Erroneous Beliefs and Political Approval: Evidence from the COVID-19 Pandemic *

Matthew Lilley† and Brian Wheaton‡

June 19, 2023

Abstract

Are politicians rewarded for good performance, as implied by retrospective models of voting? This requires that public perceptions of performance are accurate. Examining the case of the COVID-19 pandemic, we conduct an incentivized survey asking respondents how pairs of states have performed relative to one another in terms of deaths per-capita. We compute the *erroneous* component of beliefs and find that it strongly drives governor approval. This result is robust to instrumenting for the erroneous component of beliefs with the level of attention focused on each state, measured by (pre-COVID) internet search volume. We also find evidence that erroneous beliefs about state performance distort social-distancing behavior, suggesting both that our measure of beliefs is accurate and that erroneous beliefs are costly to society. We replicate our findings in an identical follow-up survey later in the pandemic and in an additional survey with experimental variation.

*We are grateful to Harvard Business School for grant funding and to Robert Barro, Ben Enke, Mo Fiorina, Ed Glaeser, Justin Grimmer, and Gautam Rao for excellent comments and advice. We also thank the participants of the Behavioral Economics workshop at Harvard and the Democracy and Polarization Lab workshop at Stanford. Matthew Lilley and Brian Wheaton have no relevant or material financial interests that relate to the research described in this paper.

†Australian National University, Research School of Economics. Contact: matthew.lilley@anu.edu.au

‡UCLA, Anderson School of Management. Contact: brian.wheaton@anderson.ucla.edu

1 Introduction

The question of whether politicians are rewarded for good performance and penalized for bad performance is a matter of paramount political-economic importance. This question – central to models of retrospective voting – is crucial because the existence of such rewards/penalties may incentivize elected leaders to pursue socially beneficial outcomes, helping ensure the accountability of elected government to its constituents and the healthy functioning of democracy. A large literature - beginning with Key (1966) - has focused on the effects of politician performance on public approval and voting patterns. However, the extent to which politicians can be rewarded or penalized for their performance depends on the accuracy of public perceptions. A government that is able to generate perceptions of good performance despite poor actual performance may be able to evade responsibility for its actions. Accordingly, this paper investigates how *false* beliefs about politician performance affect political approval.

The role of false beliefs is important to consider because it is often difficult for voters to objectively measure performance. First, there are a multitude of dimensions of both the policies pursued by politicians and the outcomes over which they preside – many of which may be difficult to measure in any objective sense. Second, it can be unclear what role politicians have on each of these dimensions. For example, a growing literature studies the extent to which leaders have actual effects on economic growth, and its findings have been mixed. All of these factors may lead to imperfectly-accurate perceptions of performance.

The precise questions that emerge from these observations are (i) whether voters do actually have accurate beliefs about performance, (ii) whether politicians are rewarded for having good outcomes or merely for being perceived as having good outcomes, and (iii) whether inaccurate beliefs yield any cost to society. To answer these questions, we study the first year of the COVID-19 pandemic, which we regard as a setting highly amenable to the investigation of our research questions. During the early stages of the pandemic, the entire apparatus of state government shifted its priorities toward managing and mitigating COVID-19. Plentiful data on per-capita COVID-19 cases, testing, and deaths was available at the state level (and finer geographies) on a daily basis, and there was substantial variation across states in these outcomes. Governors possessed an extraordinarily wide degree of latitude to implement policy responses of their choosing, with comparatively little encumbrance from legislatures. Meanwhile, they also became the highly-visible public faces of their states' efforts, with some – such as New York's Andrew Cuomo and California's Gavin Newsom – holding daily or weekly COVID-19 briefings. Furthermore, many opinion polls throughout the period focused specifically on public approval of their governor's handling of the COVID-19 pandemic. All this renders the pandemic an ideal setting for studying the accuracy of public perceptions about the performance of their leaders – and the implications of that accuracy.

We conducted an incentivized mTurk survey at the end of July 2020 (during that year's "summer wave" of the pandemic), primarily asking respondents to provide their best guess of how pairs of states performed relative to one another in terms of deaths per-capita. We additionally asked a variety of demographic questions, questions about political identification, and benchmarking questions designed to gauge respondents' perceptions of how well the states should have performed, given pre-existing characteristics such as their population density and setting aside factors of leadership/political competence. The survey consisted of approximately 400 mTurk participants located in the United States, each of whom was compensated a base rate of $1.50 along with a potential incentive bonus for answering the primary questions correctly. We subsequently ran an identical survey three months later, at the end of October 2020 (during the beginning of that year's fall/winter wave) with approximately 200 additional mTurk respondents.

We find that individuals perform better than random guessing in their pairwise comparisons of state performance – but not substantially better. Respondents only correctly guessed which state performed worse 63.4% of the time. Their performance was an identical 63.4% in the case of pairs involving their home state. Respondents tended to think that states like Florida and Texas – which received substantial critical media coverage – performed substantially worse than they actually did. We investigated whether there existed any in-group bias in beliefs, finding at most weak evidence of Republican (Democratic) respondents exhibiting a small-magnitude bias about how well Republican (Democratic) states performed, in relative terms. These results were fairly stable across both the July and October waves of the survey.

We compute the erroneous component of beliefs by subtracting the true death rates from the death rates guessed by respondents. Using data from SafeGraph and the Understanding America Study on social distancing behavior, we show that these erroneous beliefs about state's performance have strong bearing on social distancing behavior. Individuals engage in less social distancing when their state is erroneously perceived to have performed better in terms of COVID-19 deaths. We take this both as evidence that the measured beliefs are real and that erroneous beliefs may distort behavior in a way potentially harmful to society.

Next, we turn to the question of whether politicians are rewarded for good outcomes or merely perceptions of good outcomes. To do this, we regress the erroneous component of individuals' beliefs on state fixed-effects in order to provide a measure of the size and direction of the error in perceptions about how each state is doing. Next, using opinion-polling data from The COVID States Project on state-level approval of governor handling of the pandemic, we regress these measures of approval on the actual state death rate (i.e., the correct component of beliefs) and these aforementioned fixed effects that capture the erroneous component of beliefs. We find that the erroneous component of beliefs drives governor approval at least as strongly as the correct component. This remains true - with equal levels of statistical significance - even after controlling for respondents' perceptions of how well each state should have performed given pre-existing characteristics such as population density, international travel exposure, age distribution, and the like. Additionally, the result is robust to the addition of a broad variety of demographic, pandemic-related, and political control variables, which should both net out effects due to pre-existing attitudes towards politicians and ensure that our results are not biased by perceptions of governor performance on other issues.

As an alternative approach, instead of using the COVID States opinion-polling data on approval of governor COVID-19 handling, we use an identical question internal to the survey and run individual-level regressions of approval on respondents' beliefs about deaths in their home state. This yields the same result – incorrect beliefs strongly affect political approval. Our main specifications jointly use data from both the July and October waves of the survey, but when analyzed separately, the results are stable across waves. Furthermore, when analyzing approval-rating outcomes from September 2021, we find effects of July-October 2020 pandemic beliefs that remain significant and only slightly reduced in size - suggesting that perceptions formed early in the pandemic tend to endure.

To gain more certainty about the causal nature of the result, we utilize a two-stage least-squares instrumental variables (IV) strategy. We instrument the erroneous component of beliefs about each state's death rates with a measure of (pre-COVID) attention paid to each state: Google Search Volume Index (SVI) for each state in 2019. We argue such attention has the capacity to affect the extent of erroneous beliefs about a state's death rates, but should not affect governor pandemic approval ratings either directly or indirectly through other channels. Indeed, the first stage is strong, and the second stage yields results that, again, are strongly statistically significant. Another benefit of our IV approach is that it deals with concerns about sampling error. Because we estimate beliefs about death rates, there is some measurement error, which can cause attenuation bias. IV estimation helps resolve this issue. We present an exercise that allows us to measure the likely magnitude of the attenuation of the OLS estimates, and we show that this matches well with the difference between our OLS and IV effect sizes.

Additionally, to provide yet further evidence on causality, we ran a new survey in December 2020 – this one leveraging experimental variation. First, given that respondents are imperfectly informed about state performance, we elicit governor approval conditional on different hypothetical levels of performance in terms of COVID-19 deaths. We find that conditional governor approval is falling sharply in the hypothetical death rate. Second, we shock respondent beliefs about their state's performance (by eliciting their priors and providing them with the true information), and elicit their ex post governor approval. Exogenously reducing the erroneous component of beliefs about the number of deaths induces higher governor approval ratings. That is, in both experiments, respondents' approval of their governor moves in the expected direction. Taken as a whole, these results suggest complications for retrospective models of voting relying on accurate perceptions of performance.

