Hindsight Bias and Trust in Government*

Holger Herz[†] Deborah Kistler[‡] Christian Zehnder[§] Christian Zihlmann[¶]

January 8, 2024

DISCLAIMER: AUTHORS' FINAL VERSION, ACCEPTED FOR PUBLICATION IN THE REVIEW OF ECONOMICS AND STATISTICS

Abstract

We empirically assess whether hindsight bias affects citizens' evaluation of their political actors. Using an incentivized elicitation technique, we demonstrate that people systematically misremember their past policy preferences regarding how to best fight the Covid-19 pandemic. At the peak of the first wave in the United States, the average respondent mistakenly believes they supported significantly stricter restrictions at the onset of the first wave than they actually did. Exogenous variation in the extent of hindsight bias, induced through a randomized survey experiment, indicates that hindsight bias has a negative causal impact on the change in trust in government.

Keywords: Hindsight bias, Trust in Government, Evaluation distortion, Biased Beliefs

JEL Classification Codes: D72, D83, D91

*This study was approved by the IRB of the University of Lausanne, Switzerland. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. We thank Björn Bartling, Alain Cohn, Holly Dykstra, Michael Kosfeld, Lydia Mechtenberg and numerous seminar and conference participants for helpful comments and discussions.

[†]University of Fribourg, Switzerland, holger.herz@unifr.ch

[‡]ETH Zurich, Switzerland, deborah.kistler@econ.gess.ethz.ch

[§]University of Lausanne, Switzerland, christian.zehnder@unil.ch

[¶]University of Fribourg, Switzerland, christian.zihlmann@unifr.ch; Bern University of Applied Sciences, Switzerland, christian.zihlmann@bfh.ch

1 Introduction

Hindsight bias — also known as the "I-knew-it-all-along" effect — captures peoples' tendency to believe ex post that an outcome or event was evident from the very beginning (Fischhoff, 1975). Kahneman (2011, p.202) aptly describes the phenomenon as follows: "A general limitation of the human mind is its imperfect ability to reconstruct past states of knowledge, or beliefs that have changed. Once you adopt a new view of the world [...], you immediately lose much of your ability to recall what you used to believe [...]."

In this paper, we empirically assess whether hindsight bias affects the evaluation of political actors by their citizens. Our study is based on an original data set that we collected in the early phase of a major crisis: the Covid-19 pandemic. When it became increasingly likely that Covid-19 would result in a world-wide outbreak, policymakers around the globe had to decide about the extent to which they implement restrictions that would not only slow down the spread of the virus, but would also substantially curtail citizens' freedom. However, at that time, there was still a lot of uncertainty about many aspects of the pandemic. Information only became available as the pandemic evolved, so that citizens (and policymakers) constantly adjusted their beliefs about optimal policy, leaving ample potential for hindsight bias.

We use this situation to assess whether citizens exhibit hindsight bias in this context and whether it has a causal impact on their trust in government.[1] Understanding whether trust in government is undermined by hindsight bias in times of crisis is important: it is a key determinant of a state's legitimacy (Acemoglu, Cheema, Khwaja, & Robinson, 2020) and affects citizens' policy compliance (see Bargain & Aminjonov, 2020, for Covid-19 restrictions, and Lazarus et al., 2021, for vaccine acceptance); a decrease in trust in government during an ongoing crisis may therefore weaken a state's capacity to act effectively.

To address these questions, we simultaneously measured hindsight bias in policy preferences as well as changes in trust in government over time in an online randomized survey experiment

[1] See Madarász (2011) for a theoretical development of the argument that hindsight bias negatively affects the evaluation of agents. In the public health literature, it is a long-standing conjecture that hindsight bias undermines trust in authorities (see e.g. Redelmeier & Shafir, 2020).

with exogenously induced variation in hindsight bias. On March 15, 2020, in the very early days of the outbreak in the United States, we conducted the first stage of the experiment in which we elicited respondents' policy preferences about how to fight the pandemic.

A month later, in mid-April 2020, when the pandemic was at the peak of the first wave, we launched the second stage and re-invited the same group of respondents to a follow-up survey. In this second survey, we used an incentivized procedure to elicit whether respondents correctly remembered their policy preferences stated one month earlier. In addition to these Recalled Preferences, we also collected respondents' Updated Preferences, that is, their retrospective view in mid-April about the policies that the government should have implemented as of March 15.

We find that respondents' memory is indeed systematically biased. In mid-April, respondents (wrongly) believe that on March 15, they would have preferred to implement significantly stricter policies than they actually did. When we aggregate respondents' policy preferences into a restrictiveness index, we find that the difference between the Original Preference and the Recalled Preference is highly significant concerning both the mean and the distribution of this index. We further find that respondents' Recalled Preference is highly and significantly skewed towards their current Updated Preference.

A consequence of the observed hindsight bias could thus be that respondents downward bias their evaluation of the government's past actions relative to the rational counterfactual, because they incorrectly believe that they supported stricter policies all along and think that government "should have known better". To empirically assess such a potential impact, we elicited self-reported trust in government both on March 15 and a month later. These data identify the change in trust in government across the two stages at the individual level.

Our data reveal a significant negative correlation between hindsight bias and the change in trust in government. That is, respondents who exhibit a stronger hindsight bias also tend to experience a decrease in trust in government.

Importantly, our experimental design allows us to go beyond correlational evidence. In the second stage of our survey, respondents were randomly assigned to two groups. We exoge-

nously increased the extent of hindsight bias in one of the two groups. Respondents in the first group were first asked to indicate their Updated Preference before being incentivized to recall their past preference expressed on March 15 (we labeled this first group "UPDATED FIRST"). Respondents in the second group, in contrast, answered the questions in the reversed order ("RECALLED FIRST"). Research in psychology shows that explicitly formed outcome knowledge, in this case the Updated Preference, renders existing memory traces less accessible and serves as a reference point when reconstructing the Original Preference from memory (see e.g. Hell, Gigerenzer, Gauggel, Mall, & Müller, 1988; Stahlberg & Maass, 1997; Schwarz & Stahlberg, 2003). Thus, first reflecting on the Updated Preference may exogenously induce (stronger) hindsight bias, because reflecting on the Updated Preference affects memory, in turn biases recalls, and shifts the Recalled Preference closer to the Updated Preference. Our data confirm that respondents in UPDATED FIRST exhibit significantly larger hindsight bias than those in RECALLED FIRST. Moreover, respondents in UPDATED FIRST show a .14 standard deviations stronger decrease in trust in government compared to RECALLED FIRST. This suggests that an exogenously induced increase in hindsight bias leads to a more harsh evaluation of the government.

To move beyond correlational evidence for the link between hindsight bias and trust in government, we employ an instrumental variables approach, using the random assignment to treatments as an instrument. Results indicate that hindsight bias causally and significantly reduces the change in trust in government: a one standard deviation increase in hindsight bias leads to a decrease of the change in trust in government by .63 standard deviations.

Our data are in line with the theoretical argument that hindsight-biased principals assess the performance of agents too harshly (Camerer, Loewenstein, & Weber, 1989; Frey & Eichenberger, 1991; Madarász, 2011; Schuett & Wagner, 2011). Existing empirical studies on this topic mainly consist of laboratory experiments demonstrating that hindsight bias correlates with sub-optimally low delegation rates (Danz, Kübler, Mechtenberg, & Schmid, 2015) and causally drives excess entry in tournaments (Danz, 2020). Our paper provides evidence indicating a

causal negative impact of hindsight bias on the evaluations of agents.

A causal effect of hindsight bias on trust in government may have had substantial real-life consequences in the state's capacity to fight the pandemic, as trust in government affected citizens' policy compliance (Bargain & Aminjonov, 2020; Lazarus et al., 2021). Thus, accounting for hindsight bias when setting policy, or developing effective strategies that reduce hindsight bias, appear to be important when determining optimal policy.

2 Research Design

2.1 The experiment

We conducted our randomized survey experiment during the first wave of the Covid-19 outbreak in the United States (see Figure 1 for the timeline).[2]

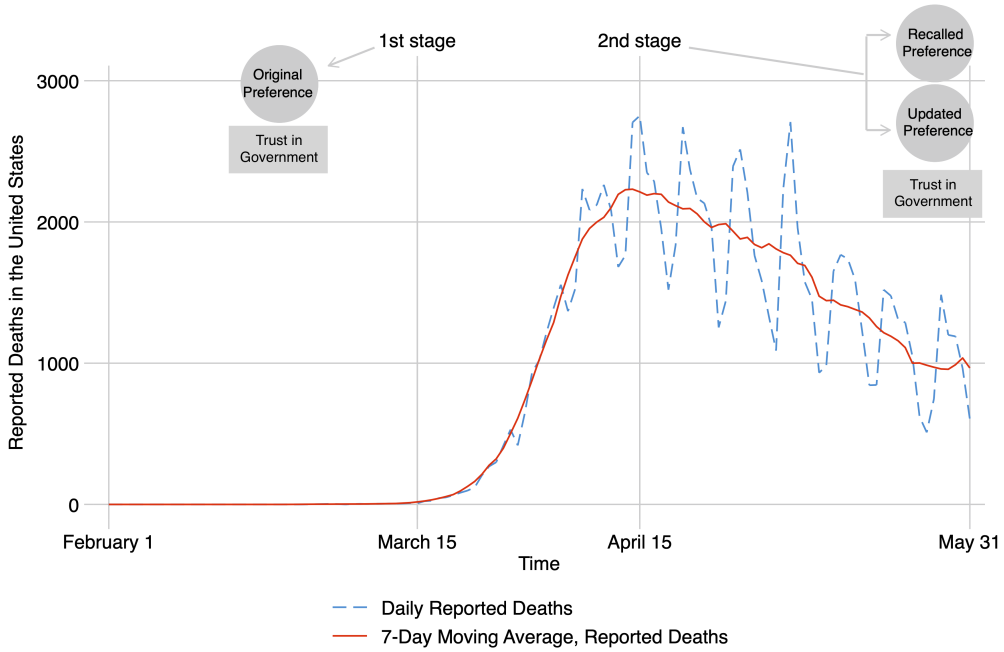
The first stage took place on March 15, 2020, when the Covid-19 outbreak was in its early days (with a total of 3600 confirmed cases and 68 confirmed deaths).[3] We asked respondents to indicate the level of restrictiveness they considered appropriate in three different policy dimensions aiming to contain the pandemic: travel restrictions, social distancing restrictions in affected states, and social distancing restrictions nationwide. We also elicited respondents' degree of approval with the measures taken by the federal government at the time (see Table 1 for details). We made it clear that the policies are ordered by restrictiveness, and that the more restrictive policies always include the measures of the less restrictive policies (i.e. respondents only could choose one policy). In addition to policy restrictions, we also measured respondents' levels of trust in government. Specifically, we asked respondents[4]: How much of the time do you think you can trust the federal government to do what is right? The answer options were: "Always", "A lot of the time", "Not very often", "Almost never".

[2]The experiment was not pre-registered, because fast implementation was critical when the idea for this study arose. For a detailed discussion, see Appendix C.

[3]The reported case and death numbers are obtained from The New York Times Company (2020) data set.

[4]This question is adapted from the American National Election Studies https://electionstudies.org/resources/anes-guide/ and the Pew Research Center https://www.pewresearch.org/politics/2021/05/17/public-trust-in-government-1958-2021/ (accessed on February 16, 2022).

Figure 1: Covid-19 deaths in the United States from February to May 2020 and the experimental timeline



Note: The graph displays the reported Covid-19 deaths in the United States on the y-axis, plotted against the timeline (February 1, 2020 to May 31, 2020). The red solid line plots the 7-day moving average while the blue dashed line plots the daily reported deaths.

A month later, from April 13 to April 16, we conducted the second stage and invited all respondents to take part in a follow-up survey.[5] As of mid-April, the first wave of the pandemic had reached its peak in the Unites States (a total of 637,056 confirmed cases and 28,582 deaths were reported on April 15), see Figure 1. The respondents had most likely acquired additional information about the Covid-19 disease and the pandemic in general. We are interested in measuring how this natural real-world feedback had affected respondents' memory of their policy preferences stated on March 15.[6]

To elicit these Recalled Preferences, we incentivized respondents to reveal their true recall of what they told us a month before. Respondents were confronted with the same choice options, and we paid a bonus of 25 cents for every correct recall. The elicitation of the Recalled Preference allows us to identify whether respondents correctly remember their past policy preference expressed a month ago.

In addition, we also measured whether and how the newly acquired information had changed respondents' views on what should have been done a month earlier. We elicited these Updated Preferences by asking respondents to indicate their current view about the extent of restrictions that should have been implemented a month earlier.[7] Finally, we re-elicited respondents' levels of trust in government. The full survey is available in the Supporting Information ("SI").

2.2 Hypotheses, Measurement, and Identification

We hypothesize that hindsight bias is present in the context of policy preferences while a major crisis unfolds: the Covid-19 pandemic.[8]

[5]The response window was longer in the second stage to maximize retention. For simplicity, we will subsequently refer to April 15 when talking about the second stage. Neither the elicited Updated Preferences, nor the Recalled Preferences differ in statistically significant ways across the days.

[6]In the first stage on March 15, respondents were not told that there will be a second survey.

[7]We asked respondents on April 15: "As of today, please select the policy that you think should have been implemented 4 weeks ago."

[8]Hindsight bias has been shown to exist across various domains and populations (see, for example, Guilbault, Bryant, Brockway, & Posavac, 2004; Harley, 2007; Biais & Weber, 2009; Roese & Vohs, 2012).

