in Columns 3 and 4, and both statements reduce belief updating in the Inconsistent treatment and increase belief updating in the Consistent group. Overall, we find that both statements cause effects in the expected direction, and we find only modest evidence of treatment effect heterogeneity. Informally, we note that the social distancing statement appears to be more inconsistent than the changing position regarding how fatal COVID is, which could drive the larger effect in the one specification. The modest heterogeneity we do find appears larger for Consistent statements rather than Inconsistent statements (Columns 2 and 4 vs. 1 and 3). Perhaps the early statements in the Inconsistent treatment were more likely to be common knowledge and therefore attenuated the difference between the statements.

B.2 Test of Additional Implication

Another implication of the theoretical framework in Appendix A is that the treatment effect is larger for people who do not have prior beliefs that the government's signal is very reliable or very unreliable.[45]

We test this prediction in two ways. First, consider heterogeneity by political party. Before treatment, we ask participants how likely they are to vote for Donald Trump in the 2020 election on a scale of 0–10. Based on this response, we form three groups: people who oppose Trump (0–2), people who support Trump (8–10), and people who are undecided (3–7). Next, we estimate Equation (1) within each group, focusing on effects on beliefs.

We find that the treatment effects are concentrated among people who do not support Trump (Appendix Table B.7). For both the update magnitude and update propensity, we find no evidence that the Inconsistent treatment has an effect on people who support Trump (Columns 1 and 4). Meanwhile, on people who are undecided or oppose Trump, we find large treatment effects (Columns 2–3 and 5–6). We find evidence that the difference in treatment effects by party is significant: F-tests of equality between those who support and do not support Trump yield that $p < 0.1$ in both cases (Columns 1 vs. 2–3 and 4 vs. 5–6). These tests therefore contribute to a growing literature that the interpretation of the response to COVID may be affected by partisan beliefs (Allcott et al., 2020; Barrios and Hochberg, 2020; Bursztyn et al., 2020a).

Interpreted through the model, the reduced treatment effect among people who support Trump may be because these participants have very strong prior beliefs that the government's information is reliable. As a result, the Inconsistent treatment does not affect beliefs about future deaths. Consistent with this interpretation, the table shows that the mean prior government credibility index is highest for people who support Trump.

We also conduct a test of this auxiliary implication by splitting the sample into people with extreme prior beliefs about the credibility of the government and people without extreme prior beliefs about the credibility. In particular, we form a sub-sample of participants whose prior attitudes are either at the top

[44] Put another way, this test addresses a concern that some ancillary feature of the statement is really responsible for the treatment effects that we find.

[45] In the notation of the theory, $\mu_1^{\mathbf{n}} \to \mu_1^{\mathbf{c}}$ if $q = 0$ or $q = 1$. This result follows immediately from Equation (A.1).

25% of the government credibility index or the bottom 25% of this index, which we denote as the sample with "extreme attitudes." We then compare the treatment effect among this sample to the sample without extreme priors. As expected, the sample with extreme attitudes exhibits an attenuated Inconsistent treatment effect relative to the rest of the sample (Appendix Table B.8).

Given the large effect by political party, one may be concerned that *partisanship*, rather than *prior beliefs about COVID-19*, explains the concentration of the Inconsistent treatment among pessimists displayed in Table 2. To address this concern, in Appendix Table B.9 we present a version of Table 2 that limits the sample only to people who do not express support for Trump. This table shows the same concentration of the treatment effects among pessimists. As a result, we find no evidence consistent with partisanship: even excluding Trump supporters (who have relatively more optimistic beliefs), we obtain the same heterogeneity by prior beliefs.

B.3 Additional Outcomes

B.3.1 Net Update and Confidence Interval

We now present the effects on the net (signed/directed) update, rather than the update magnitude (unsigned/non-directed). Appendix Table B.10 presents the results on $\ln(\text{posterior} + 1)$, using a version of Equation (1) that controls for $\ln(\text{prior} + 1)$ in addition to stratum fixed effects. We find no net effect on belief updates. Intuitively, a substantial fraction of people have prior beliefs that are more severe than the government projection. As a result, there is no average direction that participants will update. Next, we look at people whose priors are below and above the government projection. These columns highlight that the Inconsistent treatment reduces the update propensity in the direction of the information among optimists (people with priors below the government projection). For optimists, the Inconsistent treatment reduces posterior beliefs; while the coefficient is not significant at $\alpha = 0.1$, these results provide additional suggestive evidence that the Inconsistent treatment reduced the perceived credibility of the government information, since optimists exhibit reduced updating in the direction of the information. We note that for people whose priors are above the information, the Inconsistent treatment has no statistical effect on posterior beliefs, so we do not detect this phenomenon among pessimists.

In Appendix Table B.11 we show that the Inconsistent treatment has no effect on average uncertainty as measured by the participants' incentivized confidence interval about the number of deaths.[46]

B.3.2 Other Beliefs

In addition to our main elicitation about deaths, we elicit:

- Participants' prior and posterior beliefs about the death rate from COVID for 20–50 year olds and 51–80 year olds.
- Prior and posterior beliefs about the level of the Dow Jones Index (DJI) in 6 months.

These were both pre-registered as secondary outcomes and incentivized using a binarized quadratic scoring rule.

In Appendix Table B.12, we present the effect on these alternate belief updates. We show effects on belief updates and the net update. We find no effects on the magnitude of belief updates. We find that the Inconsistent treatment reduces beliefs about the death rate for ages 50. These outcomes are ancillary to the analysis because neither death rates nor the Dow Jones are tightly linked to the government projection about the number of deaths we provided.[47]

B.3.3 Behavior-Change Tasks

We conducted the following tasks to measure behavior changes.

[46] We drop one observation from the analysis of confidence intervals; see Appendix C for details.

[47] We elicited death rates as the fatality rate of COVID relative to the flu. For instance, $\hat{\beta} \approx -2$ implies the Inconsistent treatment reduces fatality of COVID by a factor of 2 (e.g., from 21 times as fatal — the average prior belief about death rate in the Consistent group — to 19 times as fatal).

- **Data entry task.** We randomly assign participants to complete an incentivized data entry task on one of two topics. In one task, we provide information about the number of COVID cases in eight states and ask participants to sort them from highest to lowest. In a second task, we provide data about the population of eight metropolitan areas and ask participants to sort them from highest to lowest.[48]

- **Demand for information.** We permit participants to choose a link to an article on one of four topics, which they receive at the end of the experiment. The participants choose between articles presenting: (i) cute animal photographs and videos; (ii) information about the Senate CARES Act's effects on health insurance; (iii) information about maintaining wellness during COVID; or (iv) information about COVID's case and death counts in the United States.

- **Risk preferences.** We enter participants in a sublottery. Conditional on winning the original incentivizing lottery, they choose between (i) a 50-50 gamble for $0 or $20, and (ii) a certain $10.

- **Willingness to pay for notifications.** We offer participants an identical notifications service as in Treatments 1 and 2. We notify about the following five goods: N95 masks, hand sanitizer, coffee, toilet paper, and sunscreen. We pre-registered willingness to pay for notifications about sunscreen availability as a placebo outcome. We aggregate willingness to pay into an index that excludes sunscreen.