2 Prior Literature

Our work relates most directly to the broad literature on retrospective voting, which originated over a half-century ago. Key (1966) seminally argued that "voters are not fools" – that is, that they update their beliefs and actions based on government performance, rewarding or punishing politicians accordingly. Key's informal intuition was subsequently formalized in models by Barro (1973) and Ferejohn (1986). In these models, by re-electing high-performing politicians and voting out poorly-performing ones, voters incentivize good performance by politicians (and thus good outcomes). These theories represented an important divergence from the theretofore standard conception of the voter as mostly lacking in information and voting entirely on the basis of promised future political outcomes rather than past performance. On the empirical front, a large subset of this literature, beginning with Kramer (1971), Fair (1978), and Fiorina (1981), has studied whether voters reward or penalize politicians for economic outcomes, which are taken as objective performance indicators.

Later theoretical frameworks enriched the mechanisms underlying retrospective voting. Persson and Tabellini (2002) and Duch and Stevenson (2008) view retrospective voters as learning about incumbent quality through incumbent performance during his/her period in office. Voters then choose between re-electing an incumbent leader of known quality or voting the incumbent out of office and taking a new draw from the quality distribution. Ashworth (2005) models the effects that such a mechanism have on politician decision-making and effort allocation over the course of a career. Recent empirical papers have exploited a variety of natural experiments (e.g., Alt et al. (2011), Gasper and Reeves (2011), Bechtel and Hainmueller (2011), Reeves and Gimpel (2012), Stokes (2016), Heersink et al. (2017), McAllister and bin Oslan (2021), Birch (2023)) and controlled experiments (e.g., Malhotra and Kuo (2008), Malhotra and Margalit (2014), Bechtel and Mannino (2022)).

An important subset of this literature has focused on how behavioral biases and cognitive limitations might interact with the concept of retrospective voting. In a complex world, voters may choose to rely on heuristics rather than evaluate all information carefully. This strand began with the observation (initially made by Kramer (1971), Fair (1978), and Tufte (1978)) that the election-year economy appeared to have larger impacts on voting behavior than conditions in other years of the incumbent's tenure, suggesting a form of availability bias. Huber et al. (2012) and Healy and Lenz (2014) examine this phenomenon in more detail. More generally, it has been argued that voters reward or punish politicians because they are happy or sad for reasons that have nothing to do with incumbent performance, such as foreign economic conditions or football games (Schwarz and Clore (1983), Achen and Bartels (2004), Wolfers (2007), Healy and Malhotra (2010), Gasper and Reeves (2011), Campello and Zucco (2016), Busby et al. (2017)). These findings are often attributed to a combination of behavioral biases and difficulties in attributing responsibility for outcomes. That is not to say that voters behave without any rationality or reason. Other recent evidence shows that the extent to which voters reward or punish a political party in the context of natural disasters depends on the extent to which the party is viewed as having ownership over environmental issues (McAllister and bin Oslan (2021), Birch (2023)). Whereas reflexively punishing incumbents for natural disasters may seem irrational and irrelevant to incumbent performance, selectively punishing those parties that minimized disaster risk is arguably much more sensible.

Our work relates most closely to this strand of the literature which probes the behavioral contours of retrospective voting. One notable contrast is that whereas many of these prior studies focus on the effects of irrelevant outcomes (e.g. football games) on voting, our setting enables us to analyse how voters respond to politician actions that they may legitimately (and plausibly) believe are able to substantially impact outcomes. More distinctively, we are able to precisely distinguish between the true and false components of beliefs held by the public about an important outcome (COVID-19 mortalities in their home state). Exploiting this, we analyze how false beliefs about relevant outcomes affect politician approval. We also contribute by examining a domain (mortalities from the COVID-19 pandemic) distinct from the standard economic outcomes most typically studied in the retrospective voting literature.

3 Data and Empirical Strategy

3.1 mTurk Survey Data

We conducted two separate surveys on Amazon Mechanical Turk. The first elicited beliefs about the pandemic, whereas the second sought to exogenously shock those beliefs.

In the first survey, participants were asked, for 10 randomly-drawn pairs of states, to guess which state in the pair had fared worse up to that point in terms of COVID-19 mortality per capita.[1] The question about each pair was immediately followed-up by a more precise question asking how much worse, as a percentage, they believe their chosen state had fared. Next, for 5 randomly-drawn pairs of states,[2] participants were asked their beliefs about which state would have performed worse (and how much worse) due to pre-existing non-political factors such as population density, population age, presence of international travelers, and anything else they deemed relevant. Also included in the survey were demographic questions on sex, age, race, education, income, and state of residence. Political questions were also asked, including respondents' Presidential election vote in 2016, their party identification, and the level of their approval for their governor's handling of the pandemic. The timing of the initial wave of our survey roughly corresponded to the peak of the summer 2020 wave of COVID-19 cases and deaths. Later, to validate our results and check for consistency, we subsequently ran an identical second wave of the survey in the fall. This timing corresponded to the beginning of the fall/winter 2020 increase in COVID-19 cases and deaths.

In the second survey, we conducted an information-revelation experiment. In the experiment, we randomly assigned participants to either a treatment group, a control group, or a hypothetical group. We asked the control group of participants for their guess of COVID-19 deaths and pandemic employment declines within their home state, followed by a question on the level of their approval of their governor's handling of the pandemic. The treatment group was provided with information on the true figures before being asked about their approval of their governor's handling of the pandemic. The disparity between their priors and the true information induces a shock to the beliefs of respondents in the treatment group, allowing us to discern the effect (if any) of beliefs on governor approval. The hypothetical group was asked a series of hypothetical questions: whether they would approve of their governor's handling of the pandemic if they learned that the true COVID-19 death rate (or the true decline in employment since the start of the pandemic) was X, for a variety of values of X (at least one of which is true). The timing of this second survey was late December 2020.

[1]The first 5 of the 10 pairs were constrained to include the respondent's home state as one of the states in the pair, since individuals may plausibly have a more accurate picture of the pandemic situation in their home state. The following 5 pairs involved comparing two non-home states. For a given survey participant, non-home states were drawn without replacement.

[2]Similarly, the first 2 of the 5 pairs were constrained to include the respondent's home state.

Additional technical details about the surveys are available in Appendix C.

3.2 Social Distancing Data

We primarily draw on social distancing data from two sources: SafeGraph and the USC Understanding America Study. SafeGraph compiles measures of mobility based on cellphone usage and location patterns. Since the early stages of the COVID-19 pandemic, SafeGraph has made a Social Distancing Metrics dataset freely available to academic researchers for the purpose of studying behavior during the pandemic. This dataset contains measures such as time spent at home, time spent outside the home, and distance traveled. It has been widely used in the academic literature on the COVID-19 pandemic. However, the SafeGraph social distancing data is noisy. For example, the median number of total minutes logged per day (at home plus outside the home) varies substantially across places due to factors such as variation in the amount of time people keep location services activated on their phones and variation in restrictions by service providers with regard to how frequently SafeGraph can ping cellphones. To combat these challenges, we leverage the fact that the data is available at the daily level from January 2019 onward; we generate normalized versions of the metrics - computing the change in behavior relative to the pre-COVID year 2019.

We additionally use social distancing data from the USC Understanding America Study. Beginning on March 10, 2020, USC asked their panel of Understanding America Study respondents a series of questions about the COVID-19 pandemic, including a few about social distancing behavior (such as whether the respondent has gone outside in the past 7 days or had any close contact with non-household members over the past 7 days). This individual-level data is freely available to academic researchers.

3.3 Governor Approval and Other Outcomes

Beginning in March 2020, The Covid States Project,[3] a multi-university group of multi-disciplinary researchers, released a variety of periodic reports on the status of the pandemic and related indicators at the state level. Amongst these reports have been state opinion polling data on approval of

[3] `covidstates.org`

governor handling of the pandemic, termed Executive Approval reports by the Project. This data is publicly-available online through the Project's website and we utilize it as our key outcome, using opinion-polling data from their July Wave with our July mTurk survey and opinion-polling data from their October Wave with our October mTurk survey.

As noted, we collect data on approval of governor handling of the COVID-19 pandemic in our mTurk survey as well, to obtain an additional source of this data.

We obtain data on each state's 2016 Presidential Election victor and margin of victory from Dave Leip's Election Atlas.[4].

We also obtain data on the actual number of COVID mortalities over time from the Johns Hopkins Coronavirus Resource Center and data on the intensity of each state's non-pharmaceutical interventions (NPIs) - such as lockdowns, mask mandates, and the like - over time from Oxford University.

3.4 Identification

In order to identify the effects of population-level erroneous beliefs about COVID-19 deaths on governor political approval ratings, we apply a two-step procedure. First, we aggregate individual guesses about relative COVID-19 deaths across different state-pairs to extract estimated average beliefs about each state's death rate. Comparing these with actual death rates yields erroneous beliefs. Second, we use this measure of erroneous beliefs as an explanatory variable of governor pandemic approval (obtained from external opinion polls), controlling for actual death rates and other relevant variables including beliefs regarding differences across states in inherent exposure to the pandemic.

More concretely, we apply the following procedure. We first note that our data on our main survey question is at the state-pair level. Respondents are asked which of State s and State r they believe has experienced a higher COVID-19 death rate and, subsequently, the proportion by which they think deaths are higher in their chosen state. From this, we construct the logarithm of the guessed ratio of the COVID-19 death rates in State s relative to State r (X_{isr}). The logarithmic

[4] https://uselectionatlas.org/

transformation ensures the data is coherently normalised[5] and also ensures the coefficients in our second-step regression are easily interpretable as the effect of a proportional increase in believed deaths.