Table 1: Survey questions used to elicit participants' original preferences (March 15, 2020)

Policy Dimension	Question	Choices
Social distancing affected States	Please choose the policy that should, according to your opinion, now be implemented in states with 300 or more cases (currently: Washington State, California, New York State).	1 No social distancing restrictions 2 Prohibiting events with more than 250 people 3 Prohibiting events with more than 50 people 4 Closing all schools and childcare facilities 5 Close all non-indispensable businesses to the public 6 Statewide lockdown with mandatory self-confinement
Social distancing nationwide	Please choose the policy that should, according to your opinion, now be implemented in the entire United States (nationwide)	Same choice options as above (nationwide)
Travel restrictions	Please choose the policy that should, according to your opinion, now be implemented in the United States.	1 No travel restrictions 2 Requesting all travelers arriving from China or Europe to self-quarantine for 14 days 3 Requesting all arriving international travelers to self-quarantine for 14 days 4 Banning flights between the U.S. & Europe and the U.S. & China 5 Close borders to end all international travel 6 Ban all interstate travel from & to all states with more than 300 confirmed infected cases 7 Ban all interstate travel
Approval of U.S. Govt. Actions	Do you think that the actions taken by the U.S. government regarding the Coronavirus pandemic as of March 14th are...?	Likert scale (7-point), with 1=far too restrictive and 7=far too unrestrictive

Note: The table displays the four survey questions that elicit respondents' belief about the appropriate extent of Covid-19 restrictions to implement. Policies were ordered from least to most restrictive, and it was made clear to the respondents that the more restrictive policies always also include the proposed less restrictive policies.

Hypothesis 1 (Existence of hindsight bias)**.** *Respondents systematically misremember their Original Preferences on how to best fight the Covid-19 pandemic. Their Recalled Preferences are biased towards their Updated Preferences.*

To obtain an individual measure of hindsight bias, we first min-max normalize each of the four preference indications to a range from 0 to 1 (with 0 representing the least restrictive policy and 1 representing the most restrictive policy). We then calculate the degree of hindsight bias at the individual level for each measure, and finally average it across the four measures.[9]

One way to quantify the degree of hindsight bias is to take the absolute value of the difference between the Original and the Recalled Preferences (the so-called "shift index"). However, this index has several weaknesses (Pohl, 2007). In particular, it ignores the Updated Preference. An alternative quantification of hindsight bias that takes the Updated Preference into account is the so-called "proximity index" (Blank, Fischer, & Erdfelder, 2003; Pohl, 2007), which is computed as follows:

$$HB_i = |\text{Updated Pref}_i - \text{Original Pref}_i| - |\text{Updated Pref}_i - \text{Recalled Pref}_i| \tag{1}$$

Note that the proximity index is identical to the shift index as long as the Updated Preference is not in-between the Original and the Recalled Preference. Empirically, however, one sometimes observes this constellation. For example, the Recalled Preference may be even more restrictive than the Updated Preference, which in turn is more restrictive than the Original Preference. The shift index would quantify individuals with such a pattern as even more hindsight biased than an individual for whom the recall coincides with the Updated Preference. The proximity index, on the other hand, assumes that hindsight bias is maximal when the recall coincides with the Updated Preference, because the current state is fully projected into

[9]For each of the three elicited preferences—the Original, Recalled and Updated Preference—, there is a strong inter-item correlation across the four policy dimensions (Cronbach's $\alpha \geq .80$). However, since three policy dimensions propose explicit policies, but the fourth measures the preference relative to the policies in place as of March 14 (see Table 1), we additionally report and make the results available separately, see the SI. The results remain qualitatively very similar.

the recollection of the past. In contrast to the shift index, the proximity index thus considers the Updated Preference as an important reference point in the quantification of hindsight bias: An individual is hindsight-biased whenever the Recalled Preference is closer to the Updated Preference than to the Original Preference. In Section 3, we present all our results using the proximity index. We replicate the results using the shift index in the SI.

The proximity index can take on values ranging from -1 to 1. The index will be zero if the Recalled Preference is identical to the Original Preference, representing a person with no systematic memory distortion. Positive values represent hindsight bias, since the Recalled Preference is closer to the Updated Preference than to the Original Preference. A person with negative index values is reverse hindsight-biased because the Recalled Preference is further away from the Updated Preference than the Original Preference. There is hindsight bias among our sample if the mean of the index is larger than zero. The existence of hindsight bias is a necessary condition in order to investigate our second research question.

Our second hypothesis posits that hindsight bias has a negative causal impact on the change in trust in government. The intuitive argument that hindsight bias distorts the evaluation of others' actions has long been recognized in the literature (Camerer et al., 1989; Frey & Eichenberger, 1991). The model of Madarász (2011) formalizes the mechanism: Hindsight-biased evaluators systematically underestimate the difference between ex post and ex ante information. Accordingly, when principals assess the quality of others' decisions that were taken based on ex ante information, their evaluations tend to be too harsh, because they misperceive the informational basis.

Applied to our setting, we hypothesize that in April 2020, hindsight-biased respondents evaluate the past policy choices of the government bleaker than respondents not suffering from the bias. This is because respondents' distorted recollection of past information makes them believe that the government should have known better. Respondents who are not subject to hindsight bias, in contrast, take into account that information has changed over the past months and therefore evaluate the government's past actions more favorably.

It is important to emphasize that this mechanism does not imply that trust in government always decreases in presence of hindsight bias. The prediction is that hindsight bias always leads to harsher evaluations of the government relative to the rational counterfactual: if an unbiased decision maker's trust in government increases over time, an otherwise identical, hindsight-biased decision maker's trust in government would increase less. Similarly, if the trust of the unbiased decision maker decreases, the trust of the hindsight-biased decision maker would decrease even more.

Hypothesis 2 (Distortion in ex post evaluations)**.** *Hindsight bias has a negative causal impact on the change in trust in government.*

The theoretical mechanism underlying our second hypothesis implies a causal impact of hindsight bias on the change in trust in government. To move beyond correlation, we thus implemented an exogenous between-subject manipulation in the second survey to obtain exogenous variation in hindsight bias. As explained in Section 1, this exogenous manipulation was achieved by randomizing the order of elicitation of the Recalled Preference and the Updated Preference: Respondents in the group RECALLED FIRST were first asked about their Recalled Preference and only then about their Updated Preference. Respondents in the UPDATED FIRST group were first confronted with the Updated Preference, implying that they were required to explicitly think about their current view before recalling their Original Preference. Because explicitly formed outcome knowledge, in our case the Updated Preference, renders existing memory traces less accessible (see e.g. Hell et al., 1988; Stahlberg & Maass, 1997; Schwarz & Stahlberg, 2003), the Original Preference should become less likely to be (correctly) retrieved from memory in UPDATED FIRST, and respondents in this treatment should thus exhibit exogenously (stronger) hindsight bias.

2.3 Procedures and Sample

The experiment was conducted on Amazon Mechanical Turk ("AMT") with the software oTree (Chen, Schonger, & Wickens, 2016). Only individuals residing in the United States were allowed to participate. Additionally, we required an approval rate of at least 95% for previous jobs, as well as a minimum of 500 completed jobs. Respondents received USD 1 for completing the first stage and the average completion time was about 5 minutes. For the second stage, respondents received a higher reward of USD 1.50 to achieve a high retention rate, plus 25 cents for each correct recall of the four Original Preferences. The average completion time in stage 2 was about 6 minutes.

When running studies on online platforms, there is a natural concern that some participants may be inattentive or bots. In Appendix B, we provide detailed arguments and data for why we believe that inattentive or bot-like responses do not threaten the validity, and are not the cause, of any of our results. First, we explain why our design rules out the possibility that our main results could have been caused by inattentive participants or bots. Second, while we did not include explicit attention checks to screen out inattentive participants, we show that our data provide several ways to detect potentially inattentive participants, and excluding them strengthens our results.

The sample size amounts to 805 respondents. 1027 respondents completed the first survey on March 15, yielding a retention rate of about 80%. We find no evidence that attrition between stage 1 and 2 is non-random. Our sample is much more diverse compared to student subject pools with regard to age, education, race, and political affiliation (see also, e.g. Snowberg & Yariv, 2021). We find that compared to the US working population, our sample is younger and better educated, which aligns with previous work (Berinsky, Huber, & Lenz, 2012; Kuziemko, Norton, Saez, & Stantcheva, 2015; Levay, Freese, & Druckman, 2016). Further details on socio-demographics and attrition are provided in Appendix A.

3 Results

3.1 Existence of hindsight bias during the Covid-19 outbreak

Our first result establishes the presence of hindsight bias during the Covid-19 outbreak in the United States and therewith provides support for Hypothesis 1.

Result 1. *People systematically misremember their Original Preference about how to fight Covid-19. In April 2020, at the peak of the first wave, the average respondent incorrectly believes that they already supported stricter restrictions at the onset of the first wave in March 2020.*

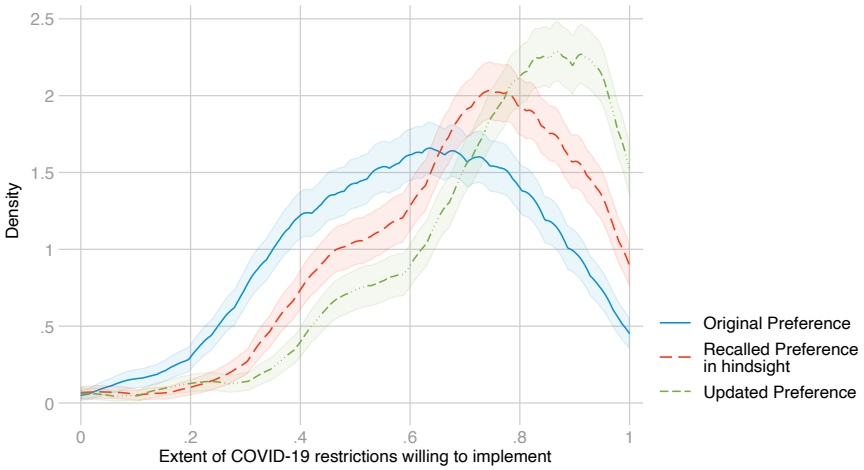
Panel 2a of Figure 2 plots the kernel density estimates of the min-max standardized preference measures regarding the Covid-19 restrictions. The solid blue line represents the distribution of the Original Preferences. At the onset of the first wave, the average respondent was in favor of implementing policies reflecting a restrictiveness index of about .61.

A month later, after having experienced the development of the first wave, respondents had updated their preferences and thought that stricter measures should have been implemented at the beginning of the pandemic. The distribution of these Updated Preferences corresponds to the dash-dotted green line. In mid-April 2020, the average respondent thought that it would have been appropriate to fight the first wave with policies reflecting a restrictiveness index of .76.

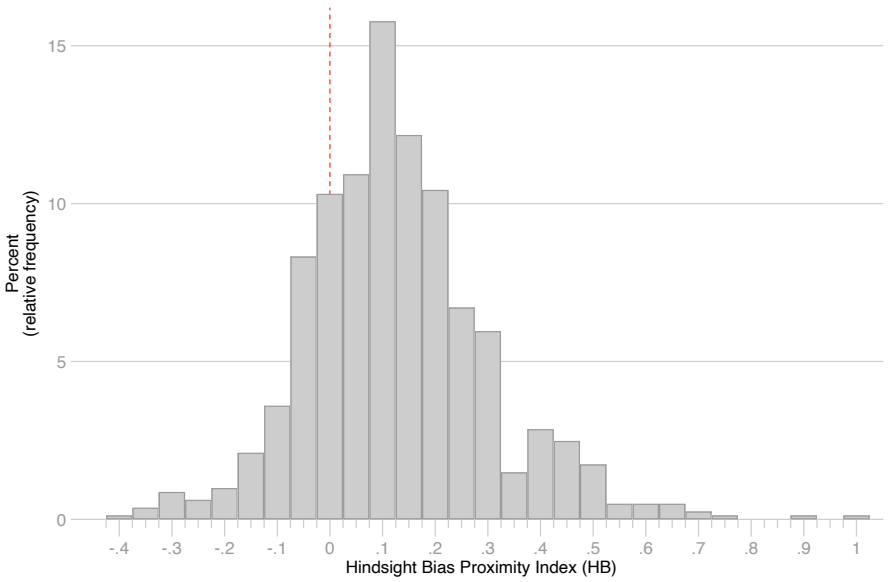
Hindsight bias suggests that peoples' Recalled Preferences should be highly skewed towards their Updated Preferences. The dashed red line representing the distribution of the Recalled Preferences confirms this prediction. The Recalled Preferences put substantially more weight on more restrictive policies and significantly differ from the Original Preferences with regard to the distribution (KS test: $p < .001$), the mean (paired t test: $p < .001$), and the median (Wilcoxon signed rank: $p < .001$). In mid-April 2020, the average respondent incorrectly believes they were already in mid-March in favor of policies reflecting a restrictiveness index

Figure 2: Existence of hindsight bias

(a) Kernel density estimates of the three preferences



(b) Histogram of the hindsight bias proximity index



Note: Panel 2a displays the kernel density estimates of the extent of Covid-19 restrictions respondents are willing to implement for the three elicited preferences, the Original Preference on March 15, the Recalled Preference on April 15 and the Updated Preference on April 15. We employ the Epanechnikov kernel with the optimal Silverman bandwidth. Shaded areas display the pointwise 95% confidence intervals. Tests of equality for the Original Preference and the Recalled Preference reveal that the two preferences differ among their location as well as their distribution (Paired t test: $p < .001$, Wilcoxon signed-rank: $p < .001$, Kolmogorov-Smirnov: $p < .001$). The histogram in Panel 2b plots the distribution of the Hindsight Bias Proximity Index (HB) as defined in Equation 1 in Section 2.1. One-sample mean and median tests against the theoretical true value of 0 both reject the null at the 0.1%-level. Sample mean $\overline{HB} = .12$, Student's one-sample t test: $p < .001$. Sample median $m = .10$, sign test: $p < .001$.

of .70. The Recalled Preference thus represents a substantial and highly significant departure from the actual Original Preference (paired t test: $p < .001$).

Panel 2b plots a histogram of the hindsight bias index as defined in Equation 1. If hindsight bias was on average absent in our sample, the index would need to be distributed with mean zero. We find that the mean is significantly larger than zero (Student's one-sample t test: $p < .001$).