Results. In Appendix Table B.13, we present the effect of the Inconsistent treatment on several outcomes that display the effects of the Inconsistent treatment on agents' economic behaviors. We present results for three groups of outcomes: a real-effort data entry task (Panel A), demand for information and risk aversion (Panel B), and willingness to pay for notifications (Panel C). First, we aggregate results from the data entry task into an index as follows. Because we reward participants based on accuracy and speed, we standardize measures relative to the Weak Treatment group, and sum standardized accuracy and standardized (negative) time to completion.[49,50] We find that the Inconsistent treatment increases performance on a COVID-related data entry task, but has no detectable effect on the task about metro areas. We find no detectable effects on demand for information, risk preferences, or willingness to pay for notifications. Given the large number of behavioral outcomes with no effect, we suggest a cautious interpretation of the effects on data entry. One interpretation is that the Inconsistent treatment makes subjects value public information less and pay more attention to the information from this data entry task (number of COVID cases in eight states).

B.3.4 Social Distancing

We elicited one question on participants' self-reported intention to adhere to social distancing:

How many people from outside your household are you planning to meet socially, in person, in the next 2 weeks? Please do not include people you may encounter doing essential errands like going to the pharmacy or grocery store.

Participants could respond in increments of 1, from 0 to 10, and we labeled 10 as "10 or more." We construct two outcomes. First, since more than 50% of the participants report that they will see 0 people, we generate the binary variable $1(\text{plans to see others})$. Second, we convert the responses to a normalized index and reverse the sign such that a positive number suggests more social distancing (fewer people planned to meet). We call this index the "social distancing index."

We collected this outcome given the policy relevant, but pre-registered it as secondary since the conceptual link with our treatment is ambiguous. It is a self-reported measure with no incentives, and might be particularly contaminated by experimenter demand effects, especially since social distancing had become a starkly partisan issue by the time of our experiment (Allcott et al., 2020).

[48] One possible concern is that the data entry task, which is randomized, could affect outcomes. Participants conduct this task after all outcomes discussed in the main text. The only outcomes following this task were two secondary outcomes: the effects on self-reported concerns about COVID and the self-reported question about social distancing guidelines.

[49] We standardize each measure and take the average of each standardized measure, as in Kling et al. (2007).

[50] We use two measures of speed: total time spent looking at the page with the data entry task, and total time between the first and last click on the page. We prefer the latter measure and include it in the index, since it does not penalize time spent reading the instructions, but show effects on the former measure in the table.

We find no statistical effect from the Inconsistent treatment on the intention to see any people socially (Appendix Table B.14, Panel A), but a *positive* effect on the social distancing index (Panel B). In our preferred specification (Column 2), the treatment increases reported social distancing by 0.082 SD (SE: 0.042, $p = 0.052$). We find no detectable heterogeneity by prior beliefs (Panel B, Columns 3-5).

The secondary nature of the outcome and its limitations emphasized above lead us to consider these results with extreme caution. We interpret them as follows. As the Inconsistent treatment reduces faith in the *government* response, people may not wish to take unsafe *private* actions (even if they do not change their beliefs). Put otherwise, if government credibility falls, one's best response may be to stay home. This interpretation is strengthened by the heterogeneity we observe: The effect is suggestively largest among people whose prior beliefs are between 100,000 and 240,000 deaths (the same as the projection). Table 4 shows these are precisely the people who most lose faith in the government's response from the Inconsistent treatment.

To further investigate this interpretation, we conduct an exploratory strategy in which we study the effect of the government credibility index on the social distancing index. In particular, we posit the structural equation:

$$y_j = \gamma \text{PosteriorGovCredibility}_j + \boldsymbol{X}_j \delta + \varepsilon_j, \tag{B.4}$$

for $y \in \{1(\text{plans to see others}), \text{SocialDistancingIndex}\}$. We then instrument for posterior government credibility with the Inconsistent treatment. We caveat this exercise by noting that the Inconsistent treatment may affect other beliefs that affect social distancing intentions, so the exclusion restriction may not hold perfectly.

Panel C presents the results. Although this strategy is somewhat underpowered, the point estimates are consistent with our interpretation: in particular, a 1 SD *increase* in the credibility index *increases* the plans to see others by 61 pp (SE: 0.65, $p = 0.348$) and reduces the social distancing index by a marginally significant 2.2 SD (SE: 1.4, $p = 0.13$).

One may be concerned that, if participants interpreted this question as about whether they would follow official guidelines about social distancing, then this result runs counter to our main finding of reduced belief updating in response to official projections. We did not present the question as asking whether participants would adhere to official guidelines or expert recommendations. By April 3–4, almost every state had an official lock-down policy (which typically prohibits seeing any person outside the household), but we find no statistical effect on the propensity to see others. As a result, this interpretation appears unlikely; rather, it appears more consistent with a reduced belief in the ability of the government or health-care system to combat the virus.

B.4 Robustness Appendix Tables

Table B.1: Effect of Inconsistent Treatment, Different Controls

	(1) Update magnitude	(2) Update magnitude	(3) Update magnitude	(4) Update propensity	(5) Update propensity	(6) Update propensity
Inconsistent	-0.539**	-0.506**	-0.485*	-0.0427*	-0.0402*	-0.0385*
	(0.252)	(0.249)	(0.250)	(0.0227)	(0.0225)	(0.0226)
Observations	1900	1900	1900	1900	1900	1900
Stratum FE	No	✓	✓	No	✓	✓
Demographic FE	No	No	✓	No	No	✓
p-value:						
Robust standard errors	.033	.043	.052	.06	.075	.088
Permutation test	.023	.039	.048	.056	.079	.08

Standard errors in parentheses

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

Columns 1–3 present the Inconsistent treatment effect on $\ln(|\text{posterior} - \text{prior}| + 1)$. Columns 4–6 present the effect on $1(\text{chooses to update})$. Columns 1 and 4 include no fixed effects. Columns 2 and 5 include stratum fixed effects and are identical to the results in Table 2. Columns 3 and 6 include stratum and demographic fixed effects for race/ethnicity, education, household income, and geographic region. We present p-values from t-tests using robust standard errors and exact p-values from permutation tests with 1,000 permutations. Stars represent p-values from t-tests.

Table B.2: Effect of Inconsistent Treatment on $|\text{posterior} - \text{prior}|$

	(1) Update magnitude: $\mid$ posterior − prior $\mid$	(2) Update magnitude: $\mid$ posterior − prior $\mid$	(3) Update magnitude: $\mid$ posterior − prior $\mid$	(4) Update magnitude: $\mid$ posterior − prior $\mid$	(5) Update magnitude: $\mid$ posterior − prior $\mid$
Inconsistent	258648.3	-7073.3	-5255.8*	-12534.1***	-7244.2**
	(961855.7)	(4474.9)	(2686.4)	(4741.4)	(3053.5)
Observations	1900	1900	1900	1796	1687
p-value	0.788	0.114	0.051	0.008	0.018
Note		Winsorized: 95th percentile	Winsorized: 90th percentile	Drop if > 95th percentile	Drop if > 90th percentile

Standard errors in parentheses

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

All columns present the Inconsistent treatment effect on the magnitude of belief updates, $\mid$ posterior − prior $\mid$. Column 1 does not adjust posterior or prior beliefs before forming the update magnitude. Columns 2 and 3 winsorizes the outcome at the 95th and 90th percentiles of the Consistent sample. Columns 4 and 5 drop observations with prior or posterior beliefs exceeding the 95th or 90th percentiles of the Consistent sample's prior or posterior beliefs. All regressions include stratum fixed effects.