By subtracting the logarithm of the *true* ratio of COVID-19 death rates in the two states from the aforementioned logged guessed ratio, we can construct a measure of the proportional extent of erroneous beliefs held by individual i about how much higher death rates are in State s relative to r. That is, in logarithms,

$$\log \tilde{X}_{isr} \equiv \log X_{isr} - \log \frac{d_s}{d_r},$$

where d_s and d_r are the respective per capita death rates.

From this data, we wish to extract an estimate of average erroneous component of beliefs about each state's death rate relative to other states. Since each observation pertains to the relative level of deaths in a state pair, for each survey wave we regress respondents' guesses about relative death rates on state indicator variables as follows, in order to estimate state fixed-effects:

$$\log \tilde{X}_{isr} = \gamma_s + \delta_r + u_{isr}, \tag{1}$$

where $\tilde{X}_{isr}$ denotes the erroneous component of the guess of respondent i about the factor by which the death rate of state s exceeds the death rate of state r. δ_r and γ_s denote state fixed-effects for state r and state s respectively, and u_{isr} is the error term. The estimated fixed effect for each state can be extracted as an estimate of beliefs regarding a state's death rate, as desired.

However, an immediate challenge arises. There is no convincing theoretical reason for any particular rotation of any particular observation, namely which state should be considered s and which should be considered r. Further, with a set of fixed effects for s and another for r, this regression generates two separate estimates of beliefs about each state. The two sets of point estimates will not in general be equal, will vary based on arbitrary rotation of datapoints, and it is unclear which (or what combination of them) should be interpreted as beliefs.

Fortunately, there exists a simple fix that works by negating the arbitrary nature of rota-

[5]Note that unlike the level of the ratio, the logarithm preserves the symmetry between comparisons of State s and State r vis a vis State r and State s.

tion decisions. That is, we duplicate each observation in the dataset, representing each observation with both rotations and weighting each by half in our analysis. By construction, this yields $\gamma_s = -\delta_s$.[6] The estimated γ_s vector thus provides a measure of how badly, on average, people think state s is doing in terms of COVID-19 death rates, with a higher value corresponding to perceptions of higher deaths.

We exploit this same procedure to extract another pertinent measure from the survey data - how people believe states should have performed in terms of COVID-19 mortality based on their characteristics (e.g. density, global connectedness) and demographics. We use our survey question on what respondents expected the relative death rate B_{isr} should be in a given state pair, taking into account factors like population density and age, while putting aside factors of political competence. In this case, we re-estimate Equation 1 replacing $\tilde{X}_{isr}$ with B_{isr} and extract the resulting state fixed effect estimates. This yields state-level measures of benchmark expected death rates abstracting from political competence, for each state. Reassuringly, the benchmark estimates we extract have sensible properties. Unsurprisingly, benchmark relative expected death rates are correlated with believed relative COVID-19 death rates, but the two measures appear distinct in practice. In particular, comparing the two survey waves, Wave 1 benchmark estimates predict future beliefs given current beliefs, but current beliefs do not predict future benchmarks given current benchmarks. This is intuitive because state death rates will tend to mean revert towards benchmarks over time (to the degree that benchmark estimates are well calibrated), whereas benchmarks should be a function of fixed characteristics about states. Accordingly, this allows us to separate out how political approval depends on erroneous beliefs regarding performance, from beliefs related to factors outside politician control.

Next, with these state level measures in hand, we regress our outcomes of interest – most notably, political approval – on these estimated fixed-effects and on the natural logarithm of the

[6]To see this, suppose the OLS estimates are $\gamma_s \neq -\delta_s$, and note that $y_{ijk} = -y_{ikj}$. This yields fitted values $\hat{y}_{ijk}$ and $\hat{y}_{ikj}$, and residuals $\hat{u}_{ijk}$ and $\hat{u}_{ikj}$. Consider alternate candidate solution vectors $\tilde{\gamma}_s = (\gamma_s - \delta_s)/2, \tilde{\delta}_s = (\delta_s - \gamma_s)/2$ (such that $\tilde{\gamma}_s = -\tilde{\delta}_s$). This yields fitted values $\hat{\tilde{y}}_{ijk} = (\hat{y}_{ijk} - \hat{y}_{ikj})/2 = -\hat{\tilde{y}}_{ikj}$ and analogously residuals $\hat{\tilde{u}}_{ijk} = (\hat{u}_{ijk} - \hat{u}_{ikj})/2 = -\hat{\tilde{u}}_{ikj}$. For any real scalars $a \neq b, a^2 + b^2 > 2 * [(a+b)/2]^2$, so this constitutes an improvement under the OLS objective function, a contradiction. Note that an analogous argument holds when $y_{ijk} = y_{ikj}$ (in which case, $\gamma_s = \delta_s$) or with an additive constant in either case ($y_{ijk} = a \pm y_{ikj}$).

actual state death rate from COVID-19. That is,

$$Y_s = \alpha + \beta_1 \log DeathsPerMil_s + \beta_2 \cdot \hat{\gamma}_s + \varepsilon_s, \tag{2}$$

where Y_s is a state-level outcome of interest (such as governor approval rate for handling of the pandemic), $DeathsPerMil_s$ is the actual COVID-19 death rate per million population (i.e., the true component of beliefs), $\hat{\gamma}_s$ are the fixed-effects estimated in the preceding regression, and ε_s is the error term. Thus the effect of a 1% increase in actual deaths (holding the error in beliefs constant) on the outcome Y is given by $\beta_1/100$. The effect of a 1% increase in the erroneous component of beliefs (holding actual deaths constant) on the outcome Y is given by $\beta_2/100$. Since we pool the first and second wave of our survey, we add a fixed-effect for the wave and report robust standard errors clustered by state. For robustness, we run additional specifications with an assortment of demographic and political control variables added to the above regression equation. This is done to partial out any correlation of beliefs with these controls, which themselves may plausibly drive political approval. Since political approval may effectively handicap each governor based on how exposed people deem their state as being due to pre-existing factors (largely) outside of government control, we also add the measure of expected death rates abstracting from political competence as a control variable to Equation 2.

Finally, for the governor approval outcome, we run an alternative one-step regression leveraging our internal mTurk survey data on respondent approval of governor handling of the COVID-19 pandemic. Because we have individual-level data on this outcome, it is not necessary to generate state fixed-effects for use in a second-stage regression. We can instead retain the observations involving the respondent's home state and run a version of Equation 2 which, in place of $\hat{\gamma}_s$, directly includes the mean log of erroneous component of the respondent's guesses $\tilde{X}_{isr}$ of how much worse his home state, s, has performed relative to some randomly-selected states r. That is,

$$Y_{is} = \alpha + \beta_1 \log DeathsPerMil_s + \beta_2 \overline{\log \tilde{X}_{is}} + \varepsilon_{is}. \tag{3}$$

4 Results

In Table 1, we first report some simple descriptive statistics about the characteristics of our sample. Average age, share male, and median household income of the sample are consistent with the U.S. general population. The sample, however, has a somewhat higher education, share of non-Hispanic whites, and share of liberals/Democrats than the U.S. general population. In certain specifications, we control for these variables in order to ensure that the deviations from representativeness have no effect on our results.

4.1 Accuracy and Bias in Beliefs

Survey respondents correctly guessed which state had performed worse (through the date of the survey) in terms of COVID-19 death rates only 63.4% of the time.[7] Restricting only to state pairs involving the respondent's home leaves this figure almost exactly unchanged (also 63.4%). Respondents beliefs about their home state are no more accurate than their beliefs about randomly-selected other states. Respondents also performed poorly when estimating states' relative per capita death rates, with a mean (absolute) error of 101 log points. For each log point higher that a state's death rate actually is, respondents estimate deaths to be only 0.28 log points higher. While this weak relationship may merely indicate imperfect knowledge, respondent beliefs are also poorly calibrated: for every log point higher respondents predict a state's death rate to be, deaths are only actually 0.38 log points higher.[8]

Figure 1 displays a scatterplot of the relative frequency with which survey respondents guess each state had a higher death rate (than states with which it is being compared in the pairwise questions) against the relative frequency it actually had a higher death rate. This reveals which states actually performed better than respondents believe (those above the 45-degree reference line) and which states performed worse. As can be seen from the scatterplot, the states with the largest positive gap between actual and perceived performance (i.e., those most erroneously perceived as performing poorly) include Texas and Florida – two states which received particularly

[7]Since completely uninformed random guessing would yield a 50% correct rate, this is consistent, for example, with respondents only truly knowing the answer in 26.8% of cases.