3.2 Hindsight bias correlates with the change in trust in government

We observe that 29% of respondents changed their reported trust in government between mid-March and mid-April 2020 (see Table D.1 in the Appendix). Respondents are more likely to decrease their trust in government (21%) than increase their trust in government (8%) (one-sample sign test: $p < .001$; test of proportions: $p < .001$). Thus, on average, there is a decline in trust in government (Student's one-sample t test: $p < .001$), in line with other public polling at the time.[10]

At the individual level, we find that the change in trust in government negatively correlates with hindsight bias (Pearson's $r = -.09$, $p = .009$; Spearman's $\rho = -.07$, $p = .037$; Kendall's $\tau_a = -.04$, $p = .037$). These results are robust to the inclusion of control variables such as how strongly someone was affected by the pandemic or the Covid-19 cases in the county of residence, see Table D.2 in the Appendix.

Certainly, this evidence cannot be interpreted causally: hindsight bias is measured after the treatment and thus potentially suffers from post-treatment bias (Montgomery, Nyhan, & Torres, 2018), and the correlational relationship may generally suffer from endogeneity bias.

[10] See, for example, the *Rasmussen Reports Daily Presidential Tracking Poll*, https://www.rasmussenreports.com/public_content/politics/trump_administration/trump_approval_index_history.

3.3 Towards a causal effect of hindsight bias on trust in government

To assess a potential causal relationship between hindsight bias and the change in trust in government, we exploit exogenous variation in hindsight bias, induced by the random assignment of respondents to UPDATED FIRST vs. RECALLED FIRST.

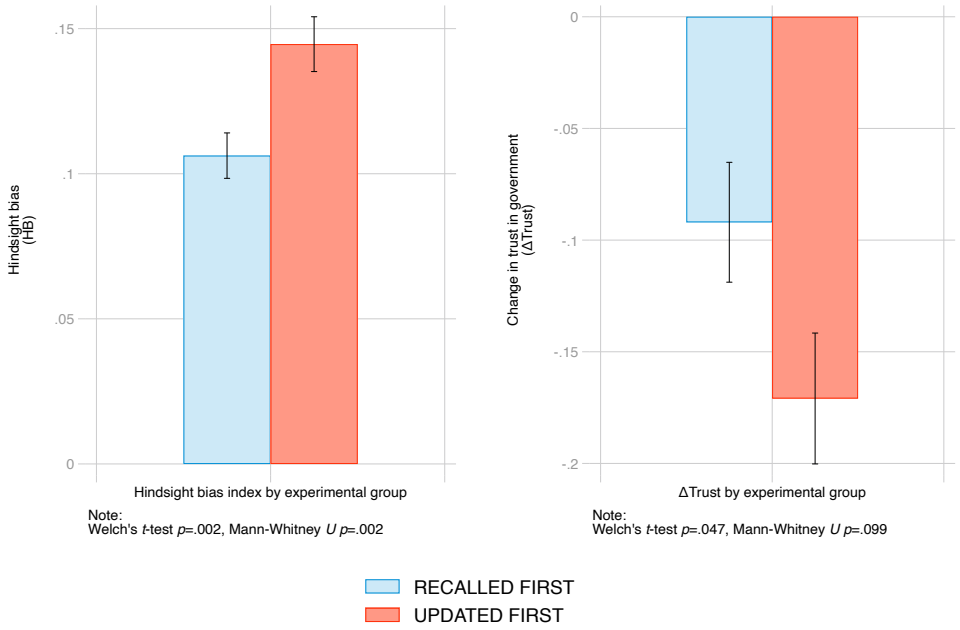
Indeed, the mean of the hindsight bias index in the UPDATED FIRST group is .145, while it is only .106 in the RECALLED FIRST group (see the left panel of Figure 3). Being confronted with the Updated Preference before recalling the Original Preference increases hindsight bias by 36%, a highly significant difference (Welch's unequal variance t test: $p = .002$, MWU test: $p = .002$). We therefore succeed in exogenously varying hindsight bias.

According to Hypothesis 2, the respondents in treatment UPDATED FIRST should in turn evaluate the government more harshly.[11] Indeed, they show a 86% stronger reduction of trust in government compared to the RECALLED FIRST group (Welch's unequal variance t test: $p = .047$, MWU test: $p = .099$), see the right panel of Figure 3. The exogenous variation of hindsight bias in UPDATED FIRST leads to a .14 standard deviations stronger decrease of trust in government.[12] This reduced form evidence suggests that respondents with an exogenously stronger hindsight bias evaluate the government more negatively than their less biased counterparts.

[11] Trust in government on March 15 does not significantly differ among the two experimental groups (Welch's unequal variance t test: $p = .259$, MWU test: $p = .359$).

[12] See Table D.3 in the Appendix for replications of this result with tobit, ordered probit and non-parametric kernel estimators.

Figure 3: Hindsight bias and the change in trust in government by experimental group



Note: The left panel depicts the mean of the hindsight bias index by experimental group. The right panel displays the mean of the change in trust in government from March 15 to April 15 by experimental group. Whiskers represent the standard error of the mean.

We can further assess this relationship using an IV approach. The randomly assigned experimental groups serve as an exogenous instrument that induces an exogenous variation in hindsight bias, allowing us to estimate the magnitude of the causal relationship between hindsight bias and the change in trust in government, conditional on the following four assumptions: the (1) existence of a first stage effect of the instrument on the endogenous variable; (2) monotonicity of the instrument; (3) exogeneity of the instrument; and (4) the exclusion restriction. As documented before, assumption (1) is satisfied in our setting, as we observe a substantial first stage effect. Second, the cumulative distribution function of UPDATED FIRST first-order stochastically dominates the distribution of RECALLED FIRST (Somers' D: $p = .002$), supporting the second assumption that the effect is monotonic, see Appendix Figure D.2. The third assumption is satisfied by design because the instrument was randomly assigned. For a more profound discussion of assumption (1)-(3), we refer to Appendix D.2.2. Assumption (4) cannot be tested empirically. We revisit and critically discuss the exclusion restriction at the end of this section.

The first stage estimation (Equation 3) regresses hindsight bias on the UPDATED FIRST group dummy, while the second stage estimation (Equation 2) regresses the change in trust in government on the first stage estimates of hindsight bias.

Second stage:

$$\Delta Trust_i = \beta_0 + \beta_1 \widehat{HB}_i + u_i \tag{2}$$

First stage:

$$HB_i = \gamma_0 + \gamma_1 UPDATED\ FIRST_i + v_i \tag{3}$$

The instrumental variable regression provides direct support for Hypothesis 2. Columns (1) and (2) in Table 2 report results from a two-stage least squares regression ("2SLS") in which both stages are estimated with least squares. The first stage regression in column (2) shows

that the change in the order of preference elicitation induces a highly significant exogenous variation in hindsight bias ($p = .002$). This result corresponds to the left panel of Figure 3 that we investigated previously.

The second stage regression in column (1) reports a negative coefficient, suggesting that hindsight bias has a causal negative impact on the change in trust in government at a statistically significant level ($p = .047$).[13] Regarding effect size, instrumented hindsight bias reduces the change in trust in government by .63 standard deviations.

Result 2. *Our data indicates that hindsight bias has a causal negative impact on the change in trust in government.*

Our analysis shows that ignoring endogeneity concerns by applying OLS leads to understating the relationship between hindsight bias and the change in trust in government. In column (5), we report the endogenous OLS model, which is statistically significant, too. When comparing the coefficient of the OLS estimation with the 2SLS estimation in column (1), we find that the OLS coefficient is smaller in magnitude. The 2SLS estimates thus suggest that some of the (positive) correlation between hindsight bias and the change in trust in government is due to endogeneity bias.

Two reasons could explain the difference between the OLS and the IV estimates.[14] First, the IV coefficient is unaffected by any potential measurement error in hindsight bias, which would bias the OLS estimates downwards. Second, IV estimates are free of any omitted variable bias. For a more pronounced discussion, please refer to Appendix D.2.4.

To assess the robustness of Result 2, columns (3) and (4) in Table 2 display coefficients of an IV regression in which the first stage is estimated with least squares and the second stage with an ordered probit estimator. A one standard deviation increase in hindsight bias decreases the

[13] Anderson-Rubin weak-instrument robust 95% confidence sets are reported in brackets, as recommended by Andrews, Stock, and Sun (2019). In presence of a single instrument, identification-robust Anderson-Rubin confidence sets are always recommended for the two-stage-least-squares estimator since these are efficient regardless of the strength of the instrument and with it, the value of the F statistic in the first stage regression.

[14] Note that OLS uses the natural variation in hindsight bias among the entire sample, irrespective of treatment, while IV estimates the local average treatment effect caused by the exogenously imposed variation of hindsight bias in the sample. The two estimands are thus different in nature.

Table 2: Change in trust in government regressed on instrumented hindsight bias

	Dependent variable: $\Delta Trust$				
	2SLS		*Ordered probit*		*OLS*
	(1)	(2)	(3)	(4)	(5)
	2nd stage	1st stage	2nd stage	1st stage	
		HB		HB	
Hindsight bias (HB)	-2.05		-3.49		-0.30
	[-6.29,-.05]		(1.43)		(0.12)
	{.047}		{.015}		{.013}
UPDATED FIRST (=1)		0.04		0.04	
		(0.01)		(0.01)	
		{.002}		{.002}	
Constant	0.13	0.11		0.11	-0.09
	(0.15)	(0.01)		(0.01)	(0.02)
N	805	805	805	805	805
F 1st stage (KP=Eff.)	9.81		9.81		
Weak iden. test (AR)	0.05		0.05		
Underidentificaton test	0.00		0.00		
Endogeneity test	0.08				
Corr. (e_v, e_u)			0.52		

Note: The table displays regression results of two instrumental variable regressions that investigate the effect of hindsight bias on the change in trust in government ($\Delta Trust$) with the accompanying OLS estimation. Model (1) and (2) report the results from a two-stage least squares estimation, regressing $\Delta Trust$ on the instrumented hindsight bias index. The first stage instruments hindsight bias with the UPDATED FIRST group dummy (column (2)). Model (3) employs an ordered probit estimator and regresses $\Delta Trust$ on the instrumented hindsight bias index. Cut-off points are not reported. Model (4) is the corresponding first stage and employs a ordinary least squares estimator to instrument hindsight bias with the UPDATED FIRST group dummy. Model (5) employs an ordinary least squares estimator and suffers potentially from endogeneity bias. For model (1), we report weak-instrument robust Anderson-Rubin 95% confidence sets for the instrumented variable in brackets. Robust standard errors are reported in column (2), (3), (4) and (5) in parentheses. p-values are reported in braces. The reported F-statistic is the Kleibergen-Paap effective F. The weak identification test reports the traditional Anderson-Rubin test based on the F-stat. The underidentification test is a Lagrange-Multiplier test based on the Kleibergen-Paap rk statistic of whether the equation is identified. The endogeneity test reports a Durbin-Wu-Hausman statistic and tests the null hypothesis whether the endogenous instrumented variable can be treated as exogenous. Corr. (e_v, e_u) indicates the correlation between the error terms of the first and second stage in the ordered probit model.

change in trust in government by .71 standard deviations. As in the 2SLS model, the coefficient is negative and significant ($p = .015$).[15] Result 2 is robust to the inclusion of controls such as self-reported experienced adverse effects of Covid-19 on own health and Covid-19 cases in the county of residence (see Appendix Table D.5), and our results remain valid across partisanship (see Appendix E).

Our instrumental variable results are valid given the premise of accepting the exclusion restriction. It requires that the instrument and the outcome are independent. The exclusion restriction is violated if being first confronted with the Updated Preference affects the change in trust in government either directly, or through a mechanism other than hindsight bias. If such confounders exist, the exclusion restriction may be violated. Note that we find significant effects in the first and second stage, the reduced form as well as the endogenous OLS model. If the exclusion restriction were fully violated (implying that the observed relationship between hindsight bias and the decrease in trust in government is not causal), the responsible mechanism would have to be able to explain these four empirical findings. Below, we briefly discuss three potential candidates for alternative mechanisms and show that none of these alternatives can explain the full set of our empirical findings. See Appendix D.2.3 for a more detailed discussion.

First, we consider the possibility that respondents might misrepresent their true preferences to appear consistent towards the experimenter (Falk & Zimmermann, 2013). Such a desire for consistency might imply that the RECALLED FIRST group feels compelled to report also a less restrictive (non-incentivized) Updated Preference compared to the UPDATED FIRST group, and in turn, again for consistency reasons, a higher trust in government compared to the UPDATED FIRST group. However, this mechanism is incompatible with the fact that we do not see a difference in the Updated Preference across treatments in our data.

Second, we explore whether a specific form of recency bias could produce our results. Reporting the Updated Preference first means that the recall of the Original Preference is more

[15] Estimating the instrumental variable models by employing trust in government on April 15 as outcome, conditional on trust in government on March 15, yields qualitatively very similar results (see Appendix Table D.4).

recent when reporting trust in government. The treatment thus reduces the interval of time and the number of questions answered between the recall of initial preferences and the reporting of trust. However, since the Recalled Preference is on average less restrictive than the Updated Preference, those who report the Updated Preference first, and the Recalled Preference second, should be more lenient with the government. Yet, we observe the opposite.

Finally, one could argue that motivated reasoning can explain the correlation between the change in trust in government and hindsight bias: participants who are increasingly frustrated with the government might be particularly prone to claim that they knew everything all along. In addition, it makes sense to assume that such an effect would be particularly strong in the UPDATED FIRST condition where participants first think about their current perspective. However, this explanation is inconsistent with the reduced-form effect showing that the randomly imposed question order affects the change in trust in government. According to this reverse-causality explanation, the change in trust in government is exogenous. A correlation between the change in trust in government and question order could then only be the result of an incidental randomization failure. However, as we show in Appendix Table A.3, treatment assignment is not predictive of trust in government on March 15 and neither of any of the four Original Preferences elicited, which speaks against such an explanation.

It is in the nature of our IV approach that we cannot completely rule out that there exist other factors able to explain all four empirical patterns in our data. However, we deem hindsight bias as the most likely (and perhaps most obvious) explanation for our results.