Table B.3: Effect of Inconsistent on Update Magnitude: Inverse Hyperbolic Sine Outcome

	(1) Update magnitude: $\sinh^{-1}(\lvert\text{posterior} - \text{prior}\rvert)$	(2) Update magnitude: $\sinh^{-1}(\lvert\text{posterior} - \text{prior}\rvert)$	(3) Update magnitude: $\sinh^{-1}(\lvert\text{posterior} - \text{prior}\rvert)$	(4) Update magnitude: $\sinh^{-1}(\lvert\text{posterior} - \text{prior}\rvert)$
Inconsistent	-0.534**	-0.380	-0.448	-0.896
	(0.264)	(0.355)	(0.476)	(0.644)
Observations	1900	909	544	447
p-value	.043	.285	.346	.165
Sample	All	Priors below info	Priors at info	Priors above info

Standard errors in parentheses

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

This table presents treatment effect coefficients $\hat{\beta}$ from the Inconsistent treatment (Equation (1)) for $\sinh^{-1}(\text{posterior} - \text{prior})$. All regressions include stratum fixed effects.

Table B.4: Effect of Inconsistent on Update Magnitude: Varying Constant Added to Log

	(1) Update magnitude: $\ln(\lvert\text{posterior} - \text{prior}\rvert + c)$	(2) Update magnitude: $\ln(\lvert\text{posterior} - \text{prior}\rvert + c)$	(3) Update magnitude: $\ln(\lvert\text{posterior} - \text{prior}\rvert + c)$	(4) Update magnitude: $\ln(\lvert\text{posterior} - \text{prior}\rvert + c)$
Inconsistent	-0.506**	-0.599**	-0.691**	-0.784*
	(0.249)	(0.299)	(0.350)	(0.401)
Observations	1900	1900	1900	1900
p-value	.043	.046	.048	.051
Value of constant c	1	.1	.01	.001

Standard errors in parentheses

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

This table presents treatment effect coefficients $\hat{\beta}$ from the Inconsistent treatment (Equation (1)). All columns present effects on the magnitude of belief updates, $\ln(|\text{posterior} - \text{prior}| + c)$, for various c. All regressions include stratum fixed effects.

Table B.5: Specification with Continuous Interaction Term

	(1) Update propensity: $1(\text{post.} \neq \text{prior})$	(2) Update mag.: $\ln(\lvert\text{post.} - \text{prior}\rvert + 1)$	(3) Govt. credibility index	(4) Update propensity: $1(\text{post.} \neq \text{prior})$	(5) Update mag.: $\ln(\lvert\text{post.} - \text{prior}\rvert + 1)$	(6) Govt. credibility index
ln(deaths) × Inconsistent	-0.0120	-0.116	-0.000554	-0.0166	-0.139	-0.0113
	(0.0118)	(0.140)	(0.00846)	(0.0142)	(0.149)	(0.00957)
Inconsistent	0.0946	0.784	-0.0309	0.140	0.985	0.0749
	(0.134)	(1.539)	(0.0960)	(0.158)	(1.618)	(0.107)
Observations	1900	1900	1900	1762	1762	1762
p-value of interaction	0.310	0.408	0.948	0.245	0.351	0.239
Sample	All	All	All	Restricted	Restricted	Restricted

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table presents results from the parametric specification in Equation (B.1) to test for treatment effect heterogeneity. The variable $\ln(\text{deaths})$ is the log of the prior beliefs about deaths, $+1$. We present p-values for the interaction coefficient $(\hat{\gamma}_2)$. All regressions include stratum fixed effects. Columns 3 and 6 include controls for the prior government credibility index. Columns 3–6 restrict the sample to participants with prior beliefs about the number of deaths in 6 months that are less than 1 million.

Table B.6: Heterogeneity: Effect of Inconsistent Treatment by Statement

	(1) Update magnitude: logs	(2) Update magnitude: logs	(3) Propensity to update	(4) Propensity to update
Inconsistent: Distancing	-0.364		-0.0226	
	(0.305)		(0.0278)	
Inconsistent: Flu	-0.644**		-0.0574**	
	(0.301)		(0.0271)	
Consistent: Distancing		0.858***		0.0658**
		(0.311)		(0.0278)
Consistent: Flu		0.163		0.0152
		(0.303)		(0.0276)
Observations	1900	1900	1900	1900
p-value: test that Flu = Distancing	.417	.054	.268	.116

Standard errors in parentheses

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

This table presents treatment effect coefficients $\hat{\beta}_{\text{flu}}$, $\hat{\beta}_{\text{distancing}}$, $\hat{\gamma}_{\text{flu}}$, and $\hat{\gamma}_{\text{distancing}}$ from Equations (B.2) and (B.3), for the Inconsistent and Consistent treatments. Columns 1 and 2 present the coefficients for the outcome $\ln(|\text{posterior} - \text{prior}| + 1)$, while Columns 3 and 4 present the coefficients for the outcome $1(\text{chooses to update})$. All regressions include stratum fixed effects.

Table B.7: Test of Auxiliary Implication: Effect of Inconsistent Treatment by Support for Trump

	(1) Update magnitude	(2) Update magnitude	(3) Update magnitude	(4) Update propensity	(5) Update propensity	(6) Update propensity
Inconsistent	0.118	-0.993*	-0.818**	0.0237	-0.0928*	-0.0701**
	(0.403)	(0.572)	(0.385)	(0.0381)	(0.0518)	(0.0337)
Observations	659	381	860	659	381	860
Sample	Support Trump	Undecided	Oppose Trump	Support Trump	Undecided	Oppose Trump
Mean prior government credibility	0.331	-0.012	-0.237	0.331	-0.012	-0.237
p-value from:						
Support Trump = Not Support Trump	.06			.036		

Standard errors in parentheses

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

This table presents treatment effect coefficients $\hat{\beta}$ from Equation (1), for the Consistent and Inconsistent treatments, separately by support for Trump (elicited before treatment). We present an F-test for equality between the treatment effect coefficient for those who do vs. do not support Trump (Columns 1 vs. 2 and 3; Columns 4 vs. 5 and 6). Columns 1–3 present effects on the update magnitude, $\ln(|\text{posterior} - \text{prior}| + 1)$. Columns 4–6 present effects on the update propensity, $1(\text{chooses to update})$. All regressions include stratum fixed effects.