[8]Rational expectations imposes no restriction on the former statistic, but requires a unit change in beliefs to predict a unit difference of the actual outcome. Under perfect knowledge, both coefficients are one.

intense media coverage despite having moderate contemporaneous death rates. States with the largest negative gap between actual and perceived performance (i.e., those most erroneously perceived as performing well) are Rhode Island, Connecticut, and Massachusetts – three New England states which were quite intensely impacted by the first wave of the pandemic but received limited attention in the media relative to New York, which only performed slightly worse but was front-and-center in terms of media coverage in early months of the pandemic.

Given the politically-charged nature of discussions surrounding state performance during the pandemic, one might wonder whether there exists any partisan bias in perceptions of death rates. That is, do Democrats have unjustifiably positive views of the performance of Democratic states while Republicans have unjustifiably positive views of the performance of Republican states? To study this question, we regress the natural logarithm of the erroneous component of individual respondents' believed deaths on an indicator variable for the state's partisan alignment and a "cross-party" indicator variable for whether the state has the opposite political party alignment to the respondent. We also include an indicator variable for whether the respondent is from the state in question.

The regression analysis in Table 2 follows this approach. To begin with, in columns 1-3, we use the party of the governor as our measure of a state's partisan alignment. As seen in column (1), there is strong evidence that average beliefs about Republican-led states' death rates are erroneously high relative to beliefs about Democratic-led states' death rates. Turning to the coefficient on the cross-party indicator variable, there is at most weak evidence of modest partisan in-group bias. Respondents, when considering a state whose governor is of the opposite party affiliation, believe that the state's deaths per-capita are 5.8% higher relative to respondents who share a party affiliation with the governor. This coefficient is small in magnitude and, furthermore, it is only statistically-significant at the 10% level. Columns (2) adds actual log deaths per capita as a control variable, yielding no meaningful change in the estimates of partisan in-group bias. However, adding the control for log deaths causes the coefficient on governor party to become a tightly estimated zero. In other words the negative bias in beliefs about states with Republican governors in column (1) is an artefact of respondents being largely unaware which states had done better, combined with the average Republican-led state then having fewer per capita deaths (as of

2020). Column (3) adds state-by-wave fixed-effects, such that in-group bias is identified only from within-state variation in beliefs; namely, for each state, the partisan difference in beliefs about that state held by respondents. Neither the magnitude or significance of the estimated partisan in-group bias is meaningfully altered. Across specifications, the home state coefficient is close to zero in magnitude; in relative terms respondents are neither meaningfully overly pessimistic or optimistic about their own state's death rate.[9]

One possible concern with this approach is that individuals may not be very familiar with the partisan affiliation of governors outside their own state, but be more familiar with states' voting tendencies in presidential elections. Partisan in-group bias in beliefs about state COVID-19 deaths may accordingly operate on this alternative measure of states' partisanship. Columns (4)-(6) repeat the analysis, defining the state partisan alignment variables according to the state's vote in the 2016 presidential election. Here, there is evidence of modest systematic overestimation of deaths in Republican states even when controlling for actual log deaths per capita. However the estimates of partisan in-group bias all remain close to zero, such that the data provides weak evidence of, at most, modest partisan in-group bias.

4.2 Effects on Social Distancing Behavior

We next examine whether these erroneous beliefs translate into behavioral differences. During the COVID-19 pandemic, efforts to "flatten the curve" of COVID-19 cases by encouraging individuals to spend as much time as possible quarantining at home – as opposed to outside – were central to the public health response. Erroneous beliefs about the intensity of the pandemic might lead to distortions in behavior, potentially inflicting costs upon society. To test this, we run versions of the regression specification described in Equation 2 – in this case, with measures of social distancing as the outcome variables. We use SafeGraph's measure of percentage of time spent at home per day along with measures from the Understanding America Study (UAS) of the share of people in each state who went outside or who had close contact with a non-household member at any time

[9] The higher the population of a state, the higher share of the question pairs about its relative COVID death rate will be by home state respondents. However, since individuals appear unbiased about their own state on average compared to other states, this should not mechanically bias the elicited aggregate beliefs.

in the past 7 days.[10]

Table 3 cycles through the same regression specifications used in the preceding section, now with the median percentage of time at home outcome. The specification in column (1) corresponds to Equation 2. The results show that erroneous beliefs about deaths strongly affect the percentage of time individuals spend at home. In particular, a one standard deviation (70 log point) increase in erroneous believed deaths translates into a (roughly) 4.1 percentage-point greater increase in the percentage of time individuals spent at home (relative to the analogous period in 2019). Consequently, erroneously believing that a state has had fewer deaths than it actually has may lead to a sub-optimal amount of social distancing behavior. In column (2), we add as controls the state's 2016 Trump minus Clinton share, a rolling 7-day average of COVID-19 cases and deaths in the state, and measures of both contemporaneous and average intensity of state lockdowns/NPIs. These controls capture the conjecture that recent case rates, implementation of lockdowns measures, and a state's partisan lean may correlate both with beliefs about the cumulative death rate and with social distancing behavior. Statistical significance of the coefficient on believed deaths is retained. Columns (3) and (4) repeat this analysis with the share of individuals who went outside in the past 7 days (from the UAS data) as their outcome. Here, too, beliefs of a higher death rate (unlike a higher *actual* death rate) translate into less time spent outside. In terms of magnitudes, column (3) suggests that a one standard deviation increase in erroneous believed deaths translates into a 9.3 percentage-point decrease in the share of individuals who went outside in the past 7 days. Columns (5) and (6) find an analogous result for the share of individuals who had close contact with someone outside their household.[11]

4.3 Effects on Political Approval (Observational)

We next turn to the key question of how aggregate beliefs about state performance affect political approval ratings. Table 4 displays versions of the regression specification described in Equation

[10] We use SafeGraph/UAS data for the 30 days following the occurrence of our mTurk survey. That is, we merge observations from the July wave of our survey with corresponding SafeGraph/UAS data on social distancing behavior from the subsequent 30 days after the end of that wave; we merge observations from the October wave of our survey with SafeGraph/UAS data from the subsequent 30 days after the end of that wave.

[11] The Covidstates survey data includes some measures of social distancing as well – whether individuals had recently visited with friends, avoided contact with others, avoided crowded places, or limited contact outside their household. Running analogous regressions with these outcomes again yields very similar results.

2, with state-level average approval ratings of governor handling of the pandemic as the outcome variable. Column (1), however, begins with a univariate regression of governor COVID-19 approval on the log of the death rate. There is evidence of a positive (albeit slightly weak) association between the death rate and governor COVID-19 approval. A one standard deviation (90 log point) increase in deaths roughly translates into a 2.6 percentage-point increase in governor approval. This regression, however, masks a more complex relationship. Column (2) is the specification directly corresponding to Equation 2. It reveals that, if anything, higher erroneously believed deaths are more strongly associated with lower approval than higher true believed deaths. In particular, a one standard deviation (70 log point) increase in erroneous believed deaths translates into a 7.9 percentage-point decrease in governor COVID-19 approval. In other words, the intuitive relationship whereby voters punish politicians for bad outcomes (here, deaths) is driven at least as much by perceptions of the outcome as the actual outcome itself. This suggests potential challenges to ensuring politicians are properly incentivized through public opinion and voting.

Columns (3) repeats the exercise of column (2), with an added control capturing beliefs about benchmark death rates. These benchmarks measure how high a death rate respondents would have expected in each state given its pre-existing characteristics (e.g., population density, population age, exposure to international travelers, etc.), putting aside factors of political competence. This allows us to separate out how political approval depends on erroneous beliefs regarding performance, from beliefs related to factors outside politician control. The addition of this benchmark deaths control increases the coefficient on erroneous beliefs approximately two-fold. The positive coefficient on benchmark deaths is consistent with governors being graded on a curve based on their state's perceived inherent exposure to the pandemic. The overall conclusion is substantively unchanged.

Column (4) adds a variety of state-level control variables to the specification in column (3) - including governor pre-pandemic approval rating, governor party indicator variables, and 2016 Presidential election margin (Trump minus Clinton) interacted with governor party - to capture differences in either beliefs or governor COVID-19 approval driven by a state's partisan tendency or pre-existing idiosyncratic governor characteristics. We also control for the natural logarithm of the state's past-seven-day average of new COVID-19 cases and deaths, in case recent outcomes correlate with beliefs and are also responsible for driving governor approval. The key results are robust to adding these controls, with the estimated effect of beliefs on approval becoming modestly smaller, but more tightly estimated. Finally, in column (5), we add controls for the contemporaneous and average (since the beginning of the pandemic) stringency of state non-pharmaceutical inventions (NPIs), per the Oxford COVID-19 Government Response Tracker database. Results are scarcely changed from the preceding column.

