CESifo WORKING PAPERS

9767
2022

May 2022

Hindsight Bias and Trust in Government: Evidence from the United States

Holger Herz, Deborah Kistler, Christian Zehnder, Christian Zihlmann

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9767*

Hindsight Bias and Trust in Government: Evidence from the United States

Abstract

We empirically assess whether hindsight bias has consequences on how citizens evaluate 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 random assignment to survey structures, allows us to show that hindsight bias causally reduces trust in government.

JEL-Codes: D720, D830, D910.

Keywords: hindsight bias, trust in government, evaluation distortion, biased beliefs.

*Holger Herz**
University of Fribourg / Switzerland
holger.herz@unifr.ch

Deborah Kistler
ETH Zurich / Switzerland
deborah.kistler@econ.gess.ethz.ch

Christian Zehnder
University of Lausanne / Switzerland
christian.zehnder@unil.ch

Christian Zihlmann
University of Fribourg / Switzerland
christian.zihlmann@unifr.ch

*corresponding author

May 20, 2022
This study was approved by the IRB of the University of Lausanne. We thank Björn Bartling, Alain Cohn, Michael Kosfeld, Lydia Mechtenberg and numerous seminar and conference participants for helpful comments and discussions.

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).[1] 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 has real consequences on 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 clear that a world-wide outbreak could not be prevented, policymakers around the globe had to decide about the extent to which they would implement restrictive measures 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 details 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 an impact on their trust in government.[2] 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 e.g. Bargain & Aminjonov, 2020, for public health 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.[3]

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 with exogenously induced variation in hindsight bias. On March 15, 2020, we conducted the first stage of the experiment in which we elicited respondents' preferences over possible policies with different degrees of restrictiveness to fight the pandemic in three different policy dimensions as well as their approval with the restrictiveness of the policies implemented by the government

[1] Hindsight bias exists across various domains and populations (Harley, 2007; Biais & Weber, 2009). The phenomena has been robustly documented in the laboratory (see, e.g., the meta-analysis by Guilbault, Bryant, Brockway, & Posavac, 2004) and in real event settings (Fischhoff & Beyth, 1975; Leary, 1982; Bryant & Brockway, 1997; Bryant & Guilbault, 2002; Danz, Kübler, Mechtenberg, & Schmid, 2015).

[2] For example, Redelmeier and Shafir (2020) have posited at the beginning of the pandemic that hindsight bias may lead to castigating authorities.

[3] See Madarász (2011) for a theoretical development of the argument that hindsight bias negatively affects the evaluation of agents.

at this point in time.[4]

A month later, in mid-April 2020, when the pandemic was at the peak of the first wave[5], 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 right level of restrictive policies that the government should have implemented as of March 15.

We find that respondents' memory is indeed systematically biased. When we asked respondents to state their Recalled Preference in mid-April, they (wrongly) believe, for each of the three policy dimensions as well as for their approval of the restrictiveness of the policies implemented by the government, 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 skewed towards their current Updated Preference.

The presence of hindsight bias suggests that our respondents systematically underestimate how difficult it was to foresee the severity of the crisis when it started. As a consequence, biased respondents might evaluate the government's past actions more negatively than is justified because they incorrectly believe that they supported stricter policies all along and think that government "should have known better". To empirically assess the potential impact of hindsight bias on evaluations of the government, we elicited respondents' self-reported trust in government both on March 15 and a month later.[6] These data allows us to identify the change in trust in government across the two stages of our data collection period 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 strong hindsight bias also tend to experience a decrease in trust in government. Furthermore, our experimental design allows us to go beyond correlational evidence and explore whether the effect is causal. In the second stage of our survey conducted in mid-April, respondents were randomly assigned to 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

[4]There were 3600 confirmed cases and 68 confirmed deaths as of March 15, 2020. All reported case and death numbers in this article are obtained from The New York Times Company (2020) data set.

[5]Cumulative deaths exhibited a 420-fold increase compared to the situation one months earlier. There were 637,056 confirmed cases and 28,582 confirmed deaths as of April 15, 2020.

[6]In both survey stages, the elicitation of trust in government followed after the policy preference elicitation.

"UPDATED FIRST"). Respondents in the second group, in contrast, answered the question in the reversed order ("RECALLED FIRST"). The random assignment to these two groups is helpful because research in psychology shows that explicitly formed outcome knowledge (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 is predicted to increase hindsight bias, because it is expected to shift the Recalled Preference closer to the Updated Preference. This hypothesis is confirmed by our data. We observe that respondents in UPDATED FIRST exhibit significantly larger hindsight bias than those in RECALLED FIRST.

This exogenously induced variation allows us to use our treatments as instruments for hindsight bias. The instrumental variable estimation confirms that hindsight bias causally and significantly reduces trust in government. In standardized terms, a one standard deviation increase in hindsight bias leads to a sizeable decrease of trust in government by .63 standard deviations. The reduced form effect confirms this observation.

Hence, our data provides support for the theoretical argument that hindsight-biased principals inappropriately assess the performance of agents (Camerer, Loewenstein, & Weber, 1989; Frey & Eichenberger, 1991; Madarász, 2011; Schuett & Wagner, 2011): in ex-post evaluations, distorted memories induce hindsight-biased principals to evaluate agents too harshly. To the best of our knowledge, the few existing empirical studies on this topic so far are laboratory experiments demonstrating that hindsight bias correlates with sub-optimally low delegation rates (Danz et al., 2015) and that hindsight bias causally drives excess entry in tournaments (Danz, 2020). Our work is thus the first to provide direct field evidence that hindsight bias causes worse evaluations of agents.

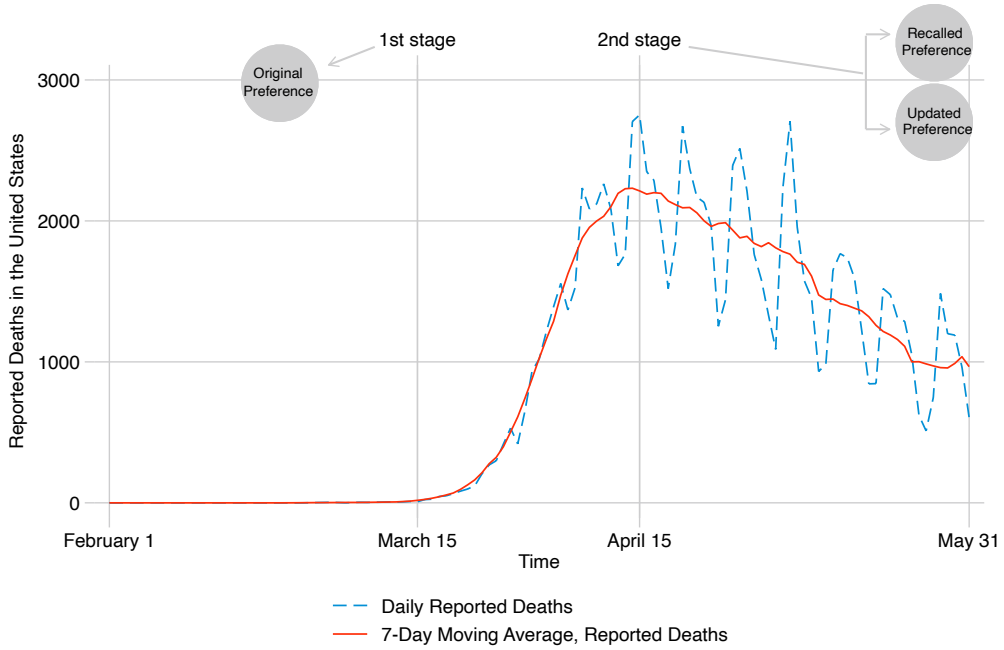
Our findings have several potentially wide ranging implications. A lack of trust in government jeopardizes the state's legitimacy. This is why trustworthy institutions are argued to be a requirement for democracy to work (see, e.g., Acemoglu et al., 2020; Fukuyama, 1995; Putnam, Leonardi, & Nanetti, 1993), which is arguably of particular importance in times of crises. Thus, the 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. Accounting for hindsight bias when setting policy, or developing effective strategies that reduce hindsight bias, appear thus 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. Figure 1 displays the timeline underlying our two-stage design.

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.

The first stage took place on March 15, 2020, when the Covid-19 outbreak in the United States was in its early days (with a total of 3600 confirmed cases and 68 confirmed deaths).[7] We elicited respondents' stated preferences regarding different public policies aiming to contain the pandemic. Specifically, we asked respondents to indicate the level of restrictiveness they considered appropriate in three different policy dimensions: travel restrictions, social distancing restrictions in affected states, and social distancing restrictions nationwide. In addition, we also elicited respondents' degree of approval with the measures taken by the federal government at the time. The precise survey questions and all possible response options are shown in Table 1.

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.[8] As of mid-April, the first wave of the pandemic

[7] The full survey for the first stage can be found in Appendix C.1.

[8] The full survey for the second stage can be found in Appendix C.2. We allowed for a longer response window for the second stage to minimize attrition. For simplicity, we will subsequently refer to April 15 when

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.

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 over the duration of a month had affected respondents' memory of their policy preferences stated on March 15.

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 very same choice options as four weeks earlier. We paid a bonus of 25 cents for every correct recall.[9] 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 in mid-April about the extent

talking about the second stage. Neither the elicited Updated Preferences, nor the Recalled Preferences differ in statistically significant ways across the days.

[9]Respondents were instructed as follows. "On March 15th, we asked you about the policy that you thought should be implemented at that time. Please try to remember the policy that you thought should be implemented at that time. For every correct recall, you will receive a bonus payment of 25 cents."

of restrictions that should have been implemented on March 15, 2020.[10]

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.

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.*

Given that our data consists of four different measures of the preferred restrictiveness of policy, we define one individual measure of hindsight bias that aggregates the degree of hindsight bias across these four measures. To this end, we first min-max normalize all 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.[11]

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.

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

[11] 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 to respondents, but the fourth measures the preference relative to the policies in place as of March 14 (see Table 1), we additionally report in the Appendix the results separately, when only considering the three policy dimensions or when only considering the preference relative to the implemented policies. see Section B.4. The results are qualitatively and quantitatively very similar to the analysis presented in the main body of the paper, which uses the index based on all four measures throughout.