Table B.8: Test of Auxiliary Implication: Effect of Inconsistent Treatment by Prior Attitudes toward Government Credibility

	(1) Update magnitude	(2) Update magnitude	(3) Update propensity	(4) Update propensity
Inconsistent	-0.342	-0.677*	-0.0188	-0.0619*
	(0.354)	(0.354)	(0.0317)	(0.0323)
Observations	950	950	950	950
Sample	Extreme	Not Extreme	Extreme	Not Extreme
p-value from:				
Extreme = Not Extreme	.499		.337	

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table presents treatment effect coefficients $\hat{\beta}$ from Equation (1), for the Consistent and Inconsistent treatments, separately by whether attitudes toward the government are or are not "extreme." We define "extreme" attitudes toward government as having a prior government credibility index that is in the top or bottom quartile. We present an F-test for equality between the treatment effect coefficient for the extreme vs. not extreme samples (Columns 1 vs. 2; Columns 3 vs. 4). Columns 1–2 present effects on the update magnitude, $\ln(|\text{posterior} - \text{prior}| + 1)$. Columns 3–4 present effects on the update propensity, $1(\text{chooses to update})$. All regressions include stratum fixed effects.

Table B.9: Effect of Inconsistent Treatment on Belief Updating (Sample: Not Support Trump)

Sample:	(1) Update propensity: $1(\text{chooses to update})$	(2) Update magnitude: $\ln(\lvert\text{posterior} - \text{prior}\rvert + 1)$	(3) N
1. All	-0.075***	-0.834***	1,241
	(0.028)	(0.317)	
2. Priors below projection (optimists)	-0.033	-0.419	549
	(0.042)	(0.440)	
3. Priors at projection	-0.080	-0.715	366
	(0.051)	(0.568)	
4. Priors above projection (pessimists)	-0.094*	-1.230*	326
	(0.055)	(0.726)	
Inconsistent mean	0.417	4.53	626
Consistent mean	0.489	5.35	615

This table presents treatment effect coefficients $\hat{\beta}$ from the Inconsistent treatment (Equation (1)) for the indicated outcome. All columns include strata fixed effects. We limit the sample only to participants who do not support Trump prior to treatment. The sample in Row 2 comprises participants whose prior beliefs about deaths in 6 months (elicited before treatment) were between 0–99,999. The sample in Row 3 comprises participants whose prior beliefs were between 100,000–240,000 (i.e., at the level of the government projection). The sample in Row 4 comprises participants whose prior beliefs were 240,001 or more deaths. Stars present p-values: $*$ indicates $p < 0.1$; $**$ indicates $p < 0.05$; $***$ indicates $p < 0.01$.

Table B.10: Effect on Inconsistent Treatment on Net Update

	(1) Net update: ln(posterior)	(2) Net update: ln(posterior)	(3) Net update: ln(posterior)	(4) Net update: ln(posterior)
Inconsistent	-0.0250	-0.0893	0.0549	0.00528
	(0.0511)	(0.0832)	(0.0583)	(0.127)
Observations	1900	909	544	447
Sample:	All	Priors below info	Priors at info	Priors above info

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table presents treatment effect coefficients $\hat{\beta}$ from the Inconsistent treatment for the net (signed) update. The regression is a version of Equation (1), where the outcome is ln(posterior +1) and we additionally control for ln(prior +1). All regressions include stratum fixed effects and control for $\ln(\text{prior} + 1)$.

Table B.11: Effect on Inconsistent Treatment on the Subjective Confidence Interval

	(1) Width of confidence interval (log)	(2) Width of confidence interval (log)	(3) Width of confidence interval (log)	(4) Width of confidence interval (log)
Inconsistent	-0.0149	-0.0253	0.0671	-0.136
	(0.0753)	(0.119)	(0.107)	(0.169)
Observations	1899	908	544	447
Sample:	All	Priors below info	Priors at info	Priors above info

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table presents treatment effect coefficients $\hat{\beta}$ from the Inconsistent treatment (Equation (1)) for the width of participants' posterior subjective confidence interval; the outcome variable is $\ln(\text{width} + 1)$. All regressions include stratum fixed effects and control for the log width of the prior confidence interval. We drop one observation from the analysis of confidence intervals because the participant did not report usable data about priors; see Appendix C.

Table B.12: Effect of Inconsistent Treatment on Additional Beliefs

	(1) Update magnitude: death rate (for ages: 50–80)	(2) Update magnitude: death rate (for ages: 20–50)	(3) Update magnitude: Dow Jones Index	(4) Net update: death rate (for ages: 50–80)	(5) Net update: death rate (for ages: 20–50)	(6) Net update: Dow Jones Index
Inconsistent	0.824	0.295	-0.180	-2.685**	-0.348	-0.0135
	(1.233)	(0.697)	(0.160)	(1.220)	(0.636)	(0.0433)
Observations	1900	1900	1900	1900	1900	1900

Standard errors in parentheses

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

This table presents treatment effect coefficients $\hat{\beta}$ from the Inconsistent treatment (Equation (1)) for the indicated outcome. Columns 1–3 present effects on the magnitude of belief updates. The outcomes in Columns 1 and 2 are $|\text{posterior} - \text{prior}|$. Because the Dow Jones has a large range, the outcome in Column 3 is $\ln(|\text{posterior} - \text{prior}| + 1)$. In Columns 4–5, we present the effect on posterior death rates, controlling for prior death rates. In Column 6, we present the effect on the log posterior beliefs about the Dow Jones, controlling for log prior beliefs about the Dow Jones. All regressions include stratum fixed effects.

Table B.13: Effect of Inconsistent Treatment on Behavior-Change Outcomes

Panel A: Effort Task

	(1) Index (sd)	(2) Accuracy (sd)	(3) Speed clicks (sd)	(4) Speed time on page (sd)	(5) Index (sd)	(6) Accuracy (sd)	(7) Speed clicks (sd)	(8) Speed time on page (sd)
Inconsistent	0.0783*	0.0883	0.0682	-0.0102	0.0198	-0.0143	0.0539	-0.0402
	(0.0412)	(0.0648)	(0.0616)	(0.0791)	(0.0417)	(0.0656)	(0.0594)	(0.0829)
Observations	916	916	916	916	984	984	984	984
Task	COVID	COVID	COVID	COVID	Metro	Metro	Metro	Metro

Standard errors in parentheses

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

Panel B: Demand for Information and Risky Gamble

	(1) Article: cute animals	(2) Article: wellness	(3) Article: health insurance	(4) Article: death counts	(5) Risky gamble
Inconsistent	-0.00538	0.0125	0.00596	-0.0130	-0.00832
	(0.0174)	(0.0211)	(0.0154)	(0.0224)	(0.0192)
Observations	1900	1900	1900	1900	1900
p-value: joint test (articles)	.882				

Standard errors in parentheses

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

Panel C: Willingness to Pay for Notifications

	(1) Index (excl. sunscreen)	(2) N95	(3) Purell	(4) Toilet Paper	(5) Coffee	(6) Sunscreen
Inconsistent	-0.0132	-0.0452	0.0299	-0.0101	-0.0662	-0.0570
	(0.0676)	(0.0547)	(0.0535)	(0.0535)	(0.0554)	(0.0513)
Observations	1900	1900	1900	1900	1900	1900

Standard errors in parentheses

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

This table presents treatment effect coefficients $\hat{\beta}$ from the Inconsistent treatment (Equation (1)) for the indicated outcome. The index in Panel A is an index of accuracy and speed for performance on the data task; the index only includes the first measure of speed (Columns 3 and 5). The distribution is normalized such that the group that received Weak Action has mean 0, standard deviation 1. Participants were randomized between the tasks. The outcomes in Panel B are an indicator for if the participant requested an article about the described topic or demanded the risky gamble vs. the sure reward. We present an omnibus F test for whether all coefficients on demanding the article are 0. The outcomes in Panel C are $\ln(\text{willingness to pay} + 1)$ for the indicated product. We pre-registered that our primary outcome would be an index of the total willingness to pay across all products excluding sunscreen, a placebo. All regressions include stratum fixed effects.