While our regressions cluster errors by state, a concern with any spatial analysis is the possible existence of spatial correlation in the regression variables, which may cause the estimated regression standard errors to be too conservative. This is potentially pertinent for COVID-19 since disease spread is inherently a spatial phenomenon. A standard approach for investigating such concerns is to compare the estimated coefficient against a placebo distribution obtained by randomly re-assigning the variable of interest (here, erroneous beliefs) across states.[12] In our case, such a permutation procedure has the additional benefit of accounting for the fact that our main regressor is generated with estimation error, albeit modest.[13] Accordingly, we perform 1000-repetition permutation tests, randomly re-assigning erroneous beliefs about death rates across states. Comparing the distribution of the placebo regression coefficients that result from these permutations to our actual regression coefficient, we obtain a measure of how likely it is that our result is purely due to chance. Figure 2 plots the result of a permutation test for the specification in column (2) of Table 4. The red distribution originates from regressing governor COVID-19 approval on (placebo) erroneous beliefs about deaths. The green distribution results from regressing governor COVID-19 approval on the placebo erroneous beliefs about deaths - orthogonalized against actual deaths. The vertical line represents the true estimate, which is quite extreme relative to the distribution of placebo estimates. Consequently, the result remains strongly significant. Figure 3 repeats this process for the specification in column (5) - the one saturated with control variables. Again, the result remains strongly significant. The implied p-value is considerably less than 0.01 in all cases.

In Table 5, we test whether our measures of erroneous beliefs from July-October 2020 have

[12]Where spatial correlation is a problem, the variance of the placebo distribution will be greater than the classical standard error estimates.

[13]See Appendix B.2.

persistent effects on approval. Replacing the dependent variable with governor COVID-19 approval as of September 2021, we find that higher believed deaths in July-October 2020 continue to predict lower governor COVID-19 approval approximately a year later. The estimated coefficients are very similar, albeit slightly smaller in magnitude, than the contemporaneous effect of erroneous beliefs studied in Table 4. The seemingly highly persistent nature of the negative effect of beliefs on governor approval suggests that perceptions formed early in the pandemic tend to endure. This in turn raises the possibility of a particularly important role for initial public impressions.

Next, we exploit the individual-level data on governor COVID-19 approval that we collected in our mTurk survey in order to estimate whether beliefs predict governor approval at the individual level, as per Equation 3. Since we run these regressions at the individual level, they focus on beliefs about one's home state. As noted previously, the accuracy of respondents' beliefs about death rates in their home states is identical to their accuracy about other states, so there is little reason to expect different results. As shown in Table 6, using this alternative approach yields results that differ only slightly from the main specifications. In particular, our estimate of the effect of erroneous believed deaths on governor COVID-19 approval remains statistically negative and of very similar magnitude. The individual-level data also allows us to conduct heterogeneity analysis. We were unable to detect differential effects of beliefs on approval by party identification or by whether one's governor is of the opposite party from oneself (results available upon request).

Of course, it must be noted that our measures of individuals' erroneous beliefs about state performance are observational rather than the product of experimental variation. Nonetheless, we believe that making a causal interpretation regarding the impact of beliefs on political approval is sound for two reasons. The most obvious threat to identification is some form of endogeneity or reverse causality. For example, individuals who happen to approve of a state's governor may be more likely to hold excessively optimistic beliefs about the performance of that state. However, if this was the case, then we should expect to observe substantial partisanship in beliefs about states' performance, but as shown above, we find minimal such bias in the data. Furthermore, any endogeneity associated with political affiliations of states and their governors should largely be captured by controlling for political and demographic factors about the state (e.g., the partisan lean of the state and the governor's pre-existing approval rating). As we show, our results are robust to adding these controls, and it is plausible that, conditional on these observables, beliefs are exogenous of unobserved determinants of governor approval.

4.4 IV Effects

Despite these considerations, there is nonetheless a contextual factor that makes IV estimation desirable. Our main variable of interest is estimated from survey data and thus is subject to random measurement error; we care about true mean beliefs γ_s but observe $\hat{\gamma_s} = \gamma_s + \xi_s$.[14] By adding random noise to the key independent variable, this measurement error will tend to weaken the relationship between estimated beliefs and governor approval and cause attenuation bias. Formally, when a single regressor x_j is subject to random measurement error, it is straightforward to show (see Appendix B.1) that

$$\underset{n \to \infty}{p\text{-lim}}\, \hat{\beta}_j^{\,OLS} = \beta_j \left(1 - \frac{\sigma^2_{\xi_j}}{\sigma^2_{\tilde{x}_j}}\right) \tag{4}$$

where x_j is the observed (mismeasured) variable, ξ is the measurement error and the superscript $\sim$ indicates variables that have been residualised against the other controls in the regression.

To this end, we run two-stage least-squares regression specifications that aim to use only exogenous variation in beliefs to identify the effect of beliefs on approval. As well as ameliorating the measurement error bias, this further addresses concerns regarding unforeseen channels through which beliefs are endogenous to politician approval. Specifically, we instrument for the erroneous component of beliefs with a measure of (pre-COVID) attention paid to each state: Google Search Volume Index (SVI) for each state in 2019. In effect, some states are much higher profile and are paid more attention to in national news narratives than others.

In general, a state having a higher profile need neither raise nor lower beliefs about its performance on various metrics, because information can be either positive or negative. In this vein, we argue high levels of attention paid to a state shouldn't affect governor approval either directly or through any indirect channel except by changing beliefs about the level of COVID-19 deaths,

[14] Specifically, recall that the erroneous beliefs variable is constructed from the state fixed-effect point estimates in an OLS regression, which have asymptotically normal errors.

which we assume as the instrumental variables exclusion restriction. However, in the context of COVID, information and news coverage tended to be inherently negative. Accordingly, if individuals focus more of their attention on a certain state relative to others, it would be unsurprising if this negatively affected individuals' beliefs about that state's death rate from COVID, and potentially their perception of political performance in turn. For example, New York tends to overshadow its neighbor states, New Jersey, Connecticut and Massachusetts, and this was evident in relative attention devoted to the states - by the general public and the media - particularly in the early stages of the pandemic. As shown in Section 4.1, New Jersey and Massachusetts ended up exhibiting substantial negative gaps between true and believed deaths, despite the fact their actual death rates were similar to New York. This logic motivates our instrumental variables procedure. We argue that people hear more negative COVID-related news about states that are more prominent in public conversations, and do not appropriately handicap for these inherent differences in attention when forming beliefs about pandemic deaths.[15]

Table 7 displays the results of our IV procedure, also reporting the first-stage F-statistic corresponding to each specification. Columns (1) through (3) are IV versions of our main OLS regressions from the preceding tables. The effects of erroneous beliefs remain statistically significant, and consistent with our prior of attenuation bias, the magnitudes are larger than the OLS versions. For example, in our baseline specification, a one standard deviation (70 log point) increase in erroneous believed deaths translates to a 9.8 percentage point decline in governor COVID-19 approval. Columns (4) through (6) present IV regressions for our three social distancing outcomes. These, too, remain statistically significant. Columns (7) and (8) are OLS specifications - with and without our suite of control variables - which regress the number of COVID-related news articles mentioning each state on the Google SVI measure for the state.[16] The coefficient on Google SVI is strongly significant, providing evidence that the level of attention paid to each state pre-COVID

[15] One factor that affects attention is that states with higher populations receive more search volume. One concern might be that people would under-attend to population differences if deaths were elicited in level terms, such that constructed per-capita beliefs would be mechanically biased. We avoid this by eliciting beliefs about per-capita deaths - there is no inherent reason why per-capita beliefs about high-population states would be more negative (ceteris paribus) except through the attention channel we wish to capture.

[16] Our dataset of COVID-related news articles is from AYLIEN, a news aggregator API which made this data freely-available to researchers during the COVID pandemic. The dataset focuses on the early phase of the pandemic, with coverage through July 2020.

did indeed predict the amount of coverage each state received during the pandemic.[17]

Consistent with attenuation bias concerns, the IV estimates on governor approval are larger across our various specifications. A natural question, in order to better understand what is driving our IV estimates, is how their magnitude compares to the scale of the likely attenuation bias in the OLS estimates as illustrated by Equation 4.

In most circumstances, only abstract discussions of attenuation bias are possible, because the degree of mismeasurement (σ_ξ^2) is unknown. In this setting, however, σ_ξ^2 can be characterised because these measurement errors are, in fact, OLS sampling errors, $\xi_j = \hat{\gamma}_j - \gamma_j$. Accordingly, utilising the 'first-stage' regressions from which we estimate beliefs about each state, and exploiting the fact that OLS estimate sampling variance $\widehat{Var}(\hat{\gamma}_j)$ is an unbiased estimate of the true parameter uncertainty, we can estimate σ_ξ^2 by $\hat{\sigma}_\xi^2 = \frac{1}{K}\sum_t \left(\frac{1}{N}\sum_j \widehat{Var}(\hat{\gamma}_{jt})\right)$.[18] Put simply, the sampling variation of the state FE estimates in the first-stage is the source of our measurement error, and estimated magnitude of this parameter uncertainty thus informs the degree of attenuation bias.

We further obtain the sample variance of the observed beliefs variable after residualising it against the covariates, $\hat{\sigma}^2(\tilde{\hat{\gamma}})$. Combining these two pieces (see Appendix B.2 for additional details), we obtain estimates of the theoretical attenuation in the p-lim of $\hat{\beta}_j^{OLS}$ for our primary analyses of governor approval to compare with the IV estimates in Table 7. A noteworthy tendency here is that the true beliefs γ_j can be correlated with other variables, so the residual variation in beliefs will tend to fall, and attenuation bias get worse, as more covariates are added to the regression.