4 Concluding Remarks

This article shows that people are systematically hindsight-biased concerning their policy preferences during the outbreak of Covid-19 in the United States, and indicates a causal relationship between hindsight bias and trust in government. The latter finding provides direct evidence for the hypothesis that hindsight bias among voters leads to negatively biased evaluations of the

government. This evaluation distortion is consistent with Camerer et al. (1989) who suggest that hindsight bias may lead to especially acute problems in public decision making, and with Frey and Eichenberger (1991, p.75)'s conjecture that "[...] hindsight bias may again be relevant for citizens' evaluation of the government's actions. If politics leads to unfavourable results, people wrongly believe that this was foreseeable. Therefore they blame government for having committed a grave mistake."

Indeed, the mechanism identified in our paper provides a potential explanation for empirical observations such as CEOs being dismissed after bad firm performance caused by factors beyond their control (Jenter & Kanaan, 2015), or voters regularly punishing governments for hardly foreseeable events such as floods, droughts, and oil price variations (Achen & Bartels, 2017; Wolfers, 2002). While we do not claim generalizability of the effect size to other situations of crisis, we believe that the methodology proposed in this paper can be used in future studies to shed further light on this mechanism in other settings.

More generally, our finding may also have profound and so far under-explored consequences. First, if hindsight bias is anticipated by policymakers, it will affect their incentives. While theoretical work has introduced this aspect as a traditional agency conflict (see Madarász, 2011; Schuett & Wagner, 2011), it is worthwhile to further explore the implications of hindsight bias in political economy, for example how the anticipation of hindsight bias affects politician incentives when politicians compete for policy platforms.

Second, our finding may have a direct impact on which policy is constrained welfare maximizing: The first-best policy at the beginning of a crisis without accounting for hindsight bias may not be optimal in the long run. Because hindsight bias causes a deterioration in trust in government, which in turn is known to have negative effects on citizen compliance with future policy, a trade-off may exist between choosing the optimal policy to tackle the crisis and maintaining trust in government in the long-run.

Third, our results show that anchoring individuals on either current or past views first affects hindsight bias, which in turn implies that agenda setting in discussions can be a strategic

instrument. For example, if political actors, such as opposition parties, have an interest in influencing evaluations negatively, anchoring individuals on updated policy preferences may be an effective way to induce hindsight bias and thus more negative evaluations.

Finally, because the anticipation of hindsight bias can lead to a policy distortion, it is important to consider interventions that directly aim at reducing hindsight bias. Scholars in different disciplines started to study possible strategies to reduce hindsight bias, with mixed results (see, for example, Fischhoff, 1977; Arkes, Wortmann, Saville, & Harkness, 1981; Davies, 1987; Nario & Branscombe, 1995; Herzog & Hertwig, 2009; Tetlock, 2017; Pohl & Erdfelder, 2019). Yet, investigations of debiasing interventions in the domain of public policy are absent.

References

Acemoglu, D., Cheema, A., Khwaja, A. I., & Robinson, J. A. (2020). Trust in state and nonstate actors: Evidence from dispute resolution in pakistan. *Journal of Political Economy*, *128*(8), 3090–3147.

Achen, C., & Bartels, L. (2017). Blind retrospection: Electoral responses to droughts, floods, and shark attacks. In *Democracy for realists: Why elections do not produce responsive government* (pp. 116–145). Princeton: Princeton University Press.

Andrews, I., Stock, J. H., & Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, *11*(1), 727-753.

Angrist, J. D., & Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, *90*(430), 431-442.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444-455.

Arkes, H. R., Wortmann, R. L., Saville, P. D., & Harkness, A. R. (1981). Hindsight bias among physicians weighing the likelihood of diagnoses. *Journal of Applied Psychology*, *66*(2), 252.

Bargain, O., & Aminjonov, U. (2020). Trust and compliance to public health policies in times of covid-19. *Journal of Public Economics*, *192*, 104316.

Bénabou, R., & Tirole, J. (2011). Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics*, *126*(2), 805–855.

Bénabou, R., & Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, *30*(3), 141–64.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's mechanical turk. *Political Analysis*, *20*(3), 351–368.

Biais, B., & Weber, M. (2009). Hindsight bias, risk perception, and investment performance. *Management Science*, *55*(6), 1018-1029.

Blank, H., Fischer, V., & Erdfelder, E. (2003). Hindsight bias in political elections. *Memory*, *11*(4-5), 491–504.

Brysbaert, M. (2019). How many words do we read per minute? a review and meta-analysis of reading rate. *Journal of Memory and Language*, *109*, 104047.

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, *97*(5), 1232-1254.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? a comparison of participants and data gathered via amazon's mturk, social media, and face-to-face behavioral testing. *Computers in human behavior*, *29*(6), 2156–2160.

Chen, D. L., Schonger, M., & Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*(Supplement C), 88 - 97.

Chmielewski, M., & Kucker, S. C. (2020). An mturk crisis? shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, *11*(4), 464–473.

Danz, D. (2020). Never underestimate your opponent: Hindsight bias causes overplacement and overentry into competition. *Games and Economic Behavior*, *124*, 588-603.

Danz, D., Kübler, D., Mechtenberg, L., & Schmid, J. (2015). On the failure of hindsight-biased principals to delegate optimally. *Management Science*, *61*(8), 1938-1958.

Davies, M. F. (1987). Reduction of hindsight bias by restoration of foresight perspective: Effectiveness of foresight-encoding and hindsight-retrieval strategies. *Organizational Behavior and Human Decision Processes*, *40*(1), 50–68.

Dennis, S. A., Goodson, B. M., & Pearson, C. A. (2020). Online worker fraud and evolving threats to the integrity of MTurk data: A discussion of virtual private servers and the limitations of IP-based screening procedures. *Behavioral Research in Accounting*, *32*(1), 119–134.

Dreyfuss, E. (2018, Aug). *A bot panic hits amazon's mechanical turk.* Conde Nast.

Falk, A., & Zimmermann, F. (2013). A taste for consistency and survey response behavior. *CESifo Economic Studies*, *59*(1), 181–193.

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human perception and performance*, *1*(3), 288.

Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(2), 349.

Frey, B. S., & Eichenberger, R. (1991). Anomalies in political economy. *Public Choice*, *68*(1), 71–89.

Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. J. (2004). A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology*, *26*(2-3), 103-117.

Harley, E. M. (2007). Hindsight bias in legal decision making. *Social Cognition*, *25*(1), 48-63.

Hell, W., Gigerenzer, G., Gauggel, S., Mall, M., & Müller, M. (1988). Hindsight bias: An interaction of automatic and motivational factors? *Memory & Cognition*, *16*(6), 533–538.

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science*, *20*(2), 231–237.

Huber, M., & Wüthrich, K. (2019). Local average and quantile treatment effects under endogeneity: A review. *Journal of Econometric Methods*, *8*(1), 20170007.

Huettig, F., & Pickering, M. J. (2019). Literacy advantages beyond reading: Prediction of spoken language. *Trends in Cognitive Sciences*, *23*(6), 464-475.

Isaiah, A., James, S., & Liyang, S. (2018). Weak instruments in iv regression: Theory and practice. *Annual Review of Economics*.

Jenter, D., & Kanaan, F. (2015). CEO turnover and relative performance evaluation. *the Journal of Finance*, *70*(5), 2155–2184.

Kahneman, D. (2011). *Thinking, fast and slow.* Macmillan.

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P. D., Jewell, R., & Winter, N. J. (2020). The shape of and solutions to the mturk quality crisis. *Political Science Research and Methods*, *8*(4), 614–629.

Kuziemko, I., Norton, M. I., Saez, E., & Stantcheva, S. (2015, April). How elastic are preferences for redistribution? evidence from randomized survey experiments. *American Economic Review*, *105*(4), 1478-1508.

Lazarus, J. V., Ratzan, S. C., Palayew, A., Gostin, L. O., Larson, H. J., Rabin, K., ... El-Mohandes, A. (2021). A global survey of potential acceptance of a covid-19 vaccine. *Nature Medicine*, *27*(2), 225–228.

Levay, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of mechanical turk samples. *SAGE Open*, *6*(1), 2158244016636433.

Madarász, K. (2011, 12). Information Projection: Model and Applications. *The Review of Economic Studies*, *79*(3), 961-985.

Montgomery, J. M., Nyhan, B., & Torres, M. (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, *62*(3), 760–775.

Nario, M. R., & Branscombe, N. R. (1995). Comparison processes in hindsight and causal attribution. *Personality and Social Psychology Bulletin*, *21*(12), 1244–1255.

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on amazon mechanical turk. *Behavior research methods*, *46*, 1023–1031.

Pohl, R. F. (2007). Ways to assess hindsight bias. *Social Cognition*, *25*(1), 14–31.

Pohl, R. F., & Erdfelder, E. (2019). Hindsight bias in political decision making. In *Oxford research encyclopedia of politics.*

Redelmeier, D. A., & Shafir, E. (2020). Pitfalls of judgment during the covid-19 pandemic. *The Lancet Public Health*, *5*(6), e306-e308.

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on psychological science*, *7*(5), 411–426.

Schuett, F., & Wagner, A. K. (2011). Hindsight-biased evaluation of political decision makers. *Journal of Public Economics*, *95*(11), 1621 - 1634.

Schwarz, S., & Stahlberg, D. (2003). Strength of hindsight bias as a consequence of meta-cognitions. *Memory*, *11*(4-5), 395-410.

Snowberg, E., & Yariv, L. (2021). Testing the waters: Behavior across participant pools. *American Economic Review*, *111*(2), 687–719.

Stahlberg, D., & Maass, A. (1997). Hindsight bias: Impaired memory or biased reconstruction? *European Review of Social Psychology*, *8*(1), 105-132.

Stokel-Walker, C. (2018, Aug). *Bots on amazon's mechanical turk are ruining psychology studies.* New Scientist.

Taylor, S. E. (1965). Eye movements in reading: Facts and fallacies. *American Educational Research Journal*, *2*(4), 187-202.

Tetlock, P. E. (2017). Expert political judgment. In *Expert political judgment.* Princeton University Press.

The New York Times Company. (2020). *Coronavirus (Covid-19) Data in the United States.* https://github.com/nytimes/covid-19-data. (Online, accessed 12 May 2020.)

Thomas, K. A., & Clifford, S. (2017). Validity and mechanical turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior*, *77*, 184–197.

Wolfers, J. (2002). *Are voters rational?: Evidence from gubernatorial elections.* Graduate School of Business, Stanford University Stanford.

Appendix

A Appendix: The data

A.1 Demographics characteristics of the sample

We briefly compare the workers who participated in our experiment with the U.S. working population in this section. In general, our sample is remarkably diverse and relatively similar to the representative U.S. working population.

Table A.1: Demographics of our data set compared with the U.S. working population

		in %	
Variable	Categories	Our Sample	U.S. working population (2019)
Gender	Women	43	47
	Men	56	53
	Other / Non-binary	1	-
Age	29 or younger	22	24
	30-39	35	22
	40-49	21	20
	50-59	13	20
	60 or older	9	14
Race	White or Caucasian	74	78
	Black or African American	8	12
	Asian or Pacific Islander	10	7
	Other	8	4
Education	High school or less	10	32
	Some college no degree	20	15
	Associate degree	12	11
	Bachelor's degree	42	26
	Graduate or above	17	16
State (Top 5)	California	11	11
	New York	8	5
	Pennsylvania	7	4
	Florida	7	6
	Texas	6	8
Party	Democrat	35	32
	Lean Democrat	19	18
	Lean Republican	13	13
	Republican	15	26
	Independent / Other	18	11
	N=	805	

Note: The table displays the demographic characteristics of our sample versus a representative sample for the U.S. labor market, namely characteristics of the U.S. working population. The source for all characteristics except party affiliation are the "Labor Force Statistics of the Current Population Survey" (2019) published by the U.S. Bureau of Labor Statistics, see https://www.bls.gov/cps/tables.htm. Party affiliation refers to the year 2020, the source is a Gallup survey https://news.gallup.com/poll/315734/party-preferences-swung-sharply-toward-democrats.aspx.

Table A.1 provides an overview. Our sample consists of slightly more men (56%) compared to the representative U.S. working population (53%). Our respondents are on average younger and better educated than the U.S. working population, two well-known features of AMT samples (Berinsky et al., 2012; Levay et al., 2016). Blacks/African-Americans are underrepresented while Asians are over-represented in our sample. Minorities are more common in our sample with 7% of our respondents not identifying themselves with any race ("Other"), compared to

the representative share of 4% among U.S. workers. These patterns align well with previous literature, see for example Kuziemko et al. (2015). The Top-5 states where our respondents reside are exactly the same five states where most of the U.S. working population lives. Our respondents are almost as likely as the U.S. working population to identify themselves as Democrat, Lean Democrat and Lean Republican. In contrast, we observe that our sample is less affiliated with the Republican party (15%) than the U.S. working population (26%). Our respondents identify themselves as "Independent" or "Other" more often (18% vs. 11%).

A.2 Procedures, Sample, Attrition

1027 respondents completed the survey on March 15. Of these, 813 respondents completed the follow-up survey a month later, yielding a retention rate of approximately 79%. 214 respondents dropped out. There is no evidence that the attrition is related to the outcome variables. We do not observe significant differences neither for the Original Preference for all of the four policy dimensions nor for expressed trust in government (see Table A.2). We further fail to reject the null that the experimental group assignment is not related to dropping out.