Table B.14: Effect of Inconsistent Treatment on Self-Reported Social Distancing

Panel A: 1(plans to see others)

	(1) 1(plans to see others)	(2) 1(plans to see others)	(3) 1(plans to see others)	(4) 1(plans to see others)	(5) 1(plans to see others)
Inconsistent	-0.0230	-0.0229	-0.00431	-0.00102	-0.0450
	(0.0224)	(0.0224)	(0.0455)	(0.0416)	(0.0327)
Observations	1900	1900	447	544	909
Data-task control	No	✓	✓	✓	✓
Sample	All	All	Priors above info	Priors at info	Priors below info

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Panel B: Social Distancing Index

	(1) Social distancing index (sd)	(2) Social distancing index (sd)	(3) Social distancing index (sd)	(4) Social distancing index (sd)	(5) Social distancing index (sd)
Inconsistent	0.0828*	0.0825*	0.0442	0.110	0.0683
	(0.0425)	(0.0424)	(0.0872)	(0.0771)	(0.0635)
Observations	1900	1900	447	544	909
Data-task control	No	✓	✓	✓	✓
Sample	All	All	Priors above info	Priors at info	Priors below info

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Panel C: Social Distancing and Government Credibility

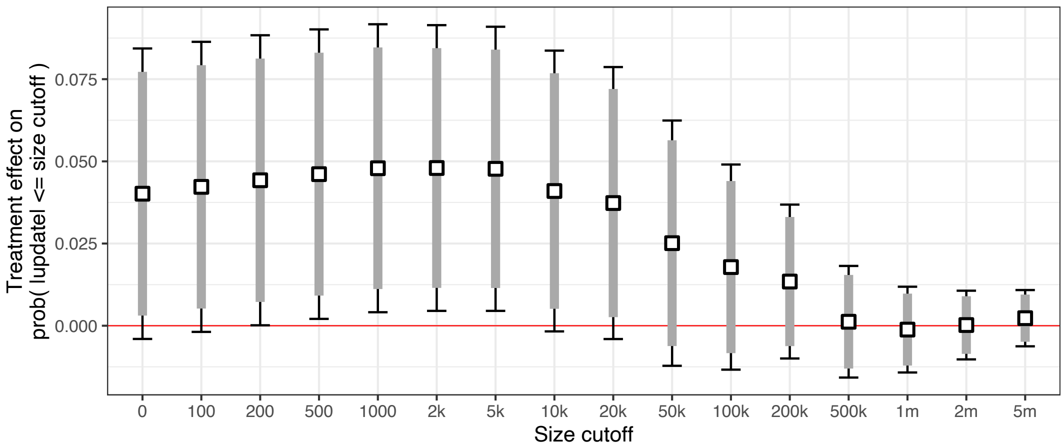
	(1) 1(plans to see others)	(2) Social distancing index (sd)
Government credibility index	0.608	-2.189
	(0.649)	(1.448)
Observations	1900	1900
Data-task control	✓	✓

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table shows the effect of the Inconsistent treatment on self-reported social distancing. Panel A presents the results for an indicator variable for whether the participant intends to see others. Panel B presents the results for a standardized variable, where positive numbers imply more intended social distancing (fewer people intended to see). Panel C presents an exploratory IV strategy (Equation B.4), where we show the relationship between intention to see others or social distancing with government credibility, instrumenting for government credibility with the Inconsistent treatment. All regressions include stratum fixed effects. Because the social distancing question was elicited after the data-entry task, we also control for the task.

B.5 Robustness Appendix Figures

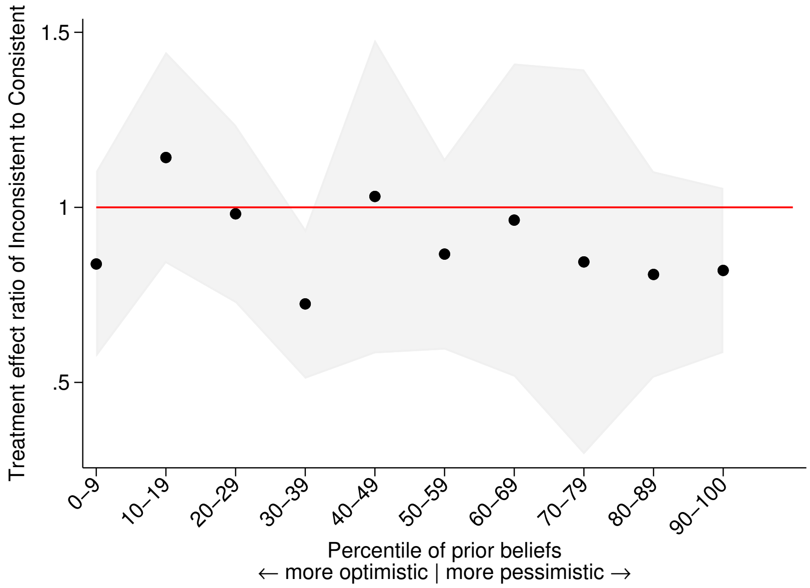
Figure B.1: Treatment Effect on Distribution of Updates



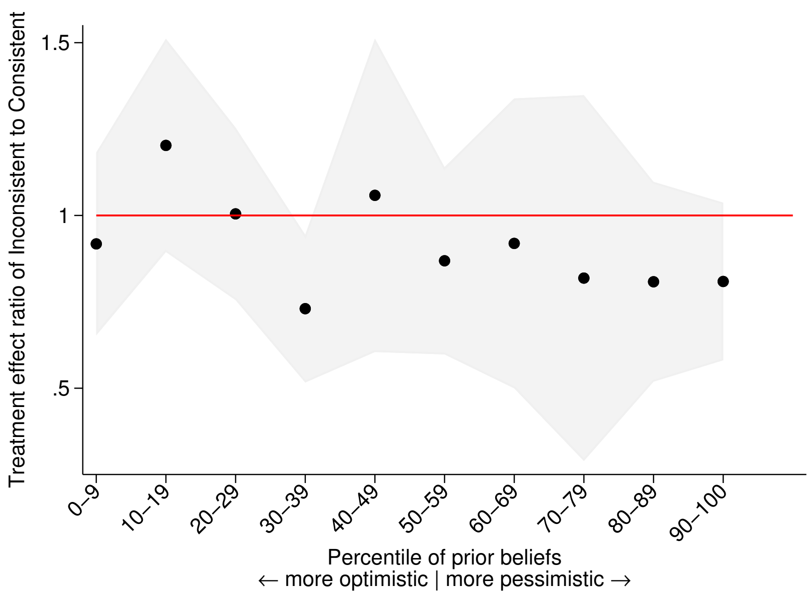
This figure presents treatment effect coefficients on $1(|\text{posterior} - \text{prior}| \leq c)$ for various cutoff values c. The x-axis varies the cutoff, and the y-axis presents the treatment effect. The black-outlined squares give point estimates, grey bars give 90% confidence intervals, and black whiskers give 95% confidence intervals using heteroskedasticity-robust standard errors. We find evidence that the Inconsistent treatment increases the propensity to make smaller updates and to make no update at all.