Indeed, this is consistent with what both our calculations of the theoretical attenuation of the OLS estimates, and the difference in magnitude between the IV and OLS estimates that we observe. Specifically, we estimate that the probability limit of the OLS estimator is attenuated by 16.2% in the baseline specification in Column 1, 37.7% in Column 2 which controls for benchmark beliefs, and 49.4% once the full set of controls are added in Column 3 (of Table 7). These theoretical calculations are broadly consistent with the relative ratios of the OLS and IV estimates in the three

[17]We note that we find very similar results for all columns if we use counts of 2019 news articles mentioning each state instead of 2019 Google SVI as our instrumental variable.

[18]This estimation also relies on the fact that measurement errors are zero expectation.

columns. This suggests the IV estimates are consistent with the amelioration of the attenuation bias in the OLS estimates.

4.5 Effects on Political Approval (Experimental)

To provide further evidence on the effects of exogenous variation in beliefs, we turn to the results of our survey experiment. Two approaches in our survey experiment yield information about the responsiveness of governor approval to beliefs. The first approach entails simply asking respondents whether they would approve of their governor if, hypothetically, they learned that deaths per capita were X, for a variety of values of X. Specifically, we elicit conditional approval for each of three different deaths rates, one of which is the true death rate for the state and the other two of which are randomly drawn from the set of other states' death rates. To the extent that approval is varying in X, individuals are admitting that their approval of their governor's handling of the pandemic is indeed responsive to their beliefs about how well the pandemic was handled in their state.

Columns (1) through (3) of Table 8 regress an indicator for individual-level approval of governor COVID-19 handling on the natural logarithm of the hypothetical value of deaths per capita the individual is presented with. Column (1) is a simple univariate regression, whereas column (2) adds controls for individual demographic characteristics. Column (3) adds person fixed-effects, identifying solely off within-person variation. In each of these specifications, the result that approval is indeed responsive to beliefs emerges – with very strong levels of statistical significance. A hypothetical one standard deviation increase in deaths per-capita is associated with an approximately 19 percentage-point decline in approval of governor COVID-19 handling in each of these specifications. In columns (4) through (6), we provide additional evidence using hypothetical questions pertaining to a different domain – the percent decline in employment since the start of the pandemic. In this domain, too, approval is responsive to beliefs. A hypothetical one standard deviation increase in employment loss is associated with an approximately 9 percentage-point decline in approval.

The second approach to identifying effects of beliefs on governor approval is a more standard information-revelation experiment. We randomize respondents into a control group or a treatment group. In the control group, they are asked to guess their state's performance in terms of deaths per capita and then about the extent to which they approve of their governor's handling of the COVID-19 pandemic. In the treatment group, they are asked to guess their state's performance – and then told their state's true performance – before being asked about approval of their governor's COVID-19 handling. The treatment group thus receives a shock to their beliefs, allowing us to measure the effect of a shift in these beliefs on governor approval. We do precisely this in Table 9. Column (1) displays the results of the simplest version of such a specification, regressing approval on a treatment group indicator, state (log) deaths per-capita, and the interaction term thereof. The interaction term isolates how the information revealed affects governor approval for treatment group members (relative to control group individuals). This is accordingly the key variable of interest. A one standard deviation shock to believed deaths per capita is estimated to lead to approximately a 14 percentage-point decline in approval of governor COVID-19 handling. Column (2) adds controls for individual demographics, while column (3) utilizes only within-state variation.[19] It also controls for prior beliefs, namely believed deaths and benchmark deaths per-capita, both of which are interacted with treatment status. The results are broadly stable across specifications. Columns (4) through (6) repeat the exercise for employment instead of deaths, again finding an analogous effect.

5 Conclusion

In order to shed light on whether the public rewards (or penalizes) politicians for their performance in office and thereby contribute further to the literature on retrospective voting, we study public perceptions of COVID-19 death rates and governor approval ratings during the early stages of the COVID-19 pandemic. We note that, in order for the public to reward or penalize politicians for their performance, it is necessary for the public to have an accurate understanding of that performance. Errors or biases may ameliorate this ability – and thus undermine the incentive structure for politicians to continue performing well. We ran an incentivized survey on Amazon Mechanical Turk in July (Wave 1) and October (Wave 2) of 2020 asking respondents to provide their best guesses, for 10 randomly-drawn pairs of states, which state had the higher death rate

[19] Note that addition of State FE makes controlling for the level of deaths per capita in the state redundant.

(and by how much). We find that respondents choose the correct state 63.4% of the time. We find little to no evidence of partisan in-group bias, though respondents systematically overestimate death rates in Texas and Florida, states which received substantial media attention despite moderate death rates. Using data on social distancing behavior, we also show that these erroneous beliefs about state performance translate into altered social-distancing behavior.

Turning to the question of how these partially-erroneous beliefs translate into governor approval, we find that approval of governor COVID-19 handling is strongly affected by the erroneous component of beliefs. This remains true if one controls for individuals' perceptions of how well the states should have performed, setting aside factors of leadership/political competence. It is robust to the addition of a variety of state-level political and pandemic controls and to an instrumental variables specification leveraging variation in the amount of attention paid to each state. We obtain a similar result when we leverage experimental variation in another survey. We thus conclude that considerations related to imperfect information on the part of the public may generate frictions in the operation of retrospective voting models and the ability of voters to reward (penalize) good (bad) performance on the part of politicians.

References

Achen, Christopher H and Larry M Bartels, "Blind Retrospection: Electoral Responses to Drought, Flu and Shark Attacks," June 2004.

Alt, James, Ethan Bueno de Mesquita, and Shanna Rose, "Disentangling Accountability and Competence in Elections: Evidence from U.S. Term Limits," *The Journal of Politics*, January 2011, *73* (1), 171–186.

Ashworth, Scott, "Reputational Dynamics and Political Careers," *Journal of Law, Economics, & Organization*, 2005, *21* (2), 441–466. Publisher: Oxford University Press.

Barro, Robert J., "The Control of Politicians: An Economic Model," *Public Choice*, 1973, *14*, 19–42. Publisher: Springer.

Bechtel, Michael M. and Jens Hainmueller, "How Lasting Is Voter Gratitude? An Analysis of the Short- and Long-Term Electoral Returns to Beneficial Policy," *American Journal of Political Science*, 2011, *55* (4), 852–868. Publisher: Midwest Political Science Association, Wiley.

Bechtel, Michael M and Massimo Mannino, "Retrospection, Fairness, and Economic Shocks: How Do Voters Judge Policy Responses to Natural Disasters?," *Political Science Research and Methods*, 2022, *10* (2), 260–278.

Birch, Sarah, "The Electoral Benefits of Environmental Position-taking: Floods and Electoral Outcomes in England 2010–2019," *European Journal of Political Research*, 2023, *62* (1), 95–117.

Busby, Ethan C., James N. Druckman, and Alexandria Fredendall, "The Political Relevance of Irrelevant Events," *The Journal of Politics*, 2017, *79* (1), 346–350. Publisher: The University of Chicago Press.

Campello, Daniela and Cesar Zucco, "Presidential Success and the World Economy," *The Journal of Politics*, April 2016, *78* (2), 589–602.

Duch, Raymond M. and Randolph T. Stevenson, *The Economic Vote: How Political and Economic Institutions Condition Election Results*, New York: Cambridge University Press, March 2008.

Fair, Ray C., "The Effect of Economic Events on Votes for President," *The Review of Economics and Statistics*, 1978, *60* (2), 159–173. Publisher: The MIT Press.

Ferejohn, John, "Incumbent Performance and Electoral Control," *Public Choice*, 1986, *50* (1/3), 5–25. Publisher: Springer.

Fiorina, Morris P., *Retrospective Voting in American National Elections*, New Haven, CT: Yale University Press, 1981.

Gasper, John T. and Andrew Reeves, "Make It Rain? Retrospection and the Attentive Electorate in the Context of Natural Disasters," *American Journal of Political Science*, April 2011, *55* (2), 340–355.

Healy, Andrew and Gabriel S. Lenz, "Substituting the End for the Whole: Why Voters Respond Primarily to the Election-Year Economy," *American Journal of Political Science*, 2014, *58* (1), 31–47. Publisher: Midwest Political Science Association, Wiley.

— **and Neil Malhotra**, "Random Events, Economic Losses, and Retrospective Voting: Implications for Democratic Competence," *Quarterly Journal of Political Science*, August 2010, *5* (2), 193–208. Publisher: Now Publishers, Inc.

Heersink, Boris, Brenton D Peterson, and Jeffery A Jenkins, "Disasters and Elections: Estimating the Net Effect of Damage and Relief in Historical Perspective," *Political Analysis*, 2017, *25* (2), 260–268.

Huber, Gregory A., Seth J. Hill, and Gabriel S. Lenz, "Sources of Bias in Retrospective Decision Making: Experimental Evidence on Voters' Limitations in Controlling Incumbents," *American Political Science Review*, November 2012, *106* (4), 720–741. Publisher: Cambridge University Press.