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 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 Appendix Section B.5.

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 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 the second research question.

Our second hypothesis posits that hindsight bias reduces 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.[12] 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 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 less negatively.

Hypothesis 2 (Distortion in ex post evaluations)**.** *Hindsight bias causally decreases trust in government.*

We measured respondents' trust in government on March 15 and a month later so that we can assess the change in trust in government at the individual level. Specifically, we asked respondents[13]: How much of the time do you think you can trust the federal government to do

[12] See also Schuett and Wagner (2011) for an alternative formal model.

[13] This question is adapted from the American National Election Studies https://electionstudies.org/

what is right? Answer options were: "Always", "A lot of the time", "Not very often", "Almost never".

The theoretical mechanism underlying our second hypothesis implies a causal impact of hindsight bias on trust in government. Various endogeneity concerns make it impossible to interpret a potential correlation between our individual measures of hindsight bias and trust in government as supportive evidence for such a causal relation. Thus, to obtain exogenous variation in hindsight bias, we implemented an exogenous between-subject manipulation in the second survey conducted on April 15. We randomized 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 then about their Updated Preference. For respondents in the UPDATED FIRST group, the order of preference elicitation was reversed. As discussed in the introduction, this randomization is expected to create exogenous variation in hindsight bias, because research in psychology suggests that explicitly formed outcome knowledge, in our 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 et al., 1988; Stahlberg & Maass, 1997; Schwarz & Stahlberg, 2003). This exogenous variation thus allows us to use an instrumental variable approach to circumvent the endogeneity 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. We further required an approval rate of at least 95% for past jobs as well as a minimum of 500 completed jobs.

Respondents received USD 1 for completing the first stage. The average completion time was approx. 5 minutes, resulting in an average hourly pay of approx. USD 12. For the second experimental stage, respondents were paid a fixed reward of USD 1.50. The 50% increase in the reward compared to stage 1 was implemented to achieve a high retention rate. In addition, respondents received a variable bonus payment of 25 cents for each correct recall of the Original Preference regarding the four policy dimensions expressed in March. Average completion time in stage 2 was approx. 6 minutes. Together with the variable compensation, this yields an average hourly compensation of approx. USD 18.90.

1027 respondents completed the survey on March 15. Of those, 813 respondents completed

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).

the follow-up survey a month later, yielding a retention rate of roughly 79%.[14] 214 respondents dropped out. The attrition seems to be random with regard 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 4 in the Appendix). We further fail to reject the null that the experimental group assignment is not related to dropping out. Refer to the Appendix for a more detailed analysis.

Out of the 813 respondents who completed both stages, we excluded 8 respondents from the data set due to irregular, non-matching responses with regard to demographic characteristics (those were elicited in both survey stages to check consistency). Therefore, the final sample size amounts to 805 respondents.

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, 2018). Some recent work investigated the demographics of AMT workers (Berinsky, Huber, & Lenz, 2012; Kuziemko et al., 2015; Levay, Freese, & Druckman, 2016). We find very similar patterns. In a nutshell, compared to the U.S. working population, our sample is younger and better educated. Refer to Table 3 in the Appendix, displaying the characteristics of our sample.

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.*

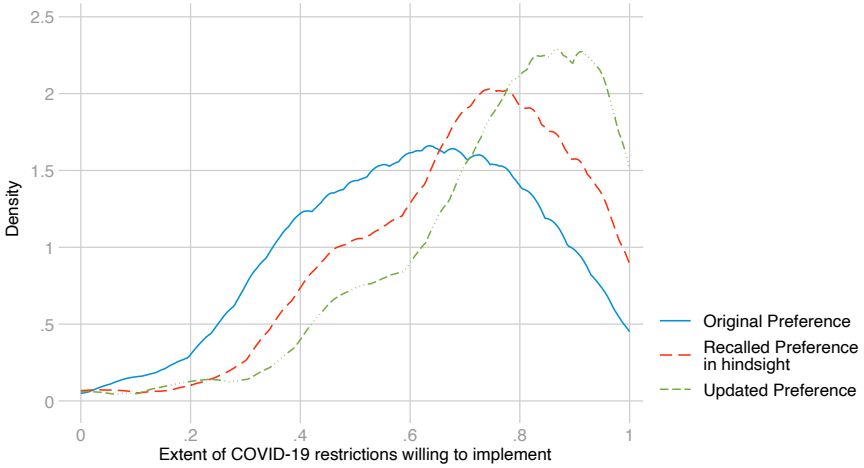
Panel 2a of Figure 2 plots the kernel density estimates of the min-max standardized preference measures regarding the Covid-19 restrictions (a value of 0 represents the least restrictive policy, a value of 1 the most restrictive policy).The solid blue line represents the distribution of the Original Preferences (elicited on March 15). At the onset of the first wave, the average respondent was in favor of implementing policies reflecting a restrictiveness index of about 0.6.

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

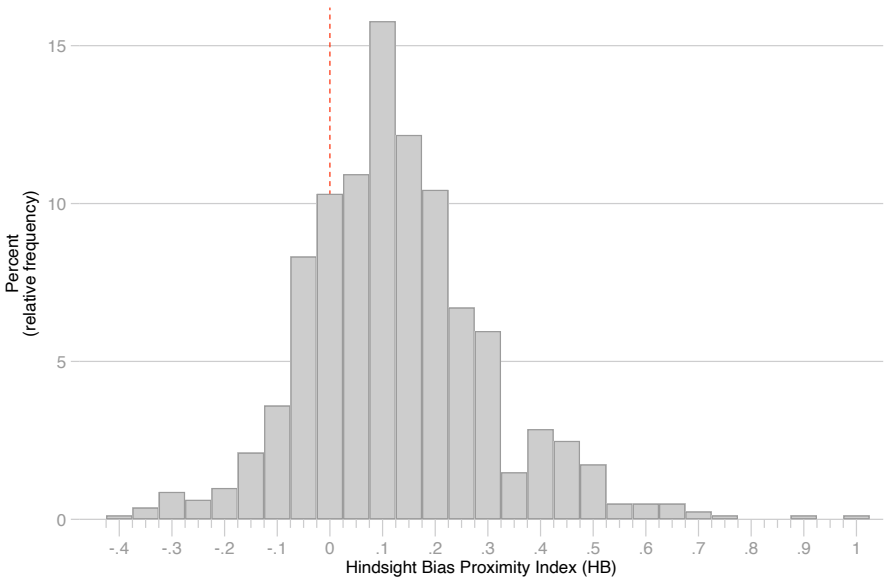
[14] This is a very similar rate as for example in Kuziemko, Norton, Saez, and Stantcheva (2015).

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 bandwidth. 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: $p < .001$.

the beginning of the pandemic. The distribution of these Updated Preferences corresponds to the dash-dotted green line. In retrospect 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 0.82.

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 of 0.7 on average. The Recalled Preference represents a substantial and highly significant departure from the truly expressed Original Preference which corresponded to a restrictiveness index of 0.6 (paired t test: $p < .001$).

Panel 2b plots a histogram of the hindsight bias index as defined in Equation 1, which provides a measure of the magnitude of hindsight bias on the individual level. If hindsight bias was absent in our sample, the index would need to be distributed with mean zero.[15] However, we find that the mean is significantly larger than zero (Student's one-sample t test: $p < .001$).

3.2 Hindsight bias correlates with a reduction 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 5 in the Appendix). Respondents are more likely to reduce their stated trust in government (21%) than to express higher trust in government (8%) (one-sample sign test: $p < .001$; test of proportions: $p < .001$). This implies that, on average, there is a decline in trust in government (Student's one-sample t test: $p < .001$). This decline in trust in government is in line with other public polling.[16]

At the individual level, we find that the change in trust in government correlates with hindsight bias. Respondents who exhibit a stronger degree of hindsight bias tend to also report a decrease in trust in government (Pearson's $r = -.09$, $p = .009$; Spearman's $\rho = -.07$, $p = .037$; Kendall's $\tau_a = -.04$, $p = .037$).[17] This evidence is purely correlational and cannot be

[15] Noteworthy, we do not impose perfect memory on individual level. Yet, the sample population on average should not exhibit a systematic error if hindsight bias is non-existent.

[16] For example, the *Rasmussen Reports daily Presidential Tracking Poll* shows a decrease in approval of the federal government during the month under investigation, refer to https://www.rasmussenreports.com/public_content/politics/trump_administration/trump_approval_index_history, accessed on July 29, 2021.

[17] It is natural to ask whether the change in trust in government is dependent on party affiliation, or by how strongly someone was affected by the pandemic. It turns out that the negative relationship between the

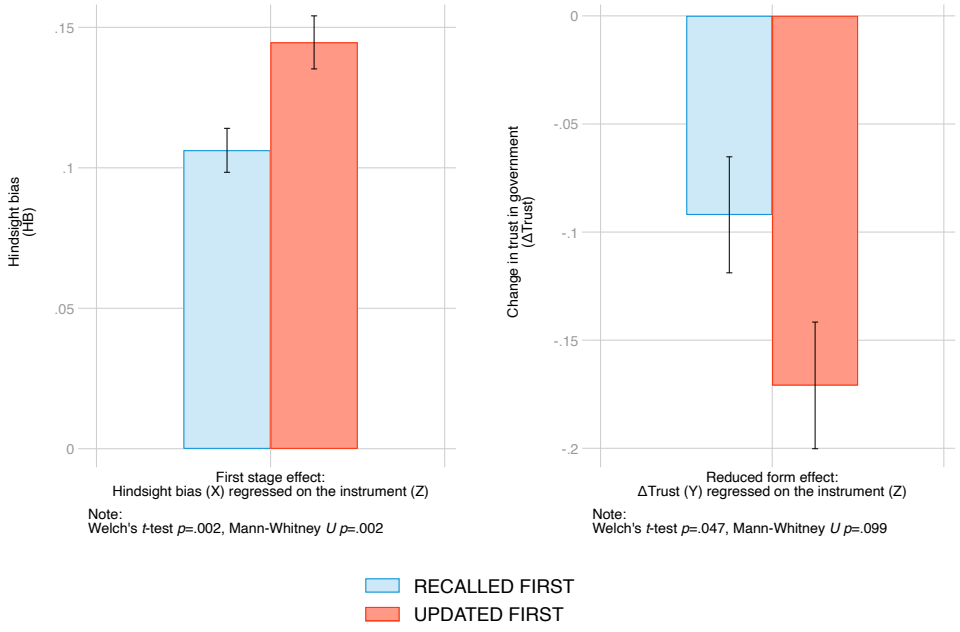
interpreted causally.