Figure B.2: Ratio of Inconsistent to Consistent Belief Updates: $\frac{\hat{\beta}_n^d+\hat{\beta}_c^d}{\hat{\beta}_c^d}$

(A) Magnitude of Belief Update: $\ln(|\text{posterior} - \text{prior}| + 1)$



(B) Propensity to Change Beliefs: $1(\text{chooses to update})$



This figure presents estimates of $\frac{\hat{\beta}_n^d+\hat{\beta}_c^d}{\hat{\beta}_c^d}$ from Equation (2) formed within each decile d of prior beliefs. The numerator is the update for the Inconsistent group. The denominator is the update for the Consistent group. Panel A presents the effect on the log belief update magnitude, while Panel B presents the effect on the update propensity. We exclude stratum fixed effects. We conduct inference on each point estimate using the delta method. The gray area represents 95% confidence intervals.

C Data Appendix

This Appendix provides some details and checks that pertain to the data cleaning steps.

C.1 Data construction details

We conduct two attention checks to ensure data quality. In one question, we ask the participants to select the number 7. In another question, we ask participants to enter the word "puce." We drop less than 100 participants who fail both checks from the analysis. About 10% of the remaining analysis sample fails either check. In Appendix Table C.1, we show that the treatment effect of Inconsistent on the update magnitude, update propensity, and the government support index are similar if we drop the participants who fail either attention check. The effects on the update magnitude and update propensity grow in magnitude, while the effect on the government support index shrinks modestly.

Table C.1: Effect of Inconsistent Treatment on Main Outcomes, Dropping People Who Fail Either Attention Check

	(1) Update magnitude: $\ln(\lvert\text{posterior} - \text{prior}\rvert + 1)$	(2) Update propensity: 1(chooses to update)	(3) Government credibility index
Inconsistent	-0.592**	-0.0470**	-0.0303*
	(0.263)	(0.0238)	(0.0165)
Observations	1702	1702	1702
p-value	.025	.049	.067

Standard errors in parentheses

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

Note: This table presents treatment effect coefficients $\hat{\beta}$ from the Inconsistent treatment (Equation (1)) for the indicated outcome. We drop participants who fail *either* attention check; the main tables in the text drop participants who fail *both*. We present the three main outcomes: update magnitude, update propensity, and the effect on the government index. All columns include strata fixed effects.

2.9% report that priors are too low or too high but then report identical posteriors as priors.[51] We code these participants as not choosing to update beliefs, and their update magnitude is 0.

We drop about 20 participants who erroneously submitted the experiment more than once; the Lucid.io platform is supposed to prevent participants from taking the experiment multiple times. We collected additional data after the ending treatment date, typically in hard-to-fill strata. We keep only the observations who begin the survey before April 4, the pre-registered end date (using GMT-7, the Qualtrics default). Treatment effects in Table 2 are larger if we keep people through e.g. April 5 (18 additional participants): people exposed to the Inconsistent treatment April 3–5 are 4.4 pp less likely to update (SE: 2.2, $p = 0.050$). Similarly, the Inconsistent treatment reduces the update magnitude by 54.7 log points (SE: 24.8, $p = 0.027$). We choose not to use data past April 4 in the main sample because that is the pre-registered end date, and because our experiment deals with a fast-moving crisis in which information changes rapidly. One observation reports "30%" as the lower bound for the percentile. We drop this observation from the analysis of confidence interval width, because we could not determine whether the participant meant "300" (and the % was a typo) or "30%" of the U.S. population. We winsorize the upper bound of the confidence interval width at 330,000,000, approximately the population of the United States. This winsorization affects fewer than 10 observations. We winsorize this variable because one response exceeded 10 trillion.

[51] We first ask participants whether their prior was too high or low. If they report that their prior was too high, e.g., they could not report a posterior *lower* than their prior, but it could be equal.

C.2 Government credibility index

We asked participants questions about whether they believe: (i) the government measures were appropriate, (ii) whether the government was over-reacting, (iii) whether the government was acting strong enough to counter the crisis, and (iv) whether the government has private information and acts on the private information. Participants responded to questions on a scale from 0–10. The precise questions were:

How much do you agree with the statements below about the federal government in the COVID-19 crisis?

- *The government is handling this crisis appropriately*
- *The government is not taking strong enough measures to contain COVID-19*
- *The government is over-reacting in trying to contain COVID-19*
- *The government has high-quality information that is not public, and bases its decisions on this information.*

We form standardized each question by normalizing relative to the Consistent group (normalizing the prior and posterior response separately). We form an aggregate index by summing the outcomes, dividing by 4, and changing the sign such that a positive response indicates that the participant believes the government has more credibility (Kling et al., 2007).

C.3 News exposure

The question we asked to generate the above- and below-median knowledge about the federal response was: *How much do you agree with the following statement below about the federal government in the COVID-19 crisis? I do not really know what measures the government is taking.* The question was elicited on a scale of 0–10, and the median was 4.

C.4 Death counts

We do not make adjustments to the raw death totals that we use to form our outcome variable. As a result, there are several extreme outliers; one observation reports posterior beliefs about the number of deaths in 6 months that exceed the population of the United States. We did not pre-register any changes to these variables, so we include the outliers in the analysis without adjustment. A natural way to handle these outliers is to winsorize. We now show that the main analysis of update magnitudes, measured as $\ln(|\text{posterior} - \text{prior}| + 1)$, holds if we winsorize. (The update propensity is not subject to concerns about outliers.)

We winsorize $|\text{posterior} - \text{prior}|$ at the 99th, 95th, and 90th percentiles of the Consistent distribution before constructing the update magnitude. Regardless of our choice of winsorization, we obtain results that are remarkably similar to the estimates presented in the body (Appendix Table C.2).

Table C.2: Effect of Inconsistent Treatment on Winsorized Update Magnitude

	(1) Update magnitude: $\ln(\lvert\text{posterior} - \text{prior}\rvert + 1)$	(2) Update magnitude: $\ln(\lvert\text{posterior} - \text{prior}\rvert + 1)$	(3) Update magnitude: $\ln(\lvert\text{posterior} - \text{prior}\rvert + 1)$
Inconsistent	-0.496**	-0.495**	-0.482**
	(0.247)	(0.243)	(0.239)
Observations	1900	1900	1900
p-value	0.045	0.042	0.044
Percentile for winsorization	99	95	90

Standard errors in parentheses

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

Note: This table presents treatment effect coefficients $\hat{\beta}$ from the Inconsistent treatment (Equation (1)) for the indicated outcome. We winsorize $|\text{posterior} - \text{prior}|$ at the 99th, 95th, and 90th percentile of the Consistent distribution before constructing the outcome. All columns include strata fixed effects.