Table A.2: Attrition between stage 1 and stage 2

Variable (predicting not dropping out after survey stage 1)	Coeff.	p
Key variables		
Original Preference: Travel restrictions	0.035	0.472
Original Preference: Restrictions relative to gvt.	0.017	0.765
Original Preference: Social distancing restrictions in affected states	0.048	0.287
Original Preference: Social distancing restrictions nationwide	0.019	0.659
Original Belief: Composite variable	0.049	0.409
Trust in government on March 15	0.031	0.571
Demographics		
Female (=1)	-0.033	0.195
Other gender or non-binary (=1)	-0.042	0.785
Age	0.050	0.000
Bachelor degree (=1)	0.028	0.275
Some college but no degree (=1)	-0.034	0.296
Graduate degree (e.g. Master degree) or above (=1)	0.021	0.521
Associate degree (=1)	-0.041	0.314
High school or equivalent (=1)	0.004	0.919
Less than high school (=1)	-0.042	0.847
White or Caucasian (=1)	0.017	0.569
Asian, or Pacific Islander (=1)	0.090	0.014
African American or Black (=1)	-0.055	0.248
Hispanic or Spanish or Latino (=1)	-0.125	0.060
Native American (=1)	0.066	0.620
Alaskan Native or American Indian (=1)	0.209	0.000
Other race or none of the listed (=1)	-0.066	0.493
Party affiliation		
Democrat (=1)	0.019	0.477
Lean Democrat (=1)	0.033	0.286
Independent or Other party affiliation (=1)	-0.051	0.130
Lean Republican (=1)	-0.009	0.808
Republican (=1)	-0.003	0.944

Note: The table displays the key outcome variables, demographic characteristics and party affiliation in the leftmost column with the goal to test the ability of these variables to predict whether respondents drop out after the first survey on March 15 (stage 1). For each row, the coefficient and p-value are obtained from a regression model of the form $FinishedBothStages_i = \alpha + \beta \times Variable_i + \varepsilon_i$, where the respective $Variable$ is listed in the leftmost column.

Continuing the attrition analysis with demographic variables, we further fail to reject the null that attrition is not random at or above the 90%-level for gender and education. We find that age predicts dropping out: Younger people are significantly more likely to drop out ($p < .001$), and the retention rate significantly increases with age. Moreover, it seems that "Asian, or Pacific Islanders" ($p < .05$) and "Alaskan Native or American Indian" ($p < .001$) have a higher probability while "Hispanic or Spanish or Latino" have a lower probability ($p < .10$)

to finish both survey stages. Note however that there does not seem a systematic pattern that minorities are either more or less likely to drop out. It is also possible that we face some false positives given the number of tests conducted.

Importantly, of those 214 who dropped out, 197 respondents dropped out before the exogenous variation in hindsight bias was induced. These 197 respondents did not even start the second survey. 17 respondents, or about 1.7% of all respondents, dropped out while participating in the second stage, that is after they were assigned to either RECALLED FIRST or UPDATED FIRST. We fail to reject the null that the experimental group assignment is not related to dropping out at the 90%-level.

Of the 813 respondents who completed both stages, we excluded 8 respondents due to irregular, non-matching responses regarding demographic characteristics across the two stages. The final sample size thus amounts to 805 respondents.

A.3 Randomization

Our experimental group assignment was randomly performed by the computer in stage 2. Since attrition after experimental group assignment is very low (only 17 respondents dropped out in stage 2 after randomization into experimental groups, see Section A.2) and independent of experimental group, there is no reason to expect that randomization into groups did not work successfully in general—except a failure of randomization by coincidence (a false positive). Table A.3 provides a test of whether our randomization of the treatment successfully worked.

Table A.3: Randomization Check (treatment assignment)

Variable (predicting treatment assignment to UPDATED FIRST)	Coeff.	p
Policy Preferences		
Original Preference: Travel restrictions	0.028	0.677
Original Preference: Restrictions relative to gvt.	0.022	0.782
Original Preference: Social distancing restrictions in affected states	0.019	0.761
Original Preference: Social distancing restrictions nationwide	0.011	0.854
Original Preference: Composite variable	0.030	0.715
Trust in government		
Trust in government on March 15	0.029	0.257
Severeness of Covid-19		
Covid-19 cases in county of residence, 15 March	-0.000	0.849
Covid-19 deaths in county of residence, 15 March	-0.002	0.675
Demographics		
Female (=1)	0.025	0.477
Age	0.001	0.970
Education	0.008	0.559
White or Caucasian (=1)	-0.069	0.084
Asian, or Pacific Islander (=1)	0.137	0.017
African American or Black (=1)	-0.018	0.778
Hispanic or Spanish or Latino (=1)	0.044	0.613
Native American (=1)	0.315	0.080
Alaskan Native or American Indian (=1)	0.180	0.509
Other race or none of the listed (=1)	-0.114	0.351
Party affiliation		
Democrat (=1)	0.023	0.540
Lean Democrat (=1)	0.052	0.244
Independent or Other party affiliation (=1)	0.028	0.534
Lean Republican (=1)	-0.007	0.893
Republican (=1)	-0.134	0.006

Note: The table displays the key outcome variables, trust in government, Covid-19 severeness, demographic characteristics and party affiliation in the leftmost column with the goal to test the ability of these variables to predict whether respondents belong to the treatment group UPDATED FIRST, as opposed to the treatment group RECALLED FIRST. For each row, the coefficient and p-value are obtained from a regression model of the form $Treatment_i = \alpha + \beta \times Variable_i + \varepsilon_i$, where the respective $Variable$ is listed in the leftmost column.

For all key variables, that is for all Original Preferences regarding the preferred policy as

of March 15, as well as the level of trust in government, we indeed fail to reject the null that treatment assignment is random at the 95%-level. This points towards successful randomization.

The same is true for two measures of how strongly respondents were affected by Covid-19: respondents from variously affected counties were randomized evenly into the two treatment groups. Also regarding demographics, neither gender, age nor education predicts treatment assignment at the 95%-level. In treatment UPDATED FIRST, we have 13.7% more Asians than in RECALLED FIRST, which is significant at the 95%-level ($p = .017$). Also Republicans are more common in one group than in the other, and the difference is significant ($p = .006$). Note, however, that we provide additional tests of our main results by including a co-variate that controls for political party, and our results remain robust (see, for example, Table E.1). Lastly, it is possible that we face some false positives given the number of tests conducted. When we correct for multiple hypothesis testing with a Bonferroni correction, not a single co-variate predicts treatment assignment at the corrected significance level.

B Appendix: Data quality

B.1 Introduction

When using online platforms like Amazon Mechanical Turk ("AMT"), a common concern is data quality. While initial research showed that data quality is comparable to data collected in the lab (Casler, Bickel, & Hackett, 2013; Thomas & Clifford, 2017), concerns about bot-like responses and inattentive participants on ATM have been raised in 2018 by researchers on social media and in online blog posts (Dreyfuss, 2018; Stokel-Walker, 2018), and led to the so-called AMT data quality crisis (see, for example, Kennedy et al., 2020; Dennis, Goodson, & Pearson, 2020). Indeed, studies show that data quality on AMT decreased considerably in around 2018 (Chmielewski & Kucker, 2020; Kennedy et al., 2020), threatening the reliability and validity of studies using online samples such as AMT. Such behavior can lead to systematic (non-random) choice patterns that are unrelated to the questions asked. While it is rather obvious why such patterns can be problematic in studies that focus on measuring and interpreting survey responses, they can also create issues in experiments that randomly assign participants to treatments. In particular, if the same choice heuristic leads to different outcomes in different treatments, inattentive participants can even create treatment effects. However, we believe that inattentive or bot-like responses do not threaten the validity, and are not the cause, of any of the results presented in this paper. In this appendix we discuss two arguments that support this conclusion:

1. The design of our experiment rules out the possibility that the treatment effects that we observe, and thus our main results, could have been caused by inattentive participants or bots.

2. We adopt common measures to limit the number of inattentive or bot-like participants. Moreover, our data offers several ways to detect and exclude potentially inattentive participants. When excluding those participants based on stringent criteria, the measured treatment effects get stronger. These results provide strong additional support for our claim that our results are not caused by inattentive participants, but rather that inattentive participants attenuate the treatment effects.

We discuss both aspects in more detail below.

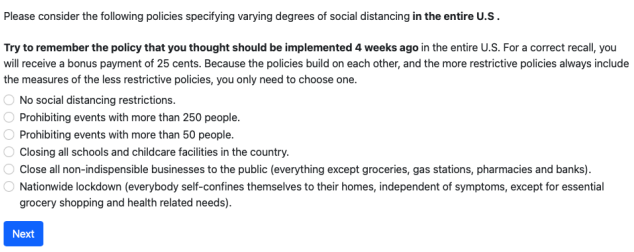
B.2 Design of the Experiment

To see why our design excludes that our findings were created by inattentive participants or bots, it is important to recall the data structure underlying our results. Result 1 (people *systematically* misremember their Original preference) relies on a within-subject comparison across our two data elicitation waves. Result 2 (hindsight bias has a negative causal impact on the change in trust in government) relies on between-subject comparison across exogenously varied treatment conditions in the second data elicitation wave.

For ease of exposition we first discuss Result 2, i.e. we first focus on the second wave of our experiment (in April 2020). The second step of our study involves a randomized controlled trial with two treatment groups. As participants re-entered our experiment, approximately half of the participants were randomized into the RECALLED FIRST treatment, and the other half into the UPDATED FIRST treatment. The survey for the two treatment groups differed between pages 4-14, where in RECALLED FIRST, on pages 4-8 participants were asked to "try to remember the policy that you thought should be implemented 4 weeks ago", and on pages 10-14 they were asked to "As of today, select the policy that you think should have been implemented in the U.S. 4 weeks ago." (p. 9 is a transition page). In treatment UPDATED FIRST, the order of these two blocks is reversed. To illustrate the similarity of the respective survey pages, Figure B.1 displays the 7th page of the survey for the RECALLED FIRST group on the left, and the UPDATED FIRST group on the right (all other wordings can be found in the SI).

Figure B.1: The seventh page of the second survey as an example of subtle differences between the experimental groups

(a) RECALLED FIRST



(b) UPDATED FIRST

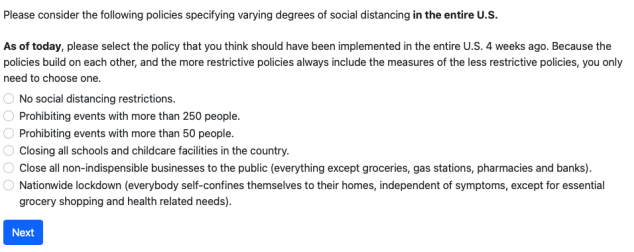


Figure B.1 shows—and it is very important to emphasize—that every aspect of the survey across the two treatments was held constant—including the precise wording and order of the answer categories—except for the nuanced content in the questions that respondents had to answer. Thus, differences in response patterns across the two treatments must imply that participants pay attention to nuanced details in the questions they were asked (and such differences are the empirical basis underlying our result 2). Otherwise, because screens did not differ in terms of the answer options, or anything else that could affect the choice behavior of a bot or inattentive participant, we would simply not observe any difference across these two treatments. However, as we show in detail in the SI, we do observe significant differences between them.

Moreover, in both treatments participants were asked about their trust in government, using the exact same wording, on the exact same page (p.17) of the respective survey. As Figure 3 in the paper shows, we find a significant effect on change in trust in government by treatment group, which again implies that the treatment affected response behavior in a way that is consistent with our behavioral hypothesis, but cannot be explained by any type of inattentive or bot-like response behavior.

Like Result 2, Result 1 is also supported by systematic and intuitively meaningful differences in responses to seemingly identical questions (identical answer categories, nuanced differences

in the wordings of the questions) within participants across the two survey waves (March and April 2020). Participants preferred less strict policies in March 2020 (Original Preference) compared to April 2020 (Updated Preference), when the first wave of the pandemic was at its peak. These shifts in policy preferences (between original and updated) occur both on average and within subject. Note also that, across the two survey waves, the placement of the questions was similar, as the Original preferences were the first questions asked in the March 2020 wave (pages 3-6).

B.3 Measures to deal with potentially inattentive participants

We adopt several measures to make sure that we survey human subjects who are attentive (see, for example, Peer, Vosgerau, & Acquisti, 2014). First, we allowed only AMT workers with an U.S. residence to participate in our task. We also required that workers have a 95% approval rate of past jobs on AMT to be eligible to participate in our survey. Lastly, we required that workers have at least completed 500 jobs on AMT, which eliminates inexperienced AMT workers.

However, despite these measures, we may still have potentially inattentive participants in our study. We therefore also attempt to directly assess the data quality of our sample. A first approach is to analyze free-form answers in which participants share their experiences regarding the impact of Covid-19 on their daily lives. This answer field allows us to determine if the responses provided are sensible and expressed in colloquial English. In a second approach, we combine answers to a free-form question about participants' primary source of information consumption with participants' self-reported political stance to see whether these two answers are consistent. Finally, we consider completion times and exclude the 10% fastest participants whose completion times were above reasonable reading speed.

B.3.1 Question about impact of Covid-19 on personal life

In stage 1, conducted on March 15, we asked respondents the following question: "Are there already any restrictions applying to your daily life due to the coronavirus? Please describe briefly." Respondents faced a free-form response field.

We conducted a manual inspection of all the responses and categorized them as attentive if the response was both sensible and written in colloquial English. Determining what constitutes a sensible response is inherently subjective. However, we believe that "please describe briefly" would typically involve more than a simple "yes", "no" or "n/a" response. Since the question asks for a brief description, it implies that the respondent should provide some additional information or details about the restrictions they are experiencing in their daily life due to Covid-19.

In fact, the average respondent provided fairly detailed descriptions of how their daily lives were affected, with the average response length being 53 characters (median: 41 characters). Examples of illustrative responses around the average length include: "Yes. Schools are closed and events are cancelled.", "Not really, we've just been told to work from home if we can.", "No, but they're coming. WFH and school closings.", "Not right now but I expect there will be in the upcoming days.", "No restrictions, but the grocery store is wiped out in some areas.", or "I'm avoiding crowded places and I stocked up at home.".

Respondents who did not provide a relevant answer to the question prompt were flagged as inattentive. Most often, these supposedly low-quality answers are a simple "No", "None", "Yes", without further description of how daily life was affected. We also flagged respondents who provided an answer in non-colloquial English. Only one respondent provided a longer but non-sensible answer to our free form question which was "IT WAS VERRY GOod", which we also labelled as inattentive.