3.3 Hindsight bias causally reduces trust in government

We exploit potential exogenous variation in hindsight bias, induced by the random assignment of participants to UPDATED FIRST vs. RECALLED FIRST, to investigate whether there is a causal effect of hindsight bias on the change in trust in government.

Indeed, the mean of the hindsight bias index in the UPDATED FIRST group is 0.145, while it is only 0.106 in the RECALLED FIRST group (see the left panel of Figure 3). Being confronted with the Updated Preference first increases hindsight bias by 36%, a highly significant difference (Welch's unequal variance t test: $p = .002$, MWU test: $p = .002$). The treatment effect appears to be homogeneous. The cumulative distribution function of UPDATED FIRST first-order stochastically dominates the distribution of RECALLED FIRST (Somers' D: $p = .002$), see Figure 6 in the Appendix.

Figure 3: First stage and reduced form effects



Note: The left panel depicts the first stage effect, that is the effect of regressing the hindsight bias index (being the endogenous explanatory variable X) on the experimental group dummy (being the exogenous instrument Z). The right panel displays the reduced form effect, that is the effect of regressing the change in trust in government from March 15 to April 15 (being the outcome variable Y of interest) on the experimental group dummy (being the exogenous instrument Z).

change in trust in government and hindsight bias is robust to controls in a regression framework. Table 6 in the Appendix shows that controlling for i) party affiliation ii) experienced adverse health effects due to Covid-19 and iii) coronavirus cases per capita in the county of residence, does neither turn hindsight bias as a predictor of change in trust in government insignificant nor does it influence its coefficient substantially.

Crucially, respondents who were first confronted with the Updated Preference 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.

In standardized terms, being first confronted with the Updated Preference leads to a .14 standard deviations stronger decrease of trust in government.[18] The reduced form effect provides the first piece of causal evidence that hindsight bias reduces trust in government.

We now turn to an instrumental variable approach.[19] Instrumenting hindsight bias with an exogenous variation provides a solution to the issue that the correlation of hindsight bias with the change in trust in government may suffer from endogeneity bias. The randomly assigned experimental groups thus serve as an exogenous instrument (Z), allowing us to establish and estimate a causal relationship between hindsight bias (X) and the change in trust in government (Y).

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:

$$\Delta Trust_i = \beta_0 + \beta_{1i} HB_i + u_i \tag{2}$$

First stage:

$$HB_i = \gamma_0 + \gamma_{1i} UPDATED\ FIRST_i + v_i \tag{3}$$

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 (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 average treatment effect we investigated previously (the left panel of Figure 3).

Result 2. *Hindsight bias causally decreases trust in government.*

The second stage (column (1)) reports a negative coefficient, meaning that hindsight bias causally reduces trust in government at a statistically significant level ($p = .047$).[20] Regarding

[18] Table 8 in the Appendix shows regressions of the change in trust in government on the two groups by employing a tobit, an ordered probit and a non-parametric kernel estimator. All three estimators confirm the observation that the UPDATED FIRST group exhibits a statistically significant larger reduction in trust in government.

[19] See Section B.3.1 in the Appendix for further elaboration.

[20] Anderson-Rubin weak-instrument robust 95% confidence sets are reported in brackets, as recommended by

effect size, instrumented hindsight bias leads to a decrease in trust in government of .63 standard deviations.

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.

Our analysis shows that ignoring endogeneity concerns by applying OLS leads to understating the relationship between hindsight bias and trust in government. In column (5), we report the endogenous OLS model. 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 than

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.

the 2SLS coefficient. The latter is in principle clean of all omitted variable bias. The 2SLS estimates 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.[21] 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 example, a changing social norm[22] or a random correlation[23] could be potential confounders.

We assess the robustness of our result and find that Result 2 is replicated when using an ordered probit estimator. Column (3) and (4) in Table 2 display results 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 trust in government by .71 standard deviations. As in the 2SLS model, the coefficient is negative and significant ($p = .015$).[24]

We further run the same instrumental variable estimations but include controls for party affiliation and self-reported experienced adverse effects of Covid-19 on own health (see Table 10 in the Appendix), as well as cases per capita in the county of residence (see Table 11 in the Appendix). In all models, the included control variable does not predict at a statistically significant level the change in trust in government. More importantly, the causal effect remains valid. Instrumented hindsight bias reduces trust in government significantly at conventional levels in all models.

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 documents 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

[21] Being aware 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 in the sample. If only a sub-population for which the decrease in trust in government is larger than the average reacts to the randomly assigned instrument, the estimated local average treatment effect will not be generalizable to the entire population.

[22] Suppose that in March 2020, the social norm was to be not too hysterical about Covid-19. People adapted to the social norm and misrepresented their preferences as more optimistic than they actually were. Suppose now that the norm broke down in April 2020 and people expressed their true honest preferences. If the government is a norm regulator, then the decrease in trust in government is to be expected even without hindsight bias.

[23] The existence of hindsight bias has been robustly documented in many contexts (Guilbault et al., 2004). Hindsight bias can thus be expected in any situation. If during the same time period as we capture hindsight bias also trust in government decreases — for any reason — , we would estimate a random non-causal relationship that omits important variables.

[24] Table 9 in the Appendix estimates the instrumental variable models by employing trust in government on April 15 as outcome, conditional on trust in government on March 15. Results are qualitatively very similar.

government. This 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."

Given that trust in government has been found to be a key determinant of citizens' compliance with public health policies (Bargain & Aminjonov, 2020) as well as vaccine acceptance (Lazarus et al., 2021) during the Covid-19 pandemic, our finding that trust in government is undermined by hindsight bias may 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, the existence of hindsight bias and its potential impact on trust in government 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.

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. Psychologists started to study possible strategies to reduce hindsight bias at the individual level, such as ex ante note-taking, with mixed results (Fischhoff, 1977; Davies, 1987). Yet, investigations of debiasing interventions in the domain of public policy evaluation are absent. For example, it would be interesting to study whether direct democracy, and thus more direct involvement in setting policies, impacts the degree of hindsight bias regarding policy.[25] We believe that these are fruitful avenues for future research.

[25]For example, Swiss citizens voted twice in referendums about enacting Covid-19 legislation, putting in place various measures to fight the pandemic.

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.

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.

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

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.

Bryant, F. B., & Brockway, J. H. (1997). Hindsight bias in reaction to the verdict in the oj simpson criminal trial. *Basic and Applied Social Psychology*, *19*(2), 225–241.

Bryant, F. B., & Guilbault, R. L. (2002). " i knew it all along" eventually: The development of hindsight bias in reaction to the clinton impeachment verdict. *Basic and applied social psychology*, *24*(1), 27–41.

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.

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.

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: Effec-

tiveness of foresight-encoding and hindsight-retrieval strategies. *Organizational Behavior and Human Decision Processes*, *40*(1), 50–68.

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.

Fischhoff, B., & Beyth, R. (1975). I knew it would happen: Remembered probabilities of once—future things. *Organizational Behavior and Human Performance*, *13*(1), 1-16.

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

Fukuyama, F. (1995). *Trust : the social virtues and the creation of prosperity.* New York : Free Press, [1995] ©1995.

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.

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

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

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

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.

Leary, M. R. (1982). Hindsight distortion and the 1980 presidential election. *Personality and Social Psychology Bulletin*, *8*(2), 257-263.

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.

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

Putnam, R. D., Leonardi, R., & Nanetti, R. Y. (1993). *Making democracy work*. Princeton University Press.

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

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. (2018, June). *Testing the waters: Behavior across participant pools* (Working Paper No. 24781). National Bureau of Economic Research.

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

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.)

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 3 provides an overview. Our sample consists of slightly more men (56%) compared to the representative U.S. working population (53%). Our participants are on average younger and better educated than the U.S. working population, two well-known features of AMT samples (Levay et al., 2016; Berinsky et al., 2012). Blacks/African-Americans are underrepresented while Asians are over-represented in our sample. Minorities are more common in our sample with 7% of our participants not identifying themselves with any race ("Other"), compared to the representative share of 4% among U.S. workers. These patterns well align with previous literature, see for example Kuziemko et al. (2015). The Top-5 states where our participants reside are exactly the same five states where most of the U.S. working population lives. Our participants 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 participants identify themselves as "Independent" or "Other" more often (18% vs. 11%).

Table 3: 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.

A.2 Attrition

Table 4: 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
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.

As elaborated in Section 2.3, we do not observe significant differences between the 214 participants who dropped out and the 813 participants who completed both stages regarding neither the Original Preference of all four policy dimension nor expressed trust in government.