Key, Vladimer Orlando, *The Responsible Electorate*, Cambridge, MA: Harvard University Press, 1966.

Kramer, Gerald H., "Short-Term Fluctuations in U.S. Voting Behavior, 1896-1964," *The American Political Science Review*, 1971, *65* (1), 131–143. Publisher: American Political Science Association, Cambridge University Press.

Malhotra, Neil and Alexander G. Kuo, "Attributing Blame: The Public's Response to Hurricane Katrina," *The Journal of Politics*, 2008, *70* (1), 120–135. Publisher: The University of Chicago Press, Southern Political Science Association.

_ **and Yotam Margalit**, "Expectation Setting and Retrospective Voting," *The Journal of Politics*, October 2014, *76* (4), 1000–1016. Publisher: The University of Chicago Press.

McAllister, Jordan H and Afiq bin Oslan, "Issue Ownership and Salience Shocks: The Electoral Impact of Australian Bushfires," *Electoral Studies*, 2021, *74*, 102389.

Persson, Torsten and Guido Tabellini, *Political Economics: Explaining Economic Policy*, Cambridge, MA: The MIT Press, 2002.

Reeves, Andrew and James G. Gimpel, "Ecologies of Unease: Geographic Context and National Economic Evaluations," *Political Behavior*, September 2012, *34* (3), 507–534.

Schwarz, Norbert and Gerald Clore, "Mood, Misattribution, and Judgments of Well-Being: Informative and Directive Functions of Affective States," *Journal of Personality and Social Psychology*, September 1983, *45* (3), 513–523.

Stokes, Leah C., "Electoral Backlash against Climate Policy: A Natural Experiment on Retrospective Voting and Local Resistance to Public Policy," *American Journal of Political Science*, October 2016, *60* (4), 958–974.

Tufte, Edward R., *Political Control of the Economy*, Princeton, NJ: Princeton University Press, 1978.

Wolfers, Justin, "Are Voters Rational? Evidence from Gubernatorial Elections," 2007.

Appendices

A Tables and Figures

Table 1 – Descriptive Statistics of mTurk Sample

	mTurk Sample	U.S. Population (18+)
Share Male	0.485	0.487
Median Age	40	47
Share White, non-Hisp.	0.726	0.628
Share w/ BA or Greater	0.491	0.306
Median HH Income	65,885	65,000
Share Clinton Voters	0.423	0.264
Share Trump Voters	0.274	0.252
Share Liberals	0.483	0.279
Share Democrats	0.426	0.354
Observations	613	-

U.S. Population data is from the 2019 American Community Survey (demographic variables), 2016 election returns data (voting variables), and 2020 American National Election Study (ideology variables).

Table 2 – Partisan In-Group Bias in Beliefs

	Erroneous Component of Believed Deaths (Log)					
	Governor Party			2016 Pres. Vote Party		
	(1) Baseline	(2) + Deaths	(3) Within State	(4) Baseline	(5) + Deaths	(6) Within State
Republican State	0.251***	0.014		0.418***	0.092***	
	(0.028)	(0.020)		(0.031)	(0.024)	
Cross Party	0.058*	0.040	0.044**	0.040	0.039	0.045*
	(0.032)	(0.025)	(0.022)	(0.036)	(0.027)	(0.024)
Home State	0.011	0.080***	-0.041*	0.017	0.082***	-0.041*
	(0.036)	(0.027)	(0.024)	(0.035)	(0.027)	(0.024)
Deaths Per Capita (Log)		-0.717***			-0.707***	
		(0.015)			(0.016)	
State x Wave FE	No	No	Yes	No	No	Yes
R-squared	0.0210	0.4434	0.5456	0.0511	0.4461	0.5475
Observations	12024	12024	12024	12240	12240	12240

Robust standard errors clustered by individual respondent. In columns (1)-(3), state party affiliation is determined by the governor's party. In columns (4)-(6), state party affiliation is determined by the state's vote in the 2016 presidential election. * p<0.10, ** p<0.05, *** p<0.01.

Table 3 – Effects of Actual and Believed COVID Deaths on Social Distancing Behavior

	Time at Home		Outside		Non-Household Contact	
	(1) Beliefs	(2) Controls	(3) Beliefs	(4) Controls	(5) Beliefs	(6) Controls
Deaths Per Capita (Log)	0.060***	0.033**	-0.070**	-0.095***	-0.153***	-0.162***
	(0.017)	(0.014)	(0.029)	(0.030)	(0.034)	(0.034)
Erroneous Believed Deaths (Log)	0.058***	0.039***	-0.133***	-0.149***	-0.191***	-0.220***
	(0.016)	(0.011)	(0.035)	(0.037)	(0.050)	(0.052)
State Controls	No	Yes	No	Yes	No	Yes
NPI Controls	No	Yes	No	Yes	No	Yes
Wave FE	Yes	Yes	Yes	Yes	Yes	Yes
Outcome Mean	-0.0142	-0.0142	0.7645	0.7645	0.6881	0.6881
R-squared	0.3943	0.6819	0.2095	0.3098	0.2570	0.3696
Observations	102	100	102	100	102	100

State controls are the state's 2016 Trump minus Clinton share and a rolling 7-day average of COVID-19 cases and deaths in the state. NPI controls are measures of both contemporaneous and average intensity of state lockdowns/NPIs. Robust standard errors clustered by state. * p<0.10, ** p<0.05, *** p<0.01.

Table 4 – Effects of Actual and Believed COVID Deaths on Governor Approval

	(1) Deaths	(2) (1) + Erroneous Beliefs	(3) (2) + Benchmark	(4) (3) + Controls	(5) (4) + NPIs
Deaths Per Capita (Log)	0.029**	-0.050	-0.152**	-0.077	-0.075
	(0.014)	(0.032)	(0.075)	(0.048)	(0.046)
Erroneous Believed Deaths (Log)		-0.113***	-0.212***	-0.131**	-0.128**
		(0.034)	(0.072)	(0.049)	(0.048)
Benchmark Deaths (Log)			0.118*	0.086*	0.082**
			(0.067)	(0.045)	(0.040)
State Controls	No	No	No	Yes	Yes
Governor Controls	No	No	No	Yes	Yes
NPI Controls	No	No	No	No	Yes
Wave FE	Yes	Yes	Yes	Yes	Yes
R-squared	0.1165	0.2141	0.2691	0.5676	0.5747
Observations	100	100	100	96	96

Outcome variable mean is 0.449. State controls are the state's 2016 Trump minus Clinton share and a rolling 7-day average of COVID-19 cases and deaths in the state. Governor controls are pre-pandemic Governor approval rating, fixed effects for party of the Governor, and 2016 Trump-Clinton net margin interacted with Governor party. NPI controls are measures of both contemporaneous and average intensity of state lockdowns/NPIs. Robust standard errors clustered by state. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 5 – Effects of Actual and Believed COVID Deaths on Governor Approval One Year Later (9/2021)

	(1) Deaths	(2) (1) + Erroneous Beliefs	(3) (2) + Benchmark	(4) (3) + Controls	(5) (4) + NPIs
Deaths Per Capita (Log)	0.035**	-0.032	-0.145**	-0.035	-0.032
	(0.015)	(0.035)	(0.069)	(0.045)	(0.042)
Erroneous Believed Deaths (Log)		-0.095**	-0.204***	-0.079*	-0.076*
		(0.040)	(0.068)	(0.045)	(0.043)
Benchmark Deaths (Log)			0.130**	0.054	0.049
			(0.061)	(0.051)	(0.044)
State Controls	No	No	No	Yes	Yes
Governor Controls	No	No	No	Yes	Yes
NPI Controls	No	No	No	No	Yes
Wave FE	Yes	Yes	Yes	Yes	Yes
R-squared	0.0949	0.1620	0.2261	0.6029	0.6113
Observations	100	100	100	96	96

Outcome variable mean is 0.449. State controls are the state's 2016 Trump minus Clinton share and a rolling 7-day average of COVID-19 cases and deaths in the state. Governor controls are pre-pandemic Governor approval rating, fixed effects for party of the Governor, and 2016 Trump-Clinton net margin interacted with Governor party. NPI controls are measures of both contemporaneous and average intensity of state lockdowns/NPIs. Robust standard errors clustered by state. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 6 – Individual-Level Specifications: Effects of Actual and Believed COVID Deaths on Governor Approval

	(1) Deaths	(2) (1) + Excess Beliefs	(3) (2) + Benchmark	(4) (3) + Controls	(5) (4) + Demographics	(6) (5) + NPIs
Deaths Per Capita (Log)	0.098***	-0.007	0.026	-0.019	-0.021	-0.022
	(0.034)	(0.032)	(0.038)	(0.032)	(0.031)	(0.032)
Erroneous Believed Deaths (Log)		-0.147***	-0.120***	-0.094***	-0.090***	-0.089***
		(0.036)	(0.038)	(0.029)	(0.028)	(0.028)
Benchmark Deaths (Log)			-0.052**	-0.058**	-0.051**	-0.051**
			(0.025)	(0.026)	(0.023)	(0.024)
State Controls	No	No	No	Yes	Yes	Yes
Governor Controls	No	No	No	Yes	Yes	Yes
Demographic Controls	No	No	No	No	Yes	Yes
NPI Controls	No	No	No	No	No	Yes
Wave FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.0216	0.0649	0.0706	0.1636	0.2876	0.2878
Observations	612	612	612	589	589	589