Table B.1: Robustness of the main results to exclusion based on manual coding

	Dependent variable:				
	$\Delta Trust$				*HB*
	2SLS		*Reduced form*	*OLS*	*Result 1*
	(1)	(2)	(3)	(4)	(5)
Full sample:					
Hindsight bias (HB)	-2.05			-0.30	
	[-6.29,-.05]			(0.12)	
	{.047}			{.013}	
UPDATED FIRST (=1)		0.04	-0.08		
		(0.01)	(0.04)		
		{.002}	{.047}		
Constant	0.13	0.11	-0.09	-0.09	0.12
	(0.15)	(0.01)	(0.03)	(0.02)	(0.01)
	{.391}	{.000}	{.001}	{.000}	{.000}
N	805	805	805	805	805
F 1st stage (KP=Eff.)	9.81				
Weak identification test (AR)	0.05				
Underidentificaton test	0.00				
Endogeneity test	0.08				
Excluding inattentive respondents based on coding free form answers:					
Hindsight bias (HB)	-3.96			-0.28	
	[...,-.83]			(0.14)	
	{.013}			{.049}	
UPDATED FIRST (=1)		0.03	-0.11		
		(0.01)	(0.04)		
		{.040}	{.013}		
Constant	0.34	0.10	-0.08	-0.09	0.12
	(0.28)	(0.01)	(0.03)	(0.03)	(0.01)
	{.230}	{.000}	{.012}	{.000}	{.000}
N	631	631	631	631	631
F 1st stage (KP=Eff.)	4.22				
Weak identification test (AR)	0.01				
Underidentificaton test	0.04				
Endogeneity test	0.02				
Including only if predicting party successful:					
Hindsight bias (HB)	-3.16			-0.37	
	[...,-.76]			(0.16)	
	{.015}			{.021}	
UPDATED FIRST (=1)		0.04	-0.13		
		(0.02)	(0.05)		
		{.016}	{.015}		
Constant	0.30	0.11	-0.04	-0.06	0.13
	(0.22)	(0.01)	(0.04)	(0.03)	(0.01)
	{.177}	{.000}	{.242}	{.074}	{.000}
N	419	419	419	419	419
F 1st stage (KP=Eff.)	5.83				
Weak identification test (AR)	0.01				
Underidentificaton test	0.02				
Endogeneity test	0.03				
Excluding fast respondents:					
Hindsight bias (HB)	-2.58			-0.27	
	[...,-.35]			(0.13)	
	{.029}			{.036}	
UPDATED FIRST (=1)		0.04	-0.09		
		(0.01)	(0.04)		
		{.006}	{.029}		
Constant	0.20	0.11	-0.08	-0.09	0.12
	(0.18)	(0.01)	(0.03)	(0.02)	(0.01)
	{.265}	{.000}	{.006}	{.000}	{.000}
N	725	725	725	725	725
F 1st stage (KP=Eff.)	7.73				
Weak identification test (AR)	0.03				
Underidentificaton test	0.01				
Endogeneity test	0.05				

Note: The table displays regression results of our main results. Model (1) and (2) report the results from a two-stage least squares estimation, regressing $\Delta Trust$ on the instrumented hindsight bias index. The first stage instruments hindsight bias with the UPDATED FIRST group dummy (column (2)). Model (3) regresses $\Delta Trust$ on the experimental group dummy, our reduced form result. Model (4) regresses $\Delta Trust$ on hindsight bias, by employing an ordinary least squares estimator and thus potentially suffering from endogeneity bias. Model (5) regresses the hindsight bias index on a constant and thus tests the average of the index against the theoretical value of mean 0, providing evidence for the existence of hindsight bias among the population—our Result 1. Robust standard errors in parentheses, weak-instrument robust Anderson-Rubin 95% confidence sets in brackets, *p*-values in braces.

Our classification adheres to a stringent approach: we adopt a conservative stance in classifying respondents as being attentive, since we assume that everyone who simply provides a "yes", "no", or similar, answer is inattentive or bot-like, which clearly needs not be the case. This stringent approach led us to classify 631 out of 805 participants (78%) as attentive.

Table B.1 presents all of our main results for the full sample (first panel), and for the subset of our sample that was classified as attentive based on the classification criteria explained above (second panel).

The main conclusion is that excluding those supposedly inattentive respondents leads to stronger effects, both in terms of magnitude but also statistical significance—despite the lower sample size and with it, lower statistical power.

Specifically, column (1) displays the result of our instrumental variable estimation. In the full sample, hindsight bias reduces the change in trust in government with a coefficient of -2.05, which translates to a standardized effect of -.63 ($p = .047$). Excluding the 174 respondents that we could not unequivocally identify as attentive leads to a stronger effect: hindsight bias reduces the change in trust in government with a coefficient of -3.96, or -1.20 standard deviations ($p = .013$).

The same holds true for the reduced form effect shown in column (3), where the coefficient increases from -.08 to -.11 (an approximately 35% increase) when excluding supposedly inattentive respondents. The statistical significance increases from $p = .047$ to $p = .007$. The correlation is displayed in column (4) and becomes slightly weaker, due to an increase in the endogeneity bias of the OLS model (see the endogeneity test reported).

Result 1, the existence of hindsight bias, is reported in column (5). The coefficient is remarkably stable, and significantly different from the theoretical value of 0, with the p-value being below the 0.1%-level in the full sample as well as in the selected sample.

B.3.2 Media consumption and political stance

In stage 1, we asked respondents about their primary source of media consumption: "In general, which source do you rely on the most for news about politics and current events?" Respondents were facing a free-form response field.

To assess the attentiveness of participants, we predict the respondents' political stance based on their chosen source of media consumption, and then validate this prediction against their self-reported political stance. For this purpose, we first map respondents' main source of

Table B.2: Predicted and actual party stance

Predicted party based on media consumption	Self-reported party		Total
	Democrat	Republican	
Democrat	308	67	375
Republican	25	111	136
Total	333	178	511

media consumption onto the political spectrum. To do so, we used the "Allsides Media Bias Rating" as a reference.[16] A research assistant manually examined each response and assigned a classification to the reported media sources of either "Democrat" or "Republican".

The two most common responses were "CNN" and "Fox", making the classification exercise rather easy. Many participants, however, also indicated several news sources. In such cases, we calculated the average bias rating of the sources mentioned. For example, if a respondent mentioned "CNN, Breitbart, and Fox News", this respondent was classified as "Republican".

[16] See https://www.allsides.com/media-bias/ratings, last accessed on 7 June 2023.

A respondent who mentioned "MSNBC, NY Times, and Fox News" would be classified as "Democrat" since two out of the three sources are Democrat-oriented according to the rating.

There were 157 respondents for whom we could not assign a predicted political stance. This was because these respondents indicated that they either consumed i) no news, ii) news through an aggregator such as Google News, Reddit, or Yahoo News, or iii) news through social media platforms like Twitter. We also excluded 137 respondents who indicated that they identify as "Independent or Other Party". The reason is we cannot reasonably infer the political spectrum of Independents and those who identify with another party.

For the remaining 511 respondents, we predict that 375 are "Democrats", and 136 as "Republicans". We then compared this predicted political stance with the respondents' self-reported political stance. The rationale behind this comparison is that if someone reports consuming say MSNBC as their primary media source, it is relatively unlikely that they identify as a Republican. Therefore, we attempt to assess the coherence of respondents' self-reported political preferences with their behavior.

Table B.2 shows the outcome of this exercise. For 419 (or 82%) out of the 511 of respondents for which we were able to predict a political stance, this stance is actually correctly predicted and matches their self-reported political preference.Our main results obtained from only this very restrictive subsample are reported in the third panel in Table B.1.

We observe again that excluding these supposedly inattentive, bot-like, or non-coherent respondents increases the effect sizes as well as the significance of our results compared to the full sample. For example, the effect of hindsight bias on the change in trust in government increases in magnitude by 54% when assessed through the 2SLS estimator, and by 63% when assessed through the reduced form result. Also Result 1, the existence of hindsight bias, remains remarkably robust.

B.3.3 Time

Another variable often used to identify supposedly inattentive participants is the time elapsed on the experimental task.

We counted the length of our survey experiment, which consists of 1689 words in total for both stages together. Furthermore, as a rule-of-thumb, the literature has identified a normal reading speed for the English language of 250-300 words per minute among adults, with 300 words being rather the upper bound (Taylor, 1965; Brysbaert, 2019; Huettig & Pickering, 2019).

We thus restrict our sample by excluding the fastest 10% in terms of completion time, as the 10th percentile in terms of completion time approximately corresponds to the above mentioned upper bound. We thus consider it to be a sensible threshold in terms of time needed to complete our experiment in an attentive manner, and respondents that were faster than 338 seconds were labelled as supposedly inattentive. The bottom panel in table B.1 provides the results. It can be seen that excluding those fast respondents again increases effect sizes and the statistical significance of the main results.

In conclusion, we found that our two main results, namely the existence of hindsight bias outside the laboratory during real crises and the reduction of trust in the agent due to hindsight bias on the principals' side, remain robust after excluding potentially inattentive participants. To the contrary, the analysis suggests that our results based on the full sample may underestimate the effect size because inattentive participants attenuate the measured treatment effects.

C Appendix: Lack of Pre-Registration

We did not pre-register the experiment because the idea to conduct this study arose right when the first wave of Covid-19 infections was about to hit the US. We were worried that the delay

caused by a careful pre-registration including a complete preanalysis plan would severely reduce the variation in the Original Preference (and thus the potential for hindsight bias).

We believe that the decision not to pre-register is acceptable in this case because our main hypotheses are straightforwardly derived from existing theory.

In response to the absence of a pre-analysis plan, we are fully transparent and conservative in our analyses, by not relying on any heterogeneity analyses, analyses on subsets of the data, or other analyses that would be suspect to p-hacking (an exception are the additional analyses on the restricted samples of "attentive" participants in Appendix B that we added in response to a comment of the editor and reviewers). When there are degrees of freedom, for example, when it comes to the choice of the hindsight bias index, we report results for both options that are used in the literature.

D Appendix: Further Results

D.1 Hindsight bias correlates with a reduction in trust in government

Table D.1 provides descriptive statistics for trust in government on March 15, on April 15 and its difference — the change in trust in government $\Delta Trust$ — between the two dates. Negative (positive) values of $\Delta Trust$ represent a decrease (increase) in trust in government.

Table D.1: Trust in government

	Expressed trust in government			
	on March 15		on April 15	
How often do you trust the federal government in Washington D.C. to do what is right?	n	%	n	%
Almost never (1)	101	12.55	146	18.14
Not very often (2)	418	51.93	436	54.16
A lot of the time (3)	268	33.29	202	25.09
Always (4)	18	2.24	21	2.61
Total	805	100	805	100

	Change in trust in government	
Δ Trust: Trust on April 15 − Trust on March 15	n	%
-3 (decrease)	1	0.12
-2	4	0.50
-1	163	20.25
0 (no change)	573	71.18
1	60	7.45
2	3	0.37
3 (increase)	1	0.12
Total	805	100.00

Note: The table displays summary statistics for the survey question "How often do you trust the federal government in Washington D.C. to do what is right?". Participants were surveyed twice about their trust in government, on March 15 and a month later. We calculate the change in trust government as the difference between expressed trust on April 15 and expressed trust on March 15 and denote the variable as $\Delta Trust$.

Table D.2: Δ Trust in government regressed on hindsight bias and controls

	(1)	(2)	(3)	(4)	(5)
	Δ Trust in government				
Hindsight Bias	-0.30	-0.29	-0.30	-0.30	-0.30
	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
	{.013}	{.017}	{.013}	{.013}	{.013}
Center		-0.00			
		(0.04)			
		{.937}			
Republican		0.09			
		(0.07)			
		{.157}			
Cases per capita (in county), March 15			172.80		
			(543.70)		
			{.751}		
Cases per capita (in county), April 15				0.89	
				(1.29)	
				{.488}	
Adversely affected: Own health					0.00
					(0.01)
					{.967}
Constant	-0.09	-0.11	-0.10	-0.09	-0.09
	(0.02)	(0.04)	(0.02)	(0.02)	(0.03)
	{.000}	{.003}	{.000}	{.000}	{.006}
r2	0.008	0.012	0.009	0.009	0.008
N	805	805	805	805	805

Note: The table reports OLS regressions that investigate the effect of hindsight bias on the change in trust in government ($\Delta Trust$). Model (1) is the raw model and regresses $\Delta Trust$ on the hindsight bias index. Model (2) to (5) add control variables: Model (2) controls for party affiliation, Model (3) for cases per capita in the county of residence as of March 15, Model (4) for cases per capita in the county of residence as of April 15 and Model (5) for how strongly a participants' health was negatively affected due to Covid-19 as of April 15. Robust standard errors reported in parentheses ,p values in braces.

D.2 Towards a causal effect of hindsight bias on trust in government

D.2.1 Robustness

Table D.3: The reduced form effect: ΔTrust in government regressed on the experimental groups

	(1)	(2)	(3)
	Tobit	Ordered Probit	Kernel
UPDATED FIRST (=1)	-0.079	-0.157	-0.034
	(0.040)	(0.083)	(0.018)
	{.046}	{.058}	{.043}
Constant	-0.092		
	(0.027)		
	{.001}		
Pseudo r2	0.003	0.003	
r2			0.005
N	805	805	805

Note: All models regress Δ Trust in government on the UPDATED FIRST group dummy. Model (1) is a tobit model, with censored lower limit set to -3 and censored upper limit set to 3, robust standard errors are reported in parentheses. Model (2) is an ordered probit model, robust standard errors are reported in parentheses. Cut-off points are omitted. Model (3) reports the results of a non-parametric kernel regression, employing a Li-Racine kernel density function. Bootstrap standard errors reported in parentheses are obtained from 500 replications.