Continuing this 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 ($p < .001$),

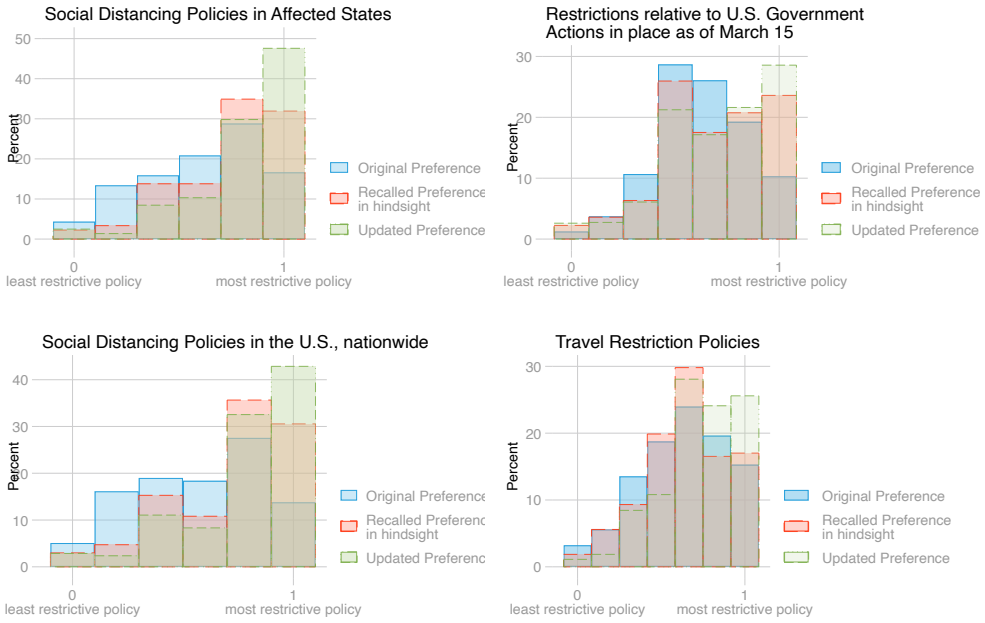
the retention rate significantly increases with the age. Moreover, it seems that "Asian, or Pacific Islanders" ($p < .05$) and "Alaskan Native or American Indian" ($p < .001$) have a higher probability while "Hispanic or Spanish or Latino" have a lower probability ($p < .10$) to finish both survey stages. 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 participants dropped out before the exogenous variation in hindsight bias was induced. These 197 participants did not even start the second survey. 17 participants or about 1.7% of all participants 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.

B Appendix: Results

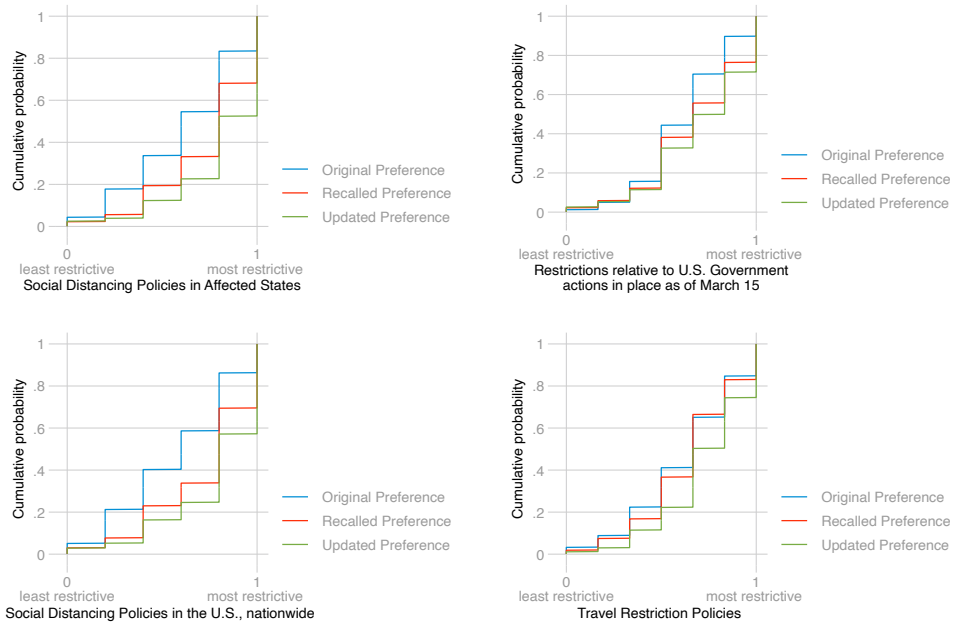
B.1 Existence of hindsight bias during the outbreak of Covid-19

Figure 4: Distribution of the Original Preference, Recalled Preference in hindsight and Updated Preference: Histograms for the four policy dimensions



Note: The graph displays a histogram of the estimates of the extent of Covid-19 restrictions participants are willing to implement for the three elicited preferences, the Original Preference on March 15, the Recalled Preference in hindsight on April 15 and the Updated Preference on April 15, separately for each policy dimension. For all four variables, 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: $p < .001$).

Figure 5: Distribution of the Original Preference, Recalled Preference in hindsight and Updated Preference: Cumulative distribution functions for the four policy dimensions



Note: The graph displays a cumulative distribution function of the estimates of the extent of Covid-19 restrictions participants are willing to implement for the three elicited preferences, the Original Preference on March 15, the Recalled Preference in hindsight on April 15 and the Updated Preference on April 15, separately for each policy dimension. For all four variables, 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: $p < .001$).

B.2 Hindsight bias correlates with a reduction in trust in government

Table 5 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 5: 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 6: 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)
Lean Democrat		-0.07			
		(0.06)			
Other party or Independent		0.08			
		(0.05)			
Lean Republican		-0.03			
		(0.06)			
Republican		0.09			
		(0.07)			
Cases per capita (in county), March 15			172.80		
			(543.70)		
Cases per capita (in county), April 15				0.89	
				(1.29)	
Adversely affected: Own health					0.00
					(0.01)
Constant	-0.09	-0.11	-0.10	-0.09	-0.09
	(0.02)	(0.04)	(0.02)	(0.02)	(0.03)
r2	0.008	0.019	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.

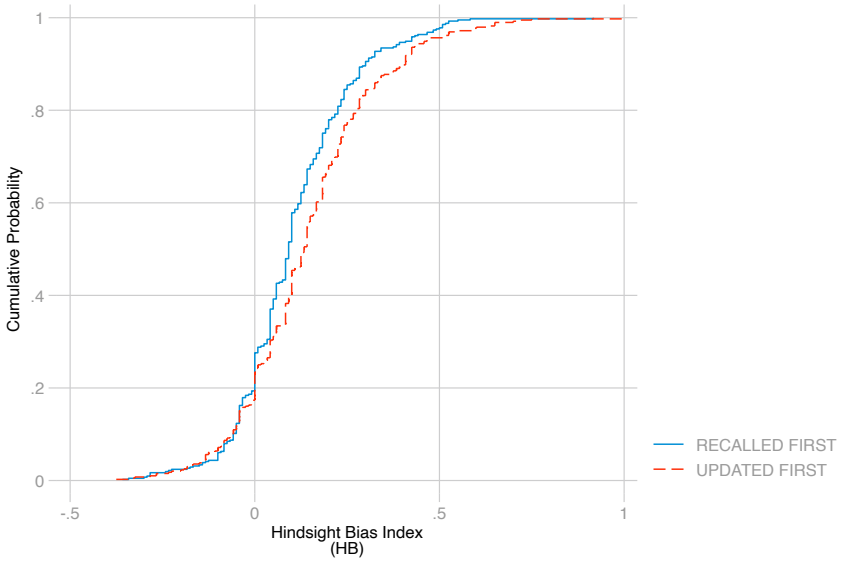
Table 7: Trust in government on April 15 regressed on Trust in government on March 15, hindsight bias and controls

	(1)	(2)	(3)	(4)	(5)
	Trust in government (April 15)				
Trust in government (March 15)	1.58	1.55	1.58	1.58	1.57
	(0.11)	(0.12)	(0.11)	(0.11)	(0.11)
Hindsight Bias	-0.63	-0.57	-0.63	-0.63	-0.61
	(0.25)	(0.25)	(0.25)	(0.25)	(0.25)
Lean Democrat		-0.08			
		(0.12)			
Other party or Independent		0.15			
		(0.12)			
Lean Republican		0.18			
		(0.13)			
Republican		0.48			
		(0.14)			
Cases per capita (in county), March 15			86.39		
			(1276.31)		
Cases per capita (in county), April 15				0.30	
				(3.18)	
Adversely affected: Own health					0.01
					(0.03)
Pseudo r2	0.292	0.302	0.292	0.292	0.292
N	805	805	805	805	805

Note: The table reports ordered probit regressions that investigate the effect of hindsight bias on trust in government on April 15, controlling for the trust in government on March 15. Model (1) is the raw model and regresses Trust in government on April 15 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 participant's health was negatively affected due to Covid-19 as of April 15. Cut-off points are not reported. Robust standard errors reported in parentheses.

B.3 Hindsight bias causally reduces trust in government

Figure 6: 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.

Table 8: The reduced form effect: ΔTrust in government regressed on the experimental groups

	(1) Tobit	(2) Ordered Probit	(3) Kernel
UPDATED FIRST (=1)	-0.08	-0.16	-0.03
	(0.04)	(0.08)	(0.02)
Constant	-0.09		
	(0.03)		
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 9: 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.

Table 10: Change in trust in government regressed on instrumented hindsight bias and control variables

	Dependent variable: $\Delta Trust$				
	2SLS		*2SLS*		*OLS*
	(1)	(2)	(3)	(4)	(5)
	2nd stage	1st stage	2nd stage	1st stage	
		HB		*HB*	
Hindsight bias (HB)	-1.96		-1.94		-0.29
	[-6.56, .12]		[-5.48, -.10]		(0.12)
	{.072}		{.045}		{.016}
UPDATED FIRST (=1)		0.04		0.04	
		(0.01)		(0.01)	
		{.003}		{.003}	
Lean Democrat	-0.09	-0.02			-0.07
	(0.07)	(0.02)			(0.06)
Other party or Independent	0.06	-0.01			0.08
	(0.07)	(0.02)			(0.05)
Lean Republican	-0.09	-0.03			-0.03
	(0.08)	(0.02)			(0.06)
Republican	0.03	-0.04			0.09
	(0.09)	(0.02)			(0.07)
Adversely affected:			-0.02	-0.02	
Own health			(0.02)	(0.00)	
Constant	0.13	0.12	0.16	0.13	-0.11
	(0.18)	(0.01)	(0.17)	(0.01)	(0.04)
N	805	805	805	805	805
F 1st stage (KP=Eff.)	8.95		11.39		
Weak identification test (AR)	0.07		0.05		
Underidentificaton test	0.00		0.00		
Endogeneity test	0.12		0.08		

Note: The table shows the results of two instrumental variable regressions that investigate the effect of hindsight bias on the change in trust in government ($\Delta Trust$), and a accompanying OLS model. Model (1) and (2) report the results from a two-stage least squares estimation, regressing $\Delta Trust$ on the instrumented hindsight bias index and controlling for party affiliation. Model (3) and (4) report the results from a two-stage least squares estimation, regressing $\Delta Trust$ on the instrumented hindsight bias index and controlling for how strongly a participants' health was negatively affected due to Covid-19 as of April 15. The first stage instruments hindsight bias with the UPDATED FIRST group dummy and the respective control variable (column (2) and (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 the second stage regressions, we report weak-instrument robust Anderson-Rubin confidence sets for the instrumented variable. Robust standard errors are reported in column (2), (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.