D Additional Details about the Experiment

This Appendix provides additional details about the experiment. First, we present additional tests for attrition (Appendix D.1). Next, we provide information on incentives (Appendix D.2). We address possible concerns about experimenter demand (Appendix D.3) and discuss how we implemented a sequential stratification procedure (Appendix D.4). We discuss minor deviations from our pre-registration (Appendix D.5). See the authors' websites for survey instruments.

D.1 Attrition

Table D.1: Attrition Check

		(1) Consistent arm attrition		(2) Inconsistent arm attrition		(3) p-value for
		N	Rate	N	Rate	difference in rates
		409	0.305	385	0.285	0.273
Age Group	18-29	71	0.262	69	0.267	0.965
	30-39	58	0.259	74	0.300	0.380
	40-49	75	0.336	54	0.245	0.045
	50-59	78	0.339	52	0.225	0.009
	60-69	78	0.316	79	0.298	0.736
	70+	49	0.333	57	0.435	0.105
Sex	Female	232	0.313	249	0.332	0.446
Political Party	Democrat	127	0.281	136	0.298	0.632
	Republican	129	0.289	112	0.250	0.213
	Other	153	0.345	137	0.306	0.253
Census Region	Midwest	79	0.284	84	0.313	0.514
	Northeast	76	0.278	74	0.263	0.762
	South	177	0.333	154	0.280	0.066
	West	77	0.296	73	0.289	0.926
Race	White	299	0.294	269	0.267	0.186
	Black	56	0.421	48	0.336	0.181
	Asian	14	0.264	20	0.274	1.000
	Native American	9	0.360	5	0.278	0.812
	Other race	24	0.312	31	0.425	0.206
	Prefer not to answer	7	0.184	12	0.324	0.259
Hispanic Origin	Hispanic	45	0.326	36	0.300	0.752
Household Income	$14,999 or less	79	0.357	66	0.363	0.997
	$15,000-$24,999	51	0.367	43	0.293	0.225
	$25,000-$34,999	44	0.284	62	0.367	0.141
	$35,000-$49,000	67	0.342	55	0.279	0.218
	$50,000-$74,999	60	0.282	57	0.249	0.501
	$75,000-$99,999	31	0.237	33	0.220	0.850
	$100,000-$149,999	25	0.214	26	0.239	0.774
	$150,000-$199,999	13	0.310	8	0.182	0.260
	$200,000 or more	9	0.257	3	0.094	0.155
	Prefer not to answer	30	0.323	32	0.344	0.876
Education	Some high school or less, or other	29	0.426	24	0.421	1.000
	High school graduate	109	0.340	115	0.378	0.355
	Associate's degree, some college, or vocational training	140	0.329	141	0.312	0.630
	Bachelor's degree	91	0.272	73	0.209	0.068
	Graduate or professional degree	40	0.207	32	0.168	0.400
					Omnibus *F*-Test	$p = 0.349$

Note: This table presents attrition rates for both treatment arms. Attritors either manually exited the survey or took longer than 4 hours to complete the questionnaire, at which point the session automatically terminates. Columns 1 and 2 present counts and rates for attrition by treatment arm. Column 3 presents p-values for χ^2 tests of differences in proportions, comparing attrition rates between treatment and control. The top row shows no evidence of differential attrition by treatment arm: $p > 0.2$ in a test for whether those in the Inconsistent group are either more or less likely to exit. The bottom row presents an omnibus F-test for internal validity in the presence of attrition, per Ghanem et al. (2020) and using the R package *systemfit* (Henningsen and Hamann, 2008), which does not include a heteroskedasticity correction. In this test, we fit a system of seemingly unrelated regressions to jointly test for treatment-control demographic differences among complete and incomplete responses. We find no evidence of compositional differences in attrition: $p > 0.3$ in our joint test. We further note that differential attrition is unlikely to drive our main findings, as Table 1 shows that the sample we analyze — respondents who complete the survey — is demographically balanced.

D.2 Incentives

Overview. We incentivize participants based on their performance on several exercises. In addition to payment for survey completion paid by Lucid.io, all participants are entered into a lottery with 1/1,000 probability. If they win the lottery, they win a $10 Amazon gift card, and we implement all incentives.

There are two principal concerns about our incentive scheme. We note that it is implausible that either concern about our incentives are correlated with the treatment. As a result, they may attenuate our results but are unlikely to introduce bias.

First, one may fear that incentives allocated with 1/1,000 probability are too small to be relevant to participants. Using incentives with lotteries is a commonly used method in experimental economics to make incentives salient (e.g., Cullen and Perez-Truglia (2020) implement a Becker-Degroot-Marschak design with .01 probability). Charness et al. (2016) find that implementing incentives via lotteries can save on implementation costs without reducing the effectiveness of incentives. We reminded participants when we elicited beliefs that more accurate beliefs were associated with a higher reward. Feedback from our participants shows that people internalized the incentives (and, indeed, expressed disappointment if they did not win the lottery). Finally, if participants did not internalize the incentives and just reported noise, that should merely attenuate our results: it is implausible that the incentives were more or less salient by treatment status.

A second concern is that the incentives may not be credible because we promise to pay participants six months after the experiment. To pay online survey participants, we created private email accounts and informed participants that, if they win the lottery, they will receive additional bonus payments at that account in the future. We have no reason to think this method of delivering future payments was viewed skeptically by participants. Moreover, we informed participants about our university affiliation prior to obtaining consent, which should confer some credibility.

Finally, the literature (reviewed in Schotter and Trevino (2014) and Schlag et al. (2015)) on the practical importance of incentivizing belief elicitations is mixed. For instance, Trautmann and van de Kuilen (2015) find that non-incentivized introspection can yield beliefs that are similarly accurate compared to providing incentives, and Sonnemans and Offerman (2001) do not find a difference between effort and beliefs when beliefs are incentivized. Hollard et al. (2016) find that non-incentivized elicitations can perform well in subjective belief elicitation, although the probability matching method does better. On the other hand, Gächter and Renner (2010) find that incentives improve the accuracy of belief elicitations in a public goods experiment, and Armantier and Treich (2013) find that incentivized elicitations reduce noise. Schlag et al. (2015)'s review concludes by recommending using incentives if possible, but the authors also note that incentives may be less important in settings where "subjects are fresh, have no clear incentive to misreport, and face a straightforward elicitation task." Our experiment was short enough that the participants should not be fatigued, and there is no reason for participants to misreport. While the belief elicitations were somewhat unusual (in asking participants to predict future death counts), they were not complex: participants simply typed a number.

Altogether, given that (i) we did provide incentives, (ii) we clearly explained how participants would be incentivized, (iii) the fact that noisy reports due to lack of credible incentives would only attenuate our results, and (iv) the mixed evidence that incentives play a large role in belief elicitation, we find limited reason for concern about our incentives.

Binarized quadratic scoring rule. Incentives for predictions of numbers (such as the number of deaths or cases from COVID-19 in 6 months in the US, the death rates, or the Dow Jones Index) are calculated using a quadratic scoring rule, "binarized" as in Hossain and Okui (2013) to keep the scoring rule proper even if participants are not risk neutral. We explain these incentives in simple terms, with examples such as:

- A perfectly correct guess yields a prize of $50 for sure.
- If your error from the truth is 10, you would get the prize of $50 with 99% chance and $10 with 1% chance.
- If your error from the truth is 20, you would get the prize of $50 with 96% chance and $10 with 4% chance.
- If your error from the truth is 50, you would get the prize of $50 with 75% chance and $10 with 25% chance.
- If your error from the truth is 70, you would get the prize of $50 with 49% chance and $10 with 51% chance.
- If your error from the truth is 90, you would get the prize of $50 with 19% chance and $10 with 81% chance.