Table D.4: Trust in government on April 15 regressed on instrumented hindsight bias, conditional on trust in government on March 15

	Dependent variable: Trust (April 15)				
	2SLS		*Ordered probit*		*OLS*
	(1)	(2)	(3)	(4)	(5)
	2nd stage	1st stage	2nd stage	1st stage	
		HB		*HB*	
Hindsight bias (*HB*)	-1.64		-3.34		-0.29
	[-5.36, .24]		(1.50)		(0.11)
	{.088}		{.026}		{.009}
Trust (March 15)	0.72		1.38		0.71
	(0.03)		(0.27)		(0.03)
	{.000}		{.000}		{.000}
UPDATED FIRST (=1)		0.04		0.04	
		(0.01)		(0.01)	
		{.002}		{.002}	
Constant	0.72	0.11		0.11	0.55
	(0.15)	(0.01)		(0.01)	(0.07)
N	805	805	805	805	805
F 1st stage (KP=Eff.)	9.66		9.66		
Weak iden. test (AR)	0.09		0.09		
Underidentificaton test	0.00		0.00		
Endogeneity test	0.15				
Corr. (e_v, e_u)			0.49		

Note: The table shows the results of two instrumental variable regressions that investigate the effect of hindsight bias on trust in government on April 15, conditional on trust in government on March 15, and the accompanying OLS model in (Model (5)). Model (1) and (2) report the results from a two-stage least squares estimation, regressing *Trust* (*April* 15) on the instrumented hindsight bias index. The first stage instruments hindsight bias with the UPDATED FIRST group (column (2)). Model (3) employs an ordered probit estimator and regresses *Trust* (*April* 15) on the instrumented hindsight bias index. Cut-off points are not reported. The first stage employs a ordinary least squares estimator and instruments hindsight bias with the UPDATED FIRST group (column (4)). The Durbin-Wu-Hausman endogeneity test is not rejected in model (1), favoring the OLS instead the 2SLS model. Therefore, model (5) reports the standard OLS model that does not instrument hindsight bias. For model (1), we report weak-instrument robust Anderson-Rubin confidence sets for the instrumented variable. Robust standard errors are reported in column (2), (3), (4) and (5). The reported F-statistic is the Kleibergen-Paap effective F. The weak identification test reports the traditional Anderson-Rubin test based on the F-stat. The underidentification test is a Lagrange-Multiplier test based on the Kleibergen-Paap rk statistic of whether the equation is identified. The endogeneity test reports a Durbin-Wu-Hausman statistic and tests the null hypothesis whether the endogenous instrumented variable can be treated as exogenous. Corr. (e_v, e_u) indicates the correlation between the error terms of the first and second stage in the ordered probit model.

Figure D.1: Identification strategy

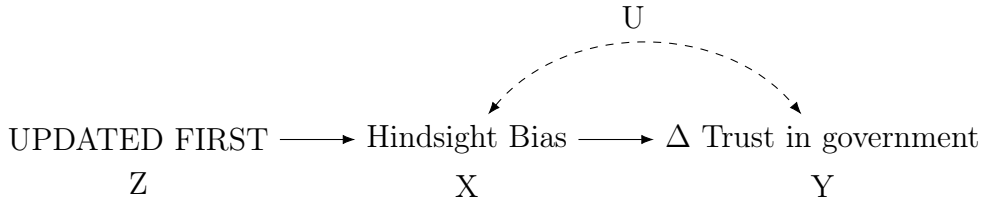


Table D.5: Change in trust in government regressed on instrumented hindsight bias and control variables

	Dependent variable: $\Delta Trust$					
	2SLS		*2SLS*		*2SLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
	2nd stage	1st stage	2nd stage	1st stage	2nd stage	1st stage
		HB		*HB*		*HB*
Hindsight bias (HB)	-2.07		-2.11		-1.94	
	[-6.41,-.06]		[-6.65,-.05]		[-5.48, -.10]	
	{.046}		{.046}		{.045}	
UPDATED FIRST (=1)		0.04		0.04		0.04
		(0.01)		(0.01)		(0.01)
		{.002}		{.002}		{.001}
Cases per capita (in county), March 15	220.41	8.31				
	(768.38)	(243.86)				
	{.774}	{.973}				
Cases per capita (in county), April 15			3.31	1.20		
			(2.85)	(0.90)		
			{.245}	{.183}		
Adversely affected: Own health					-0.02	-0.02
					(0.02)	(0.00)
					{.280}	{.000}
Constant	0.13	0.11	0.13	0.10	0.16	0.13
	(0.15)	(0.01)	(0.15)	(0.01)	(0.17)	(0.01)
N	805	805	805	805	805	805
F 1st stage (KP=Eff.)	9.81		9.42		11.39	
Weak identification test (AR)	0.05		0.05		0.05	
Underidentificaton test	0.00		0.00		0.00	
Endogeneity test	0.08		0.08		0.08	

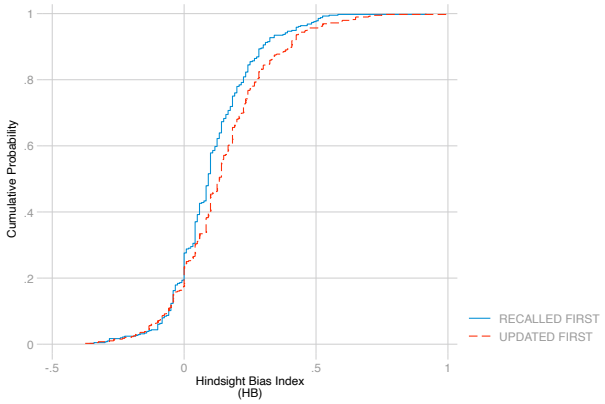
Note: The table shows the results of instrumental variables regressions (2SLS) that investigate the effect of hindsight bias on the change in trust in government ($\Delta Trust$). Model (1) and (2) regress $\Delta Trust$ on the instrumented hindsight bias index and control for for cases per capita in the county of residence as of March 15. Model (3) and (4) regress $\Delta Trust$ on the instrumented hindsight bias index and control for for cases per capita in the county of residence as of April 15. Model (5) and (6) regress $\Delta Trust$ on the instrumented hindsight bias index and and control for the severity a respondent's health has been affected by Covid-19 until April 15 (self-reported). The first stage instruments hindsight bias with the UPDATED FIRST group dummy and the respective control variable (column (2) and (4) and (6)). For the second stage regressions, we report weak-instrument robust Anderson-Rubin confidence sets for the instrumented variable. Robust standard errors are reported in parentheses, p values in braces. The reported F-statistic is the Kleibergen-Paap effective F. The weak identification test reports the traditional Anderson-Rubin test based on the F-stat. The underidentification test is a Lagrange-Multiplier test based on the Kleibergen-Paap rk statistic of whether the equation is identified. The endogeneity test reports a Durbin-Wu-Hausman statistic and tests the null hypothesis whether the endogenous instrumented variable can be treated as exogenous.

D.2.2 Instrumental Variable Assumptions

An empirical challenge is to establish a causal relationship between hindsight bias and the change in trust in government. The degree of hindsight bias is a subject-specific individual characteristic. A correlation between hindsight bias and trust in government may therefore suffer from endogeneity bias since the error term may be correlated.

The random order of preference elicitation that we introduced in the second stage of our survey induces an exogenous variation in the extent of hindsight bias. With the randomization of the order of elicitation, we exogenously vary the degree of hindsight bias. This exogenous variation in hindsight bias allows us to apply a instrumental variable approach with the aim to causally assess the effect of hindsight bias on the change in trust in government. As an instrument, we employ the randomly induced instrument Z which varies the order of elicitation between the two experimental groups, see the causal graph D.1.

Figure D.2: Cumulative Distribution Function, by experimental group assignment



Note: The graph plots the empirical cumulative distribution function separately by experimental group. The CDF of the RECALLED FIRST group is plotted in solid blue, the CDF of the UPDATED FIRST group in dashed red.

The IV approach requires some assumptions (Angrist, Imbens, & Rubin, 1996; Huber & Wüthrich, 2019). **Assumption 1: Relevance.**
First, the instrument must be relevant. The instrument Z must have a causal effect on hindsight bias X.[17] Assumption 1 is empirically testable by inspecting the first stage F-value and the underidentification test which is a Lagrange-Multiplier test based on the Kleibergen-Paap rk statistic of whether the equation is identified. The tests are reported in Table 2. The underidentification test rejects the null that the instrument is not relevant: The test shows that the first stage model is identified ($p < .01$). Regarding the instrument to be weak, we observe the F-statistic to be 9.81, a value below the rule-of-thumb of 12. However, the weak instrument robust inference test (Anderson-Rubin) rejects the null that the coefficient of hindsight bias is equal to zero, and, in addition, that the over-identifying restrictions are valid. Nevertheless, we report weak-instrument robust Anderson-Rubin confidence sets for the linear 2SLS model as recommended by Isaiah, James, and Liyang (2018). These confidence sets are efficient regardless of the strength of the first stage.
Assumption 2: Monotonicity.
A technical assumption is that the effect of the instrument on the endogenous variable is homo-

[17] In formal terms, $E[X|Z = 1] - E[X|Z = 0] \neq 0$.

geneous.[18] Our binary instrument Z should have a monotonic effect on X. To test monotonicity in a setting with a binary instrument Z and a continuous endogenous variable X, the cumulative distribution function of hindsight bias conditional on the instrument status should exhibit no crossings (Angrist & Imbens, 1995). Refer to the Figure D.2 plots the CDF of hindsight bias by experimental group. We observe that the two lines exhibit some crossings at negatives values of hindsight bias. In this range of hindsight bias, however, there are relatively few observations. Indeed, a statistical test reveals that the RECALLED FIRST group actually first order stochastically dominates the UPDATED FIRST group (Somers' D, $p = .002$). The instrument thus impacts hindsight bias monotonically and the monotonicity assumption is sufficiently satisfied.

Assumption 3: Exogeneity.

Exogeneity requires that the instrument Z is exogenous to X and Y.[19] In simple terms, the assumption states that the instrument is as good as randomly assigned. The assumption cannot be empirically tested in a just-identified model. However, in our case, the instrument is indeed randomly assigned and thus exogenous. Therefore, in a successfully conducted experiment, the randomness of Z holds by construction and the exogeneity assumption is satisfied by design.

D.2.3 Discussion of the Exclusion Restriction

Assumption 4: Exclusion restriction.

Our instrumental variable results are valid given the premise of accepting the exclusion restriction. It requires that the instrument and the outcome are independent. The exclusion restriction is violated if being first confronted with the Updated Preference affects the change in trust in government either directly, or through a mechanism other than hindsight bias. If such confounders exist, the exclusion restriction may be violated. Note that we find significant effects in the first and second stage, the reduced form as well as the endogenous OLS model. If the exclusion restriction were fully violated, the responsible mechanism for doing so would have to be able to explain these four empirical findings. Here, we discuss three potential alternative mechanisms, misrepresentation of preferences, recency effects and motivated reasoning, and argue why they cannot explain our empirical findings.

Misrepresentation of Preferences

One potential alternative channel could be misrepresentation of preferences. Respondents might like to appear consistent towards the experimenter (Falk & Zimmermann, 2013). Respondents might thus base their evaluation of trust in government on the policy preferences that we elicited before the elicitation of trust in government.

Before recalling the Original Preference, respondents in the UPDATED FIRST group needed first to report their current view, that is the Updated Preference, which on average is more restrictive than their recall of the past (the Recalled Preference). Respondents in RECALLED FIRST needed first to recall the incentivized Recalled Preference, which tends towards less restrictive policies compared to the Updated Preference, see Figure 2.

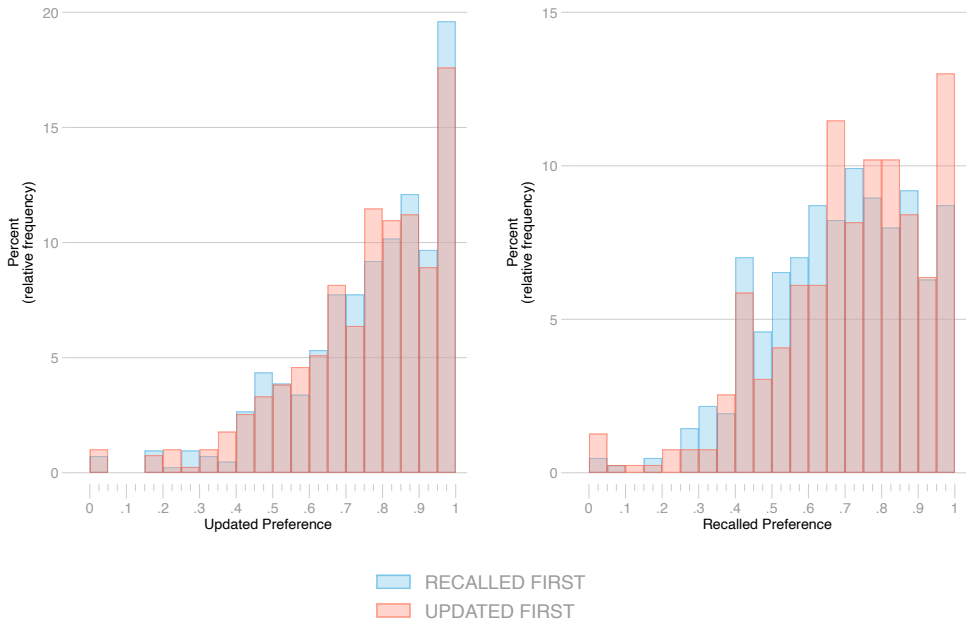
For consistency reasons, respondents in the RECALLED FIRST group may feel compelled to report also a less restrictive (non-incentivized) Updated Preference compared to the UPDATED FIRST group, and in turn, again for consistency reasons, a higher trust in government compared to the UPDATED FIRST group. As a consequence, even without the existence of hindsight bias, we would find lower trust in government in the UPDATED FIRST group.

If this explanation has some merit, the Updated Preference should differ among the two groups. However, we find that the Updated Preference does not significantly differ among the two groups, neither regarding the mean and location (Mean Updated Preference in UPDATED

[18] Formally, $Pr[(X|Z = 1) \geq (X|Z = 0)] = 1$.