Table 11: Change in trust in government regressed on instrumented hindsight bias and control variables

	Dependent variable: $\Delta Trust$			
	2SLS		*2SLS*	
	(1)	(2)	(3)	(4)
	2nd stage	1st stage HB	2nd stage	1st stage HB
Hindsight bias (HB)	-2.07		-2.11	
	[-6.41,-.06]		[-6.65,-.05]	
	{.046}		{.046}	
UPDATED FIRST (=1)		0.04		0.04
		(0.01)		(0.01)
		{.002}		{.002}
Cases per capita (in county), March 15	220.41	8.31		
	(768.38)	(243.86)		
Cases per capita (in county), April 15			3.31	1.20
			(2.85)	(0.90)
Constant	0.13	0.11	0.13	0.10
	(0.15)	(0.01)	(0.15)	(0.01)
N	805	805	805	805
F 1st stage (KP=Eff.)	9.81		9.42	
Weak identification test (AR)	0.05		0.05	
Underidentificaton test	0.00		0.00	
Endogeneity test	0.08		0.08	

Note: The table shows the results of two instrumental variable regressions that investigate the effect of hindsight bias on the change in trust in government ($\Delta Trust$). Model (1) and (2) report the results from a two-stage least squares estimation, regressing $\Delta Trust$ on the instrumented hindsight bias index and controlling for for cases per capita in the county of residence as of March 15. Model (3) and (4) report the results from a two-stage least squares estimation, regressing $\Delta Trust$ on the instrumented hindsight bias index and controlling for for cases per capita in the county of residence as of April 15. The first stage instruments hindsight bias with the UPDATED FIRST group dummy and the respective control variable (column (2) and (4)). For the second stage regressions, we report weak-instrument robust Anderson-Rubin confidence sets for the instrumented variable. Robust standard errors are reported in column (2) and (4). 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.

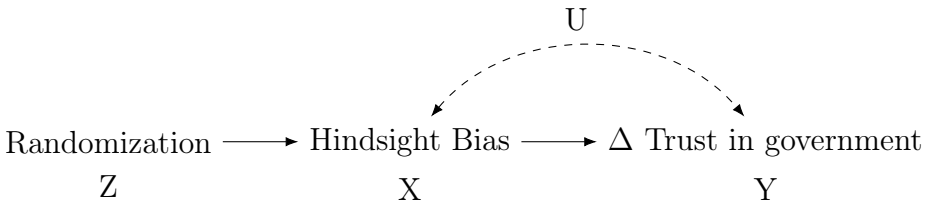
B.3.1 Instrumental Variable Assumptions

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 U 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: In the UPDATED FIRST group, participants were first confronted with their Updated Preference. After that, we asked them about their Recalled Preference. This order was reversed for the RECALLED FIRST group.

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 in Figure 7.

Figure 7: Identification strategy



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.[26] 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

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

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 homogeneous.[27] Our binary instrument Z should have a monotonous 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 6 in the Appendix that 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.[28] 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.

Assumption 4: Exclusion restriction.

The exclusion restriction is a non-testable assumption in just-identified models. It requires that the instrument Z is independent of the change in trust in government Y.[29] The exclusion restriction holds if the instrument, that is the randomization of the order of elicitation of the Recalled Original Preference and the Updated Preference, does not have a direct effect on the change in trust in government. The instrument must have only an indirect effect on the change in trust in government through affecting the amount of hindsight bias one exhibits. While empirically not testable, in our case, we deem it plausible that the exclusion restriction holds. It seems hard to find many plausible cases of how the mere randomization of the elicitation order shall affect the change in trust in government directly other than through hindsight bias.

One example we deem plausible and like to address is misrepresentation of preferences. Participants might like to appear consistent towards the experimenters. Participants might thus anchor their evaluation of trust in government on the policy preferences that we elicited

[27] Formally, $Pr[(X|Z = 1) \geq (X|Z = 0)] = 1$.
[28] Formally, for parametric models the assumption is that $E[v_i|Z_i] = 0$ and $E[u_i|Z_i] = 0$.
[29] Formally, $Y(X, Z(1)) = Y(X, Z(0)) = Y(X)$.

before trust in government.

Participants in the UPDATED FIRST group needed first to report their current view, that is the Updated Preference, which on average is more restrictive than the Recalled Preference. Participants in RECALLED FIRST need to report first the incentivized Recalled Preference, which tends towards less restrictive policies compared to the Updated Preference, see Figure 2.

For consistency reasons, participants 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.

However, if this explanation has some merit, the Updated Preference should differ among the two groups. Importantly, we find that the Updated Preference does not significantly differ among the two groups (Welch's unequal variance t test: $p = .33$).[30] It is only the incentivized Recalled Preference that differs among the two groups, which is much in line with hindsight bias.

B.4 Results separately for explicit policy choices and the policy choice relative to actions taken by the U.S. government

We asked participants about four policy dimensions, refer to Table 1 for an overview. For three policy dimensions — social distancing measures in affected States, social distancing measures nationwide and travel restrictions — participants' had the choice between a selection of explicit policy choices, as summarized in Table 1. For the fourth policy dimension, participants' were requested to indicate whether they would implement less or more restrictive policies than the policies in place as of March 14, facing a relative judgment without explicit policy choices to choose from. In this section, we report all Tables and Figures from the main body separately, first for the preferences regarding the three dimensions with explicit policy choices[31] and then for the preference regarding the approval of the U.S. government measures in place as of March 14.

We find that the two results, the existence of hindsight bias as well as the decrease in trust in government due to hindsight bias, both hold.

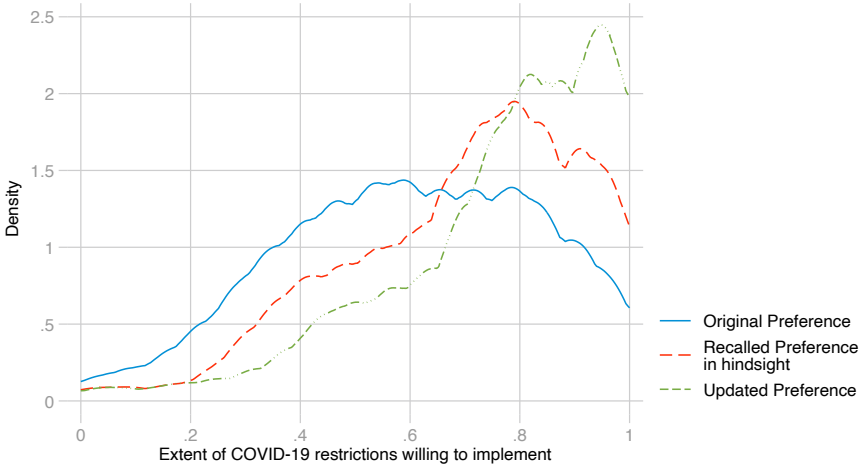
Results for the three policy dimensions with explicit choices

[30] Moreover, note that between the elicitation of the policy preferences and trust in government, we elicited a set of demographic variables. It is thus unlikely that participants anchor trust in government on the policy preferences.

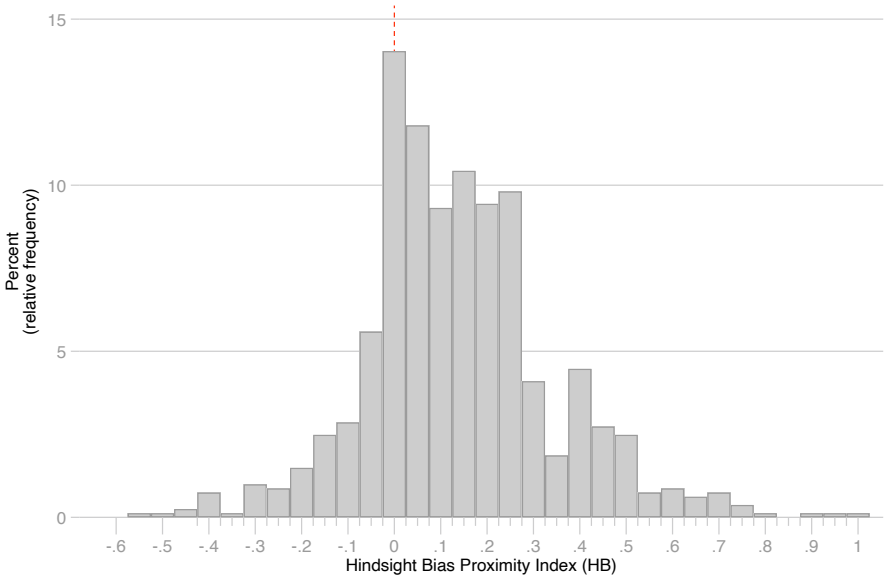
[31] Social distancing measures in affected States, social distancing measures nationwide and travel restrictions.