Incentives for data entry. If participants win the incentives lottery, we reward participants with $5 if they enter the data completely accurately and are in the top 20% of speed among people who enter the data completely accurately.

D.3 Experimenter Demand

We provide several arguments suggesting that experimenter demand is not likely to affect our results.

First, we adopted the best practices suggested by de Quidt et al. (2019) to guard against possible experimenter demand responses. In particular, we implement incentives, provide neutral instructions, and give little information about the experimenters (besides our university affiliation) to reduce social pressure. We made it clear to participants that all choices were anonymous. Finally, and most importantly, participants had no way of inferring our hypotheses about reduced belief updating or government credibility.

Second, the experimental economics literature has not typically found evidence that experimenter demand effects are important in practice (de Quidt et al., 2018, 2019). Camerer (2011) notes that participants are not generally likely to guess what the experimenter demands. The author also points out that concerns about experimenter demand are less important if evidence either in favor or against the hypothesis would still be of academic importance. This is the case in this study: if participants had instead updated more in response to Inconsistent, or updated in a given (net) direction, that could present evidence in favor of an alternate model. As a result, we had no reason to demand any behavior, even implicitly.

Third, we argue that concerns about experimenter demand are unlikely in our study in particular. The results on beliefs about COVID severity are especially unlikely to be affected by experimenter demand concerns. It is not clear how experimenter demand would affect the update magnitude or propensity. A more plausible alternate scenario would be that upon seeing the Inconsistent treatment, participants believe that the experimenters have pessimistic views about the crisis — e.g., that the government has mishandled the response. In that case, demand effects would drive participants to report more pessimistic posterior beliefs to match our supposed slant as experimenters. However, we do not find evidence of net belief updates due to the Inconsistent treatment (Appendix Table B.10), which serves as evidence against such demand effects. Additionally, given the imperfect alignment between President Trump and the CDC, it is not clear that subjects inferring a view about Trump would then use CDC information less as a result of experimenter demand. Moreover, the incentives we provide offer some confidence that it was in participants' best interest to report their true beliefs about COVID, rather than the beliefs they thought we wanted them to report (Appendix D.2).

The results on government attitudes are also unlikely to be influenced by experimenter demand. First, we find no effect on self-reported propensity to vote for Trump (Table 4). Additionally, we find a host of small or null results in Appendix B.3. For instance, we find no effect on the willingness to pay for notifications about the availability of various COVID-related goods (Appendix Table B.13). This willingness to pay would very likely be affected by experimenter demand if demand were a large factor in the results.

D.4 Sequential Stratification

The Consistent and Inconsistent treatments each consist of two messages about different topics: one about the government's social distancing recommendations, and one about the severity of COVID. The randomization in our experiment determines both the content of the messages we show a respondent (Consistent or Inconsistent), and the order we present the topics (distancing first or severity first).

We randomize the ordering of topics within treatment arms to mitigate the risk that specific information about either topic drives our findings. In our survey we deliver the first message and measure an initial set of outcomes, then deliver a second messages and measure further outcomes. Randomization balances any possible topic-specific effects on each set of outcomes, allowing cleaner inference on the effect of the Inconsistent treatment.

We assign treatment via stratified sequential randomization. We use Lucid.io's information about baseline political identification (3 levels: Democrat, Republican, other), age (3 levels: 18–39, 40–59, 60+), and sex (2 levels: male, female) to form $2 \times 3 \times 3 = 18$ strata. Within each stratum we perform a two-step randomization of participants, first into one of two treatment arms, and second into one of two topic orderings. Each step of randomization is based on arrival time and nested within the previous step. Among the first two participants to arrive in a stratum, exactly one will enter each treatment arm. Then, among the first two to enter each treatment arm, exactly one will receive each topic ordering.

To perform randomization inference, we define randomization groups based on arrival time in each stratum and permute treatment assignment within these randomization groups. Due to the sequential nature of our assignment mechanism, among the first four respondents to take the survey in a given stratum, exactly one will receive each combination of message content and topic ordering. The first four respondents in a stratum therefore form a single randomization group, and likewise for each subsequent group of four respondents. Permuting within these randomization groups generates counterfactual treatment assignment vectors that could in fact have been realized by our assignment mechanism.

To give a brief example, the first four 40-59 year old Republican males to take the survey all fall within the same stratum. Exactly two will receive each content type (Consistent or Inconsistent) and exactly two will receive each topic ordering (distancing first or severity first). Specifically, each unique combination of content and ordering will be shown to one of these four participants. When we permute treatment assignment, we interchange the treatment assignments amongst these four participants so that the previous two sentences still hold.

D.5 Pre-Registration

We pre-registered the study (without a formal pre-analysis plan) on April 2, 2020 at the AEA RCT Registry under RCT ID AEARCTR-0005639. The pre-registration includes two separate experiments studying similar outcomes. This manuscript only discusses one experiment (the Inconsistent/Consistent experiment). We now discuss our rationale for modest departures from the pre-registration. Unless stated below, we present all primary and secondary outcomes in the paper or appendix.

Experiment date The treatment was pre-registered to run on April 2–4. We required one extra day to refine the experiment, and we administered the experiment on April 3–4.

Anxiety We do not study treatment effects on anxiety because the standard anxiety test we pre-registered is proprietary, which we learned after submitting the pre-registration. Because we do not study anxiety, we also do not conduct the following related analyses:

- We do not present the secondary outcome of an "anxiety decomposition," in which we asked participants what they were anxious about. This secondary outcome was intended to shed additional insight on a primary outcome we do not study.
- We do not conduct an exploratory strategy that we pre-registered which uses the change in anxiety as an instrument for other revealed-preference outcomes.

Update measure We pre-registered our main update measure as the "update in predictions about how severe COVID-19 will be (log predicted number of deaths)." We did not specify in the pre-registration whether this would be the update magnitude or the *net* update. The paper focuses on update magnitudes: the official projection lies above the prior beliefs for a substantial fraction of people, but it also lies below the prior beliefs for a substantial fraction of people. As a result, the theory does not predict that Inconsistent treatment should reduce updating in any given direction on average. We present effects on the net update in Appendix B. We also present the update propensity as an additional measure of the "update in predictions," since we used a design that explicitly asks people whether they wish to update. This measure is easily interpretable and just represents the extensive margin of the update magnitude.

Stock market uncertainty After pre-registration, we truncated the experiment slightly and eliminated the secondary outcome of stock market uncertainty. We therefore do not present it.

Other heterogeneity To be concise, we do not present these sources of pre-registered heterogeneity:

- Baseline belief that the government acts on its own information (which is a sub-scale incorporated in the government credibility index and presented as an outcome in Table 4)
- Heterogeneity by baseline news the person has consumed (a pre-registered secondary source of heterogeneity).