[19] Formally, for parametric models the assumption is that $E[v_i|Z_i] = 0$ and $E[u_i|Z_i] = 0$.

Figure D.3: Distribution of the Updated Preference and Recalled Preference by experimental group



Note: The left panel depicts the distribution of the Updated Preference by experimental group. The right panel depicts the distribution of the Recalled Preference by experimental group.

FIRST: .753; Mean Updated Preference in RECALLED FIRST: .767; Welch's unequal variance t test: $p = .326$, MWU test: $p = .215$) nor the distribution (Kolmogorov-Smirnov equality-of-distribution test: $p = .386$, Epps–Singleton test: $p = .689$). See the left panel of Figure D.3 for a plot of the distribution of the Updated Preference by experimental group.[20]

However, we do find that the incentivized Recalled Preference differs among the two groups (Mean Recalled Preference in UPDATED FIRST: .712; Mean Recalled Preference in RECALLED FIRST: .688; Welch's unequal variance t test: $p = .087$, MWU test: $p = .043$). See the right panel of Figure D.3 and Table D.6 in the Appendix. This implies that confronting UPDATED FIRST respondents first with their Updated Preference changed those respondents' recall of the past, but not their current view.[21]

Table D.6: Means of the Preferences and corresponding t and MWU tests

	RECALLED FIRST	UPDATED FIRST	p t test	p *MWU*
Original Preference	.607	.612	.715	.717
Recalled Preference	.688	.712	.087	.043
Updated Preference	.767	.753	.326	.215

Moreover, respondents in the UPDATED FIRST group less often recall their Original Preference correctly: On average, respondents in the UPDATED FIRST group recall 1.28 out of the 4 policy preferences correctly, while respondents in the RECALLED FIRST group achieve

[20] Also note that between the elicitation of the policy preferences and trust in government, we elicited a set of demographic variables. The elicitation of trust in government thus did not immediately follow after the policy preference elicitation.

[21] Note further that the Original Preference elicited on March 15 does expectedly not differ among the two groups (Welch's unequal variance t test: $p = .715$, MWU test: $p = .717$).

1.46 out of 4 items (Welch's unequal variance t test: $p = .034$, MWU test: $p = .023$). Thus, being confronted with the Updated Preference before recalling the Original Preference reduces respondents' bonus payment by 12% on average. Both pieces of evidence are in line with the notion that confronting people with their current view of the world affects their recall of their past view of the world (Kahneman, 2011), and therefore lends support to hindsight bias being the mechanism that leads to a more harshly evaluation of the government.

Recency Effects

Another factor that may be triggered by the manipulation of the question order is a specific form of recency bias: Reporting updated preferences first means that the recall of initial preferences is more recent when reporting trust in government. The treatment thus reduces the interval of time and the number of questions answered between the recall of initial preferences and the reporting of trust. If memory were affected by this form of recency bias, this could be a channel through which the experimental manipulation affects the trust reports.

However, the Recalled Preferences are on average less restrictive than the Updated Preferences. Thus, if recency bias were at work, those who report the Updated Preference first, and the Recalled Preference second, should be more lenient with the government. However, we observe the opposite. Subjects in UPDATED FIRST show a stronger decrease in trust in government, despite reporting the Recalled Preference second. Recency bias can therefore not explain our results. If anything, they work against us and might reduce the observed effect sizes.

Motivated Reasoning

Finally, one could argue that asking for current preference first, then recalled preference, gives participants that are already unhappy with the government an easy opportunity for motivated reasoning (Bénabou & Tirole, 2011, 2016): "My beliefs haven't changed, and therefore I'm justified in being unhappy." The expressed recall thus serves a motivated purpose, caused by the pre-existing unhappiness with the government. Asking for the Updated Preference first could facilitate such motivated reasoning.

Note that this is a reverse causality argument: unhappiness with the government affects beliefs, and the question order then affects how much participants express the (non)-change in beliefs. While such motivated reasoning is consistent with some of our results, they cannot explain the full pattern. In particular, while such reversed causality could explain a correlation between the question order and hindsight bias, as well as a correlation between (the change in) trust in government and measured hindsight bias, it cannot explain the reduced-form effect.

There are two important points here: First, it is important to recall that correlation is not necessarily transitive. If the question order is correlated with hindsight bias, and hindsight bias with the change in trust in government, this does not automatically imply that the question order is also correlated with the change in trust in government. Second, it is decisive to keep track of what is assumed to be exogenous and endogenous in this explanation. In this reverse causality argument, the change in trust in government is assumed to be exogenous and to drive hindsight bias. In addition, there is another effect from the question order (which is also exogenous) on hindsight bias. The fact that treatment assignment is randomized by the computer is important in this context, because a successful randomization implies that there are no systematic differences in exogenous characteristics of participants across treatments. A correlation between the change in trust in government and the question order would therefore imply randomization failure. For some reason more people with an increasing frustration with the government would have been assigned to the UPDATED FIRST treatment.

As we show in the paper in Appendix Table A.3, treatment assignment is not predictive of trust in government on March 15 and neither of any of the four Original Preferences elicited, which speaks against randomization failure.

D.2.4 Discussion of the IV vs. the OLS estimates

Table 2 displays evidence that hindsight bias decreases the change in trust in government through an OLS estimation, which potentially suffers from an endogeneity bias, and an instrumental variable approach, which exogenously instruments hindsight bias, and thus establishes a causal effect given the IV assumptions discussed earlier. By design, the two estimations result in two different estimates: the coefficient for the OLS regression is estimated to be -.30 (column (5)), while the IV coefficient is with -2.05 much larger in magnitude (column (1)).

When comparing these estimates, first note that OLS estimates the average treatment effect and relies on the natural variation in hindsight bias among the entire sample, while IV estimates the local average treatment effect caused by the exogenously imposed variation of hindsight bias in the sample.

That is, the IV coefficient represents the effect of the exogenously imposed portion of hindsight bias on trust in government. The OLS estimate, in contrast, represents the average effect of the general existence of hindsight bias in the population on trust in government. Thus, respondents that are not hindsight-biased, but correctly remember their past policy preference, necessarily dilute the OLS estimate since their hindsight bias index is (close to) zero. For those respondents, the non-existent hindsight bias does not correlate with trust in government. The presence of non-hindsight-biased respondents therefore dilutes the OLS estimate towards zero.

The IV coefficient, however, estimates the effect of hindsight bias on trust in government among the population of respondents who react to the randomly assigned instrument, in our case the randomization of the question order, which exogenously induces hindsight bias. The estimated local average treatment effect thus represents the causal effect of hindsight bias on the change in trust in government. This implies that the two coefficients do not estimate the same thing, and are thus only comparable to a limited extent.

Moreover, the OLS estimate can suffer from endogeneity bias, potentially arising due to either measurement error or omitted variable bias. Measurement error implies that we measure hindsight bias imperfectly by also taking up some noise. When regressing trust in government on hindsight bias, this random error would bias the OLS coefficient towards zero.

Omitted variable bias implies that the regression of trust in government on hindsight bias does not explain the full relationship because other explaining variables were omitted from the regression. If this omitted variable is negatively correlated with hindsight bias, the OLS coefficient can be biased downwards. For example, we find that respondents whose finances were negatively affected by Covid-19 show lower hindsight bias, which is in line with the literature (Pohl & Erdfelder, 2019). Because being financially negatively affected by Covid-19 also reduces trust in government, the exclusion of this variable from the basic regression model thus biases the (negative) OLS coefficient upwards towards zero, while the IV estimate is unaffected by this omitted variable bias.

E Appendix: Heterogeneity by Partisanship

Republicans and Democrats may have differed in how they handled and perceived Covid-19. In the following, we analyze our two main results with regards to possible heterogeneity by party affiliation. We first create three categories out of the five political affiliations we observe: self-reported Democrats and Lean Democrats are classified as Democrats; self-reported Republicans and Lean Republicans are classified as Republicans, and the third category are self-reported Independents/Other.

E.1 Result 1: Existence of hindsight bias

At the beginning of the pandemic, Democratic and Republican preferences were relatively aligned. On March 15, we find that Democrats support significantly more restrictions than Republicans by a relatively small magnitude of .05 on the restrictiveness index, which corresponds to a standardized effect of .21 ($p = .006$). However, this difference increases during the first wave of the pandemic: Democrats' Updated Preference on the restrictiveness index is on average 0.11 larger than Republicans' on April 15, an effect of 0.57 in standardized terms ($p < .001$). Thus, preferences on how to handle Covid-19 drifted apart quite substantially during the first wave of the outbreak: Compared to Republicans, Democrats update their preference towards more restrictions about twice as strongly as Republicans.

Hindsight bias is defined as a shift of the Recalled Preference towards the Updated Preference, see Equation 1. Therefore, whether hindsight bias is heterogeneous by party affiliation is likely to depend on the Updated Preference: respondents who, in retrospect, believe that much stricter restrictions should have been implemented have a greater potential for hindsight bias than someone who, on April 15, still supports the exact same restrictions as on March 15. The greater the belief update, i.e., the distance between the Original Preference and the Updated Preference, the greater the potential for hindsight bias. Thus, Democrats should exhibit more hindsight bias than Republicans. Column (1) in Table E.1 shows regressions of hindsight bias on partisanship, and this is precisely what we find.

E.2 Result 2: Towards a causal effect of hindsight bias on trust in government

Result 2 relies on the *exogenous* part of the hindsight bias induced by the treatment to identify a causal effect of hindsight bias on the change in trust in government. Whether Result 2 is driven by Democrats boils down to a potential heterogeneous treatment effect. Column (2) in Table E.1 provides evidence that treatment effects for Democrats and Republicans are not heterogeneous. Our treatment UPDATED FIRST induces an exogenous portion of hindsight bias at statistically non-different magnitudes across partisanship (see also the discussion regarding the monotonicity of the instrument in Section D.2.2). Controlling for party affiliation in the IV regression yields qualitatively very similar results, see column (3) in Table E.1: the standardized effect of instrumented hindsight bias on the change in trust in government is with $-.65$ roughly the same magnitude as without controlling for partisanship (standardized effect: $-.63$, see Table 2). Because party affiliation is a significant predictor of hindsight bias, column (4) in Table E.1 instruments hindsight bias by the treatment dummy and party affiliation. Column (1) is thus the first stage of column (4). Instrumented hindsight bias significantly reduces trust in government, with a very similar standardized effect size of $-.64$. Column (5) instruments hindsight bias with the treatment dummy, partisanship, and its interaction. Column (2) is thus the first stage regression of column (5). We find that hindsight bias reduces the change in trust in government by .62 standard deviations, a similar effect size. Finally, Column (6) shows the reduced form effect, demonstrating that there are no heterogeneous treatment effects across partisanship on the change in trust in government.

To sum up, Result 2 holds independent of partisanship.

Table E.1: Result 1 and 2 in light of partisanship

	Dependent variable:					
	Hindsight bias		Δ *Trust*			
	Result 1		*Result 2*			
	(1)	(2)	(3)	(4)	(5)	(6)
Republican	-0.03	-0.03	0.00			0.04
	(0.01)	(0.02)	(0.07)			(0.06)
	{.016}	{.062}	{.955}			{.532}
Independent/Other	-0.01	-0.01	0.09			0.02
	(0.02)	(0.02)	(0.06)			(0.07)
	{.619}	{.577}	{.152}			{.735}
UPDATED FIRST (=1)		0.03				-0.12
		(0.02)				(0.05)
		{.055}				{.028}
Republican × UPDATED FIRST (=1)		0.01				0.05
		(0.03)				(0.10)
		{.783}				{.586}
Independent/Other × UPDATED FIRST (=1)		0.01				0.16
		(0.03)				(0.10)
		{.875}				{.104}
Hindsight bias (HB)			-2.09	-2.08	-2.00	
			[...,.04]	[...,-.30]	[...,-.11]	
			{.057}	{.028}	{.046}	
Constant	0.14	0.12	0.11	0.13	0.12	-0.11
	(0.01)	(0.01)	(0.17)	(0.12)	(0.12)	(0.04)
	{.000}	{.000}	{.512}	{.279}	{.307}	{.006}
N	805	805	805	805	805	805
F 1st stage (KP=Eff.)			8.89	4.91	3.14	
Weak identification test (AR)			0.06	0.03	0.04	
Underidentificaton test			0.00	0.00	0.01	
Endogeneity test			0.09	0.04	0.06	
Sargan-Hansen (Overidentification)				0.32	0.37	

Note: The table investigates both our main results regarding partisanship. Model (1) and (2) investigate Result 1, Model (3), (4),(5) and (6) investigate Result 2. Model (1) regresses hindsight bias on party affiliation. Model (2) tests heterogeneity in treatment response by regressing hindsight bias on party affiliation and the treatment dummy and the interaction terms. Model (3) is the instrumental variable regression that regresses $\Delta Trust$ on the instrumented hindsight bias index and controls for party affiliation. Model (4) regresses $\Delta Trust$ on instrumented hindsight bias, and hindsight bias is instrumented through the treatment dummy and party affiliation (which is the regression shown in column (1)). Model (5) also regresses $\Delta Trust$ on instrumented hindsight bias, but hindsight bias is instrumented through the treatment dummy, party affiliation, and the interaction terms (the first stage of model (5) is therefore the regression displayed in column (2)). Model (6) regresses the change in trust in government on the treatment, interacted with the party affiliation—the reduced form effect. The instrumental variable regressions all employ a two-stage least square estimator. Robust standard errors are reported in parentheses. Weak-instrument robust Anderson-Rubin confidence sets are reported in brackets for the instrumented variable in column (3), (4) and (5). p values are reported in braces. The reported F-statistic is the Kleibergen-Paap effective F. The weak identification test reports the traditional Anderson-Rubin test based on the F-stat. The underidentification test is a Lagrange-Multiplier test based on the Kleibergen-Paap rk statistic of whether the equation is identified. The endogeneity test reports a Durbin-Wu-Hausman statistic. The overidentification test reports the p value based on the Sargan-Hansen test.