Figure 8: Existence of hindsight bias

(a) Kernel density estimates of the three preferences

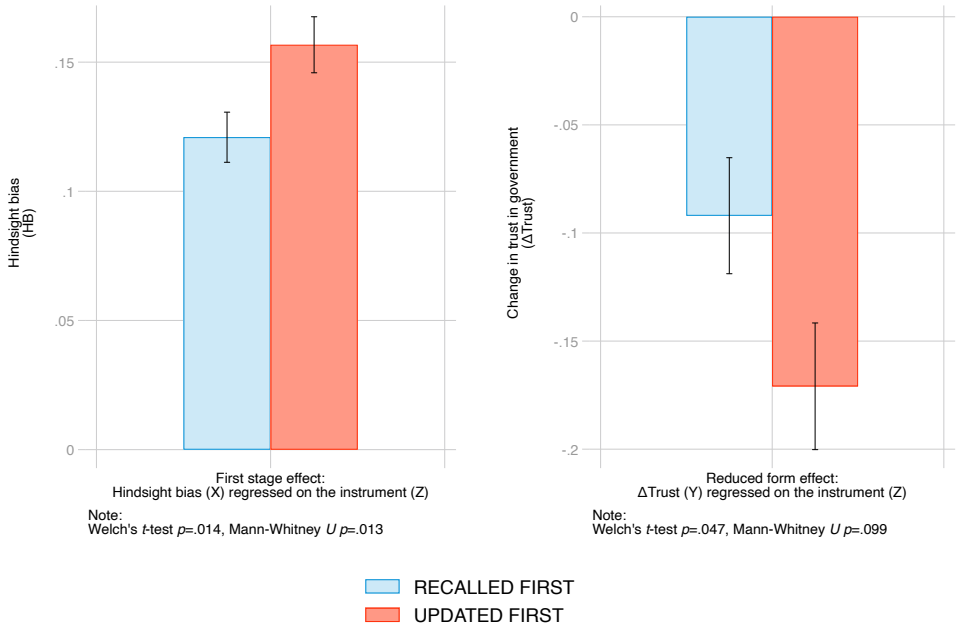


(b) Histogram of the hindsight bias proximity index



Note: Panel 8a displays the kernel density estimates of the extent of Covid-19 restrictions participants 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 bandwidth. 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 8b 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 $\overline{HB} = .14$, Student's one-sample t test: $p < .001$. Sample median $m = .12$, sign test: $p < .001$.

Figure 9: First stage and reduced form effects



Note: The left panel depicts the first stage effect, that is the effect of regressing the hindsight bias index (being the endogenous explanatory variable X) on the experimental group dummy (being the exogenous instrument Z). The right panel displays the reduced form effect, that is the effect of regressing the change in trust in government from March 15 to April 15 (being the outcome variable Y of interest) on the experimental group dummy (being the exogenous instrument Z).

Table 12: Change in trust in government regressed on instrumented hindsight bias

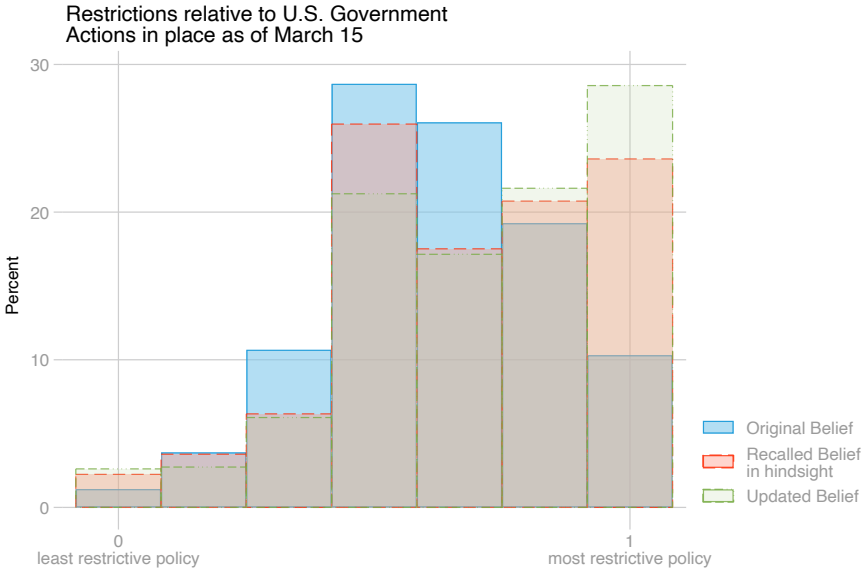
	Dependent variable: $\Delta Trust$			
	2SLS		*Ordered probit*	
	(1)	(2)	(3)	(4)
	2nd stage	1st stage	2nd stage	1st stage
		HB		HB
Hindsight bias (HB)	-2.20		-3.43	
	[...,-.10]		(1.23)	
	{.047}		{.005}	
UPDATED FIRST (=1)		0.04		0.04
		(0.01)		(0.01)
		{.014}		{.014}
Constant	0.17	0.12		0.12
	(0.19)	(0.01)		(0.01)
N	805	805	805	805
F 1st stage (KP=Eff.)	6.05		6.05	
Weak identification test (AR)	0.05		0.05	
Underidentificaton test	0.01		0.01	
Endogeneity test	0.07			
Corr. (e_v, e_u)			0.63	

Note: The table shows the results of two instrumental variable regressions that investigate the effect of hindsight bias on the change in trust in government ($\Delta Trust$). 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. For model (1), we report weak-instrument robust Anderson-Rubin 95% confidence sets for the instrumented variable. Robust standard errors are reported in column (2), (3) and (4). 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.

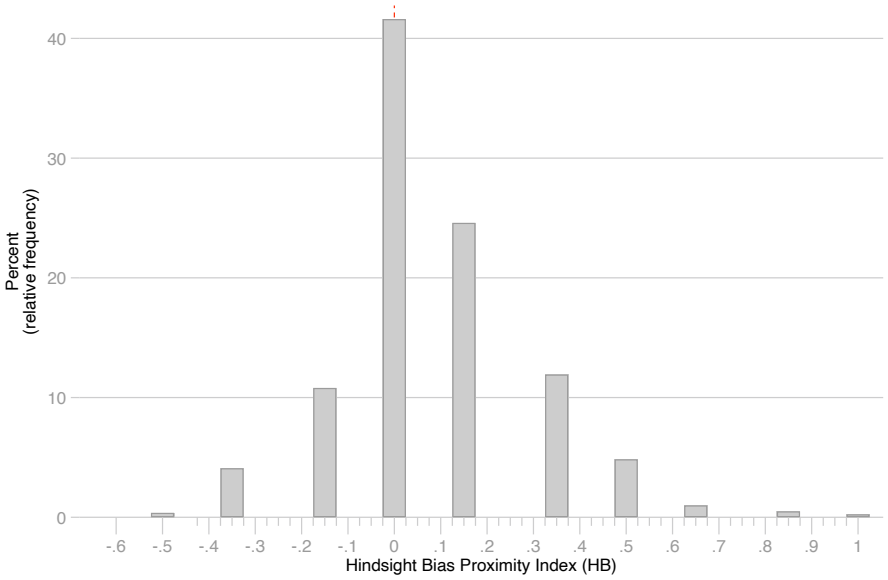
Results for the policy dimension with relative judgment

Figure 10: Existence of hindsight bias

(a) Histogram

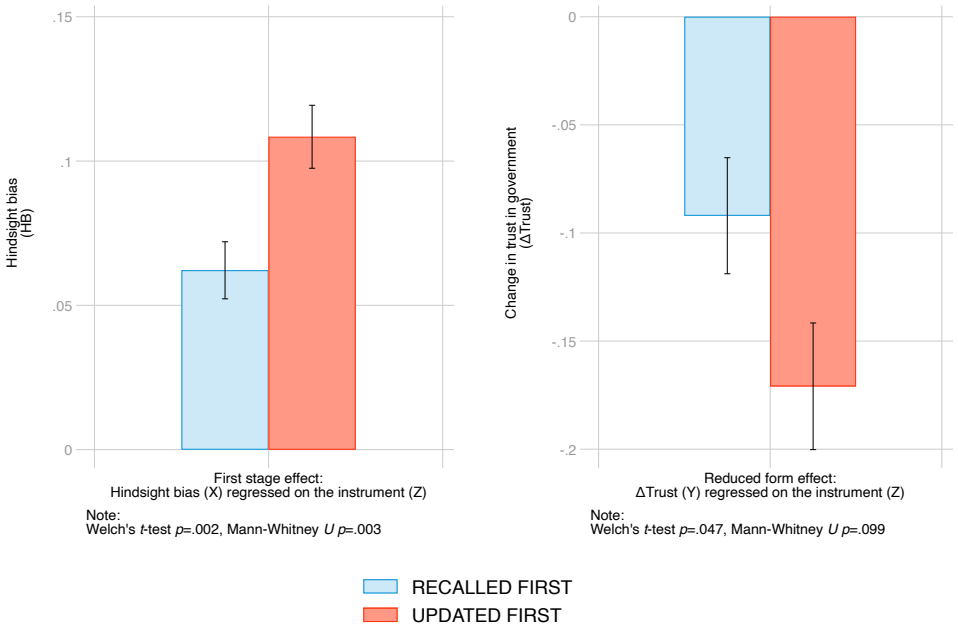


(b) Histogram of the hindsight bias proximity index



Note: Panel 10a displays the histogram of the extent of Covid-19 restrictions participants 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. 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 10b 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 $\overline{HB} = .08$, Student's one-sample t test: $p < .001$. Sample median $m = .00$, sign test: $p < .001$.

Figure 11: First stage and reduced form effects



Note: The left panel depicts the first stage effect, that is the effect of regressing the hindsight bias index (being the endogenous explanatory variable X) on the experimental group dummy (being the exogenous instrument Z). The right panel displays the reduced form effect, that is the effect of regressing the change in trust in government from March 15 to April 15 (being the outcome variable Y of interest) on the experimental group dummy (being the exogenous instrument Z).

Table 13: Change in trust in government regressed on instrumented hindsight bias

	Dependent variable: $\Delta Trust$			
	2SLS		*Ordered probit*	
	(1)	(2)	(3)	(4)
	2nd stage	1st stage	2nd stage	1st stage
		HB		HB
Hindsight bias (HB)	-1.71		-2.80	
	[-5.48,-.08]		(1.15)	
	{.047}		{.015}	
UPDATED FIRST (=1)		0.05		0.05
		(0.01)		(0.01)
		{.002}		{.002}
Constant	0.01	0.06		0.06
	(0.09)	(0.01)		(0.01)
N	805	805	805	805
F 1st stage (KP=Eff.)	9.88		9.88	
Weak identification test (AR)	0.05		0.05	
Underidentificaton test	0.00		0.00	
Endogeneity test	0.06			
Corr. (e_v, e_u)			0.56	

Note: The table shows the results of two instrumental variable regressions that investigate the effect of hindsight bias on the change in trust in government ($\Delta Trust$). 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. For model (1), we report weak-instrument robust Anderson-Rubin 95% confidence sets for the instrumented variable. Robust standard errors are reported in column (2), (3) and (4). 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.

B.5 Results for hindsight bias measured with the shift index

In the following, we report all results of the main body of the paper with hindsight bias measured by the shift index. This shift index is computed as follows (Pohl, 2007):

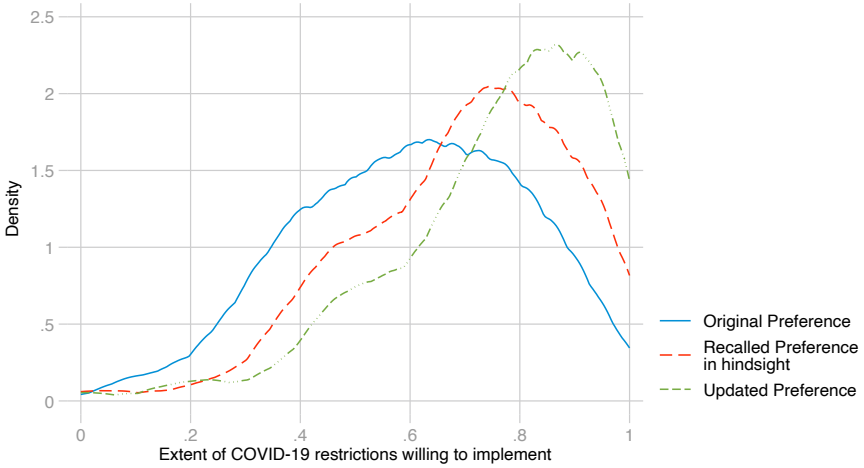
$$HB_{shift} = \begin{cases} \text{Original Preference} - \text{Recalled Preference}, & \text{if Updated Pref.} < \text{Original Pref.} \\ \text{Recalled Preference} - \text{Original Preference}, & \text{if Updated Pref.} > \text{Original Pref.} \end{cases} \tag{4}$$

HB_{shift} measures whether the Recalled Preference shifts towards the Updated Preference. The index is not defined if the Updated Preference exactly equals the Original Preference.[32] In our sample, the index is not defined for 27 participants. Therefore, the sample size for the analysis with the shift index amounts to 778 participants. Hindsight bias exists if the mean of the index is larger than zero.

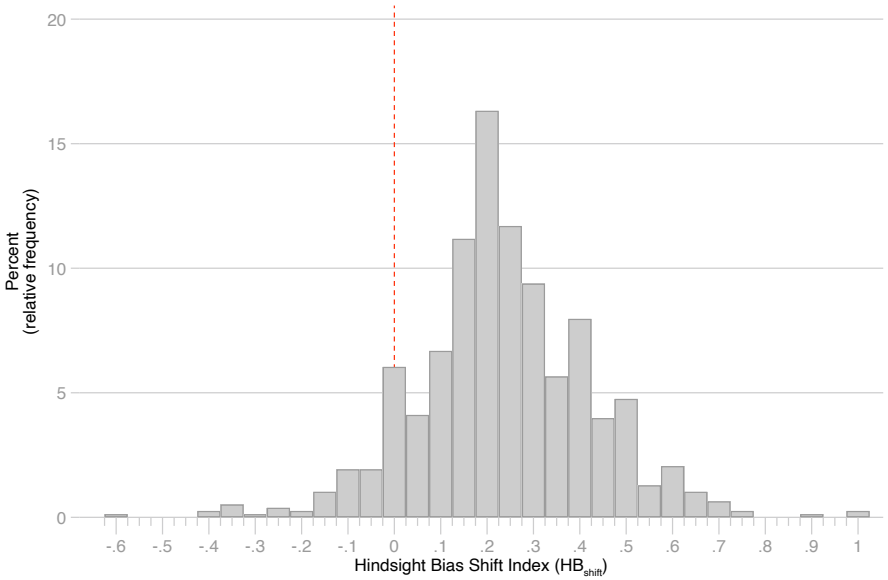
[32] Refer to Pohl (2007) for a discussion why this is reasonable.

Figure 12: Existence of hindsight bias

(a) Kernel density estimates of the three preferences



(b) Histogram of the hindsight bias shift index



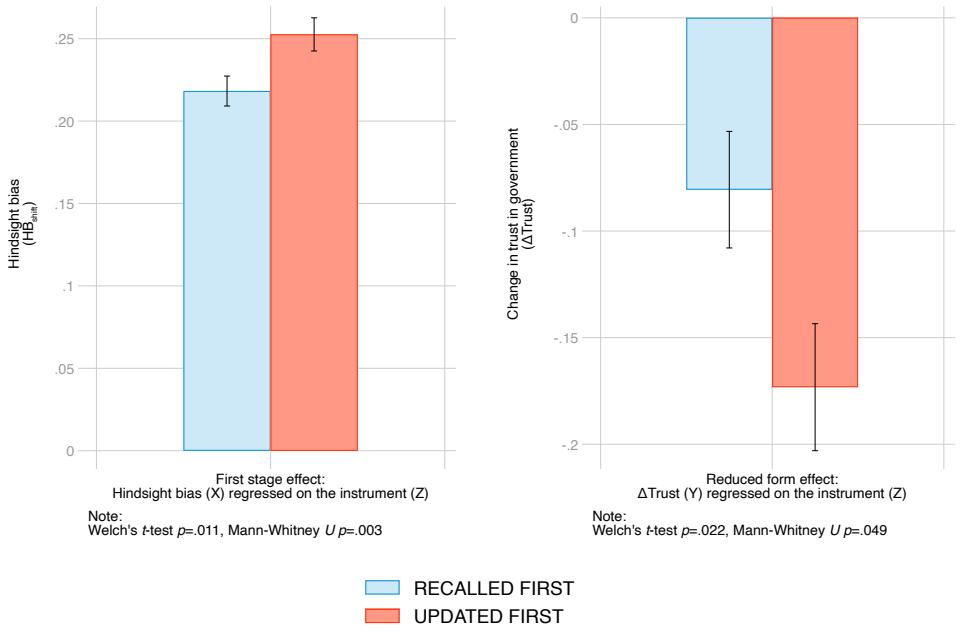
Note: Panel 12a displays the kernel density estimates of the extent of Covid-19 restrictions participants are willing to implement for the three elicited preferences, the Original Preference on March 15, the Recalled Preference on April 15 and the Updated Preference on April 15. We employ the epanechnikov kernel with the optimal bandwidth. 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 12b plots the distribution of the Hindsight Bias Shift Index (HB_{shift}) as defined in Equation 4. One-sample mean and median tests against the theoretical true value of 0 both reject the null at the 0.1%-level. Sample mean $\overline{HB}_{shift} = .24$, Student's one-sample t test: $p < .001$. Sample median $m = .23$, sign test: $p < .001$.

Table 14: 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_{shift})	-2.70		-3.89		-0.16
	[...,-.43]		(1.12)		(0.11)
	{.022}		{.001}		{.145}
UPDATED FIRST (=1)		0.03		0.03	
		(0.01)		(0.01)	
		{.011}		{.011}	
Constant	0.51	0.22		0.22	-0.09
	(0.36)	(0.01)		(0.01)	(0.03)
N	778	778	778	778	778
F 1st stage (KP=Eff.)	6.43		6.43		
Weak iden. test (AR)	0.02		0.02		
Underidentificaton test	0.01		0.01		
Endogeneity test	0.03				
Corr. (e_v, e_u)			0.69		

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 shift 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. 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 13: First stage and reduced form effects



Note: The left panel depicts the first stage effect, that is the effect of regressing the hindsight bias index (being the endogenous explanatory variable X) on the experimental group dummy (being the exogenous instrument Z). The right panel displays the reduced form effect, that is the effect of regressing the change in trust in government from March 15 to April 15 (being the outcome variable Y of interest) on the experimental group dummy (being the exogenous instrument Z).

C Appendix: Instructions

C.1 Stage 1 (conducted on 15 March 2020)

Page 1

Please read the following information carefully. In December 2019, a new Coronavirus disease emerged in Wuhan, China. In the beginning, the disease spread locally. On January 31st, 9,720 people were infected by the virus in China, and 213 had died from the virus. On February 29th, 79,394 people were infected by the virus in China, and 2,838 had died from the virus. In February, the spread of the virus to other countries and continents intensified, reaching the United States and Europe. As of March 14th, 142,539 people had been confirmed infected worldwide, and 5,393 people had died from the virus worldwide. In Europe, 36,264 people had been confirmed infected, and 1,510 people had died from the virus. A particular hotspot was in Italy, where 17,660 people had been confirmed infected, and 1,268 people had died from the virus. In the United States, 1,678 people had been confirmed infected, and 49 people had died from the virus. A particular hotspot was in Washington State, where 607 people had been confirmed infected, and 37 people had died from the virus.

Page 2

In the following, we will present you with potential policies that the U.S. government can implement to combat the Coronavirus. Given the current situation in the U.S., you are requested to select the policy that you think should now be implemented in the U.S.

Page 3

Please consider policies specifying varying degrees of travel restrictions. The policies are ordered from being the least to the most restrictive.

Please choose the policy that should, according to your opinion, now be implemented in the U.S. (because the policies build on each other, and the more restrictive policies always include the measures of the less restrictive policies, you only need to choose one).

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. In addition to 3): Banning flights between the U.S. and Europe and the U.S. and China, except for U.S. citizens and permanent residents.

5. In addition to 3): Close borders to end all international travel, except for U.S. citizens and permanent residents.
6. In addition to 3) and 5): Banning all interstate travel from and to all states with more than 300 confirmed infected cases (currently: Washington State, California, New York State).
7. In addition to 3) and 5): Banning all interstate travel.

Page 4

Please consider policies specifying varying degrees of social distancing **at the state level**. The policies are ordered from being the least to the most restrictive.

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). (Because the policies build on each other, and the more restrictive policies always include the measures of the less restrictive policies, you only need to choose one).

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 in the state.
5. Close all non-indispensable businesses to the public (everything except groceries, gas stations, pharmacies and banks).
6. Statewide lockdown (everybody self-confines themselves to their homes, independent of symptoms, except for essential grocery shopping and health related needs).

Page 5

Please consider policies specifying varying degrees of social distancing **in the entire U.S.** . The policies are ordered from being the least to the most restrictive.

Please choose the policy that should, according to your opinion, now be implemented in the entire U.S. (because the policies build on each other, and the more restrictive policies always include the measures of the less restrictive policies, you only need to choose one).

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 in the country.

5. Close all non-indispensable businesses to the public (everything except groceries, gas stations, pharmacies and banks).
6. Nationwide lockdown (everybody self-confines themselves to their homes, independent of symptoms, except for essential grocery shopping and health related needs).

Page 6

Do you think that the actions taken by the U.S. government regarding the Coronavirus pandemic as of March 14th are...? Please indicate on a scale from 1 to 7, with 1=far too unrestrictive, 4=appropriate, and 7=far too restrictive.

Page 7

- What is your gender?
- Which category includes your age? Choices: 20 or younger, 21-29, 30-39, 40-49, 50-59, 60-69, 70 or older.
- What is the highest level of education you have completed or the highest degree you have received? Choices: Less than high school degree, High school degree or equivalent (e.g. GED), Some college but no degree, Associate degree, Bachelor degree, Graduate degree (e.g. Master degree).
- What race / ethnicity best describes you? Choices: American Indian or Alaskan Native, Asian or Pacific Islander, Black or African American, Hispanic or Spanish or Latino, White or Caucasian, Native American, Other or none of the listed.
- In what state do you currently reside?

Page 8

- In general, which source do you rely on the most for news about politics and current events (e.g. CNN, Fox News, NY Post, NY Times, USA Today, etc.)? Choices: Free form response field.
- In politics today, do you consider yourself a... Choices: Democrat, Lean Democrat, Lean Republican, Republican, Independent/Other.
- How much of the time do you think you can trust the federal government in Washington DC to do what is right? Choices: Always, A lot of the time, Not very often, Almost never.

Page 9

You're done. Thank you for your participation! Do not forget to click the "Submit HIT" button below - this will submit the HIT on Mturk.

C.2 Stage 2 (conducted from 13 April to 16 April 2022)

Page 1

Please read the following information carefully. In December 2019, a new Coronavirus disease emerged in Wuhan, China. In the beginning, the disease spread locally. On January 31st, 9,720 people were infected by the virus in China, and 213 had died from the virus. On February 29th, 79,394 people were infected by the virus in China, and 2,838 had died from the virus. In February, the spread of the virus to other countries and continents intensified, reaching the United States and Europe. As of April 12, approximately 1,734,000 people had been confirmed infected worldwide, and approximately 108,000 people had died from the virus worldwide. In the United States, 555,371 people had been confirmed infected, and 22,056 people had died from the virus.

Page 2

This survey has three parts. In part 1 and 2, we will ask you about potential policies that the U.S. government could have implemented to combat the Coronavirus. Part 3 is a short demographic and opinion survey.

Page 3

Please consider the situation 4 weeks ago: As of March 14th, 142,539 people had been confirmed infected worldwide, and 5,393 people had died from the virus worldwide. In Europe, 36,264 people had been confirmed infected, and 1,510 people had died from the virus. In the United States, 1,678 people had been confirmed infected, and 49 people had died from the virus.

[NOTE: The order for the treatment UPDATED FIRST was reversed. Participants first faced Page 10 to 14, then Page 9 and afterwards Page 4 to 8.]

Page 4 (RECALLED FIRST)

On March 15th, we asked you about the policy that you thought should be implemented at that time. Please try to remember the policy that **you thought should be implemented at that time**. For every correct recall, you will receive a **bonus payment of 25 cents**.

Page 5 (RECALLED FIRST)

Please consider the following policies specifying varying degrees of travel restrictions.

Try to remember the policy that you thought should be implemented 4 weeks ago. For a correct recall, you will receive a bonus payment of 25 cents. Because the policies build on each other, and the more restrictive policies always include the measures of the less restrictive policies, you only need to choose one.

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. In addition to 3): Banning flights between the U.S. and Europe and the U.S. and China, except for U.S. citizens and permanent residents.
5. In addition to 3): Close borders to end all international travel, except for U.S. citizens and permanent residents.
6. In addition to 3) and 5): Banning all interstate travel from and to all states with more than 300 confirmed infected cases (currently: Washington State, California, New York State).
7. In addition to 3) and 5): Banning all interstate travel.

Page 6 (RECALLED FIRST)

Please consider the following policies specifying varying degrees of social distancing at the **state level**.

Try to remember the policy that you thought should be implemented 4 weeks ago in states with 300 or more cases (at that time: Washington State, California, New York State). For a correct recall, you will receive a bonus payment of 25 cents.

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 in the state.
5. Close all non-indispensable businesses to the public (everything except groceries, gas stations, pharmacies and banks).
6. Statewide lockdown (everybody self-confines themselves to their homes, independent of symptoms, except for essential grocery shopping and health related needs).

Page 7 (RECALLED FIRST)

Please consider the following policies specifying varying degrees of social distancing **in the entire U.S**.

Try to remember the policy that you thought should be implemented 4 weeks ago in the entire U.S. For a correct recall, you will receive a bonus payment of 25 cents. Because the policies build on each other, and the more restrictive policies always include the measures of the less restrictive policies, you only need to choose one.

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 in the country.
5. Close all non-indispensable businesses to the public (everything except groceries, gas stations, pharmacies and banks).
6. Nationwide lockdown (everybody self-confines themselves to their homes, independent of symptoms, except for essential grocery shopping and health related needs).

Page 8 (RECALLED FIRST)

4 weeks ago, we also asked you about what you thought about the appropriateness of the actions taken by the U.S. government regarding the Coronavirus pandemic at that time. Please try to remember these actions and what you thought about the appropriateness of these actions at that time. For a correct recall, you will receive a bonus payment of 25 cents.

4 weeks ago, I thought that the actions taken by the U.S. government regarding the Coronavirus pandemic were... (1=far too unrestrictive, 4=appropriate, and 7=far too restrictive).

Page 9 (RECALLED FIRST)

Part 1 of the survey is over. You will now continue with the second part.

Page 10 (RECALLED FIRST)

In this part, please select the policy that **you today think** should have been implemented in the U.S. 4 weeks ago.

Page 11 (RECALLED FIRST)

Please consider the following policies specifying varying degrees of travel restrictions.

As of today, please select the policy that you think should have been implemented in the U.S. 4 weeks ago. Because the policies build on each other, and the more restrictive policies always include the measures of the less restrictive policies, you only need to choose one.

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. In addition to 3): Banning flights between the U.S. and Europe and the U.S. and China, except for U.S. citizens and permanent residents.
5. In addition to 3): Close borders to end all international travel, except for U.S. citizens and permanent residents.
6. In addition to 3) and 5): Banning all interstate travel from and to all states with more than 300 confirmed infected cases (currently: Washington State, California, New York State).
7. In addition to 3) and 5): Banning all interstate travel.

Page 12 (RECALLED FIRST)

Please consider the following policies specifying varying degrees of social distancing at the **state level**.

As of today, please select the policy that you think should have been implemented in the U.S. 4 weeks ago in states with 300 or more cases at that time (at that time: Washington State, California, New York State). Because the policies build on each other, and the more restrictive policies always include the measures of the less restrictive policies, you only need to choose one.

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 in the state.
5. Close all non-indispensable businesses to the public (everything except groceries, gas stations, pharmacies and banks).

6. Statewide lockdown (everybody self-confines themselves to their homes, independent of symptoms, except for essential grocery shopping and health related needs).

Page 13 (RECALLED FIRST)

Please consider the following policies specifying varying degrees of social distancing **in the entire U.S.**.

As of today, please select the policy that you think should have been implemented in the entire U.S. 4 weeks ago. Because the policies build on each other, and the more restrictive policies always include the measures of the less restrictive policies, you only need to choose one.

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 in the country.
5. Close all non-indispensable businesses to the public (everything except groceries, gas stations, pharmacies and banks).
6. Nationwide lockdown (everybody self-confines themselves to their homes, independent of symptoms, except for essential grocery shopping and health related needs).

Page 14 (RECALLED FIRST)

Please try to remember the actions taken by the U.S. government regarding the Coronavirus pandemic as of March 14th. Today, I think that the actions taken by the U.S. government 4 weeks ago regarding the Coronavirus pandemic were... (1=far too unrestrictive, 4=appropriate, and 7=far too restrictive).

Page 15

Thank you for your responses. Part 2 of the survey is over. We will now ask you some demographic questions as well as an opinion survey before this study ends.

Page 16

- What is your gender?
- Which category includes your age? Choices: 20 or younger, 21-29, 30-39, 40-49, 50-59, 60-69, 70 or older.
- In what state do you currently reside?
- In what county (or equivalent) do you currently reside?

Page 17

- How much of the time do you think you can trust the federal government in Washington DC to do what is right? Choices: Always, A lot of the time, Not very often, Almost never.

Page 18

- Does or did the coronavirus affect you own health? 7-point Likert: 1 not all all; 7 very severely.
- Does or did the coronavirus affect the health of close friends or members of your family? 7-point Likert: 1 not all all; 7 very severely.
- Does or did the coronavirus affect the health of close friends or members of your family? 7-point Likert: 1 not all all; 7 very severely.
- Does or did the coronavirus affect the health of anybody you know? 7-point Likert: 1 not all all; 7 very severely.
- How does or did the coronavirus crisis affect your household income (incl. your own)? 7-point Likert: 1 very positively; 7 very negatively.
- If the coronavirus negatively affected your personal income, please specify the type of impact. Choices: N/A or no negative impact, job loss, reduced self-employment income, furlough, Other.
- How many people do you know whose household income was negatively affected by the coronavirus? 7-point Likert: 1 no one; 7 a large number of people. crisis?
- What do you think will be the effect of the coronavirus crisis on your household income in the future? 7-point Likert: 1 very positively; 7 very negatively.

Page 18

You're done. Thank you for your participation!

You've earned a bonus payment of $ AB.CD because you correctly remembered X out of 4 of your responses from the survey conducted 4 weeks ago. Do not forget to click the "Submit HIT" button below - this will submit the HIT on Mturk.