Outcome variable mean is 0.489. Because these are individual-level specifications, both the left-hand-side approval rating and the right-hand-side variables contain only observations pertaining to the respondents' home state. State controls are the state's 2016 Trump minus Clinton share and a rolling 7-day average of COVID-19 cases and deaths in the state. Governor controls are pre-pandemic Governor approval rating, fixed effects for party of the Governor, and 2016 Trump-Clinton net margin interacted with Governor party. Demographic controls are fixed effects for individual sex, race, household income, education, a quadratic in age, and party ID interacted with the party of an individual's state governor. NPI controls are measures of both contemporaneous and average intensity of state lockdowns/NPIs. Robust standard errors clustered by state. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 7 – Instrumental Variables: Effects of Actual and Believed COVID Deaths on Governor Approval

	IV: Governor Approval			IV: Social Distancing			OLS: Articles (Log)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Baseline	(1) + Benchmark	(2) + Controls	% Time Home	UAS: Outside	UAS: Non-HH	Baseline	Full
Deaths Per Capita (Log)	-0.068	-0.253	-0.268**	0.039***	-0.104**	-0.121**	0.312***	0.073
	(0.053)	(0.162)	(0.123)	(0.014)	(0.049)	(0.054)	(0.079)	(0.092)
Erroneous Believed Deaths (Log)	-0.140**	-0.324*	-0.345***	0.048***	-0.162**	-0.163**		
	(0.066)	(0.173)	(0.133)	(0.017)	(0.064)	(0.081)		
Benchmark Deaths (Log)		0.196	0.232**					
		(0.132)	(0.105)					
Search Volume Index (Log)							0.803***	0.647***
							(0.072)	(0.084)
State Controls	No	No	Yes	Yes	Yes	Yes	No	Yes
NPI Controls	No	No	Yes	Yes	Yes	Yes	No	Yes
Governor Controls	No	No	Yes	No	No	No	No	Yes
Wave FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First Stage F-Statistic	46.74	17.37	14.79	41.86	41.86	41.86		
Observations	100	100	96	100	100	100	100	96

Outcome variable mean is 0.449. The erroneous component of beliefs is instrumented with Google Search Volume Index (SVI) for each US state in 2019, an indicator of how much attention each state receives. Columns (1) to (3) repeat the main governor approval specifications using this instrument. Columns (4) to (6) repeat the social distancing specifications. Columns (7) and (8) show the instrument is predictive of media coverage of COVID; the outcome is log(number of COVID news articles in 2020). State controls are the state's 2016 Trump minus Clinton share and a rolling 7-day average of COVID-19 cases and deaths in the state. Governor controls are pre-pandemic Governor approval rating, fixed effects for party of the Governor, and 2016 Trump-Clinton net margin interacted with Governor party. NPI controls are measures of both contemporaneous and average intensity of state lockdowns/NPIs. Robust standard errors clustered by state. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 8 – Experimental Effects of Beliefs on Governor Approval (Hypothetical Approach)

	Deaths Hypothetical			Employment Hypothetical		
	(1)	(2)	(3)	(4)	(5)	(6)
Deaths Per Capita (Log)	-0.311***	-0.332***	-0.339***			
	(0.034)	(0.031)	(0.038)			
Employment Decline (%)				-0.030***	-0.034***	-0.038***
				(0.008)	(0.007)	(0.009)
Demographic Controls	No	Yes	No	No	Yes	No
Person FE	No	No	Yes	No	No	Yes
Outcome Mean	0.4585	0.4585	0.4585	0.5520	0.5520	0.5520
R-squared	0.1235	0.2516	0.7769	0.0327	0.1758	0.7247
Observations	615	615	615	615	615	615

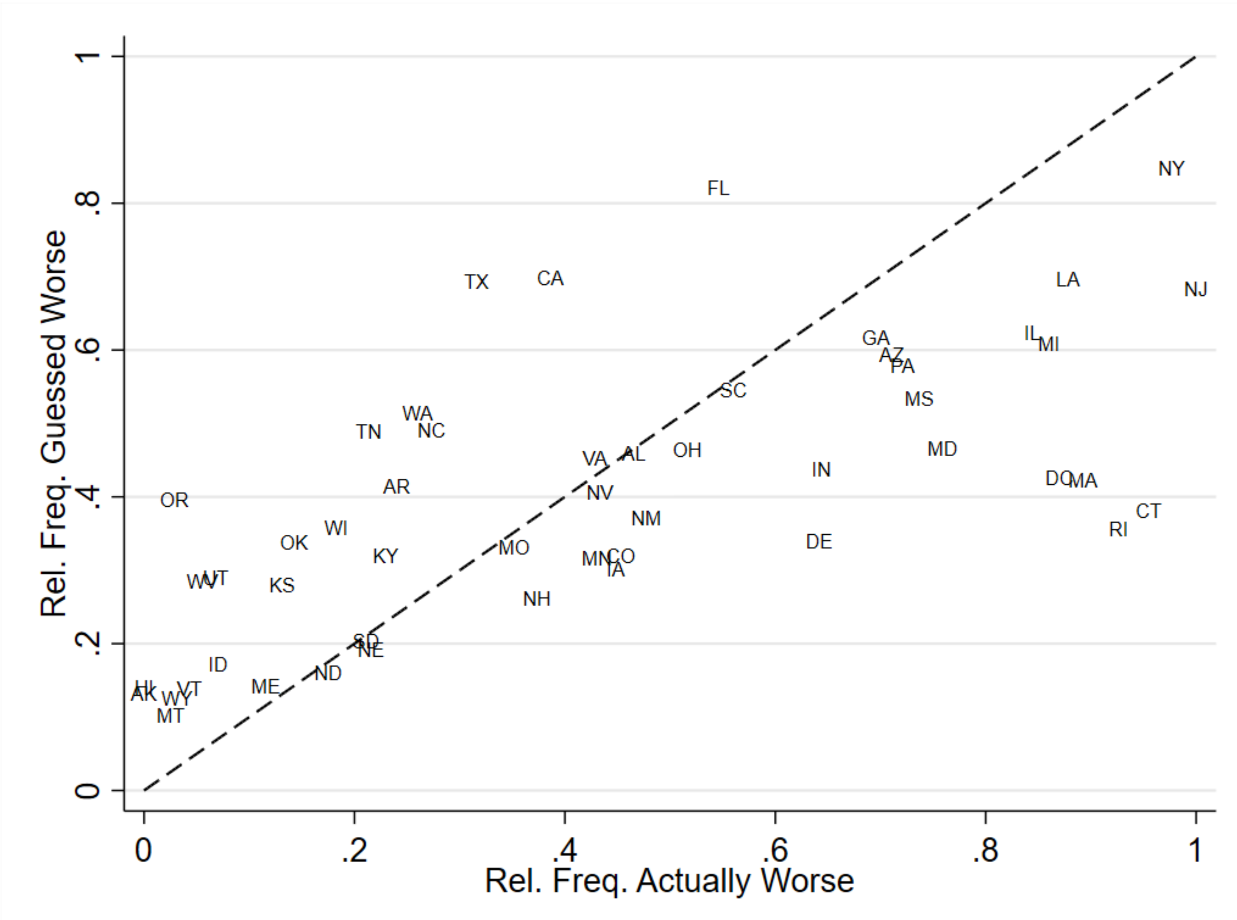
Demographic controls are dummies for individual sex, race, household income, education, a quadratic in age, and party ID interacted with the party of an individual's state governor. Robust standard errors clustered by individual respondent. * p<0.10, ** p<0.05, *** p<0.01

Table 9 – Experimental Effects of Beliefs on Governor Approval (Information Approach)

	Deaths Information			Employment Information		
	(1)	(2)	(3)	(4)	(5)	(6)
Deaths Information	-0.250**	-0.178	-0.280**			
× Deaths Per Capita (Log)	(0.117)	(0.124)	(0.121)			
Employment Information				-0.044**	-0.047**	-0.060**
× Employment Decline (%)				(0.022)	(0.021)	(0.023)
Information Treatment	Yes	Yes	Yes	Yes	Yes	Yes
State Outcome Controls	Yes	Yes	No	Yes	Yes	No
State FE	No	No	Yes	No	No	Yes
Initial Beliefs	No	No	Yes	No	No	Yes
Initial Beliefs x Treatment	No	No	Yes	No	No	Yes
Demographic Controls	No	Yes	Yes	No	Yes	Yes
R-squared	0.0191	0.1814	0.3234	0.0420	0.1735	0.3297
Observations	346	344	344	351	350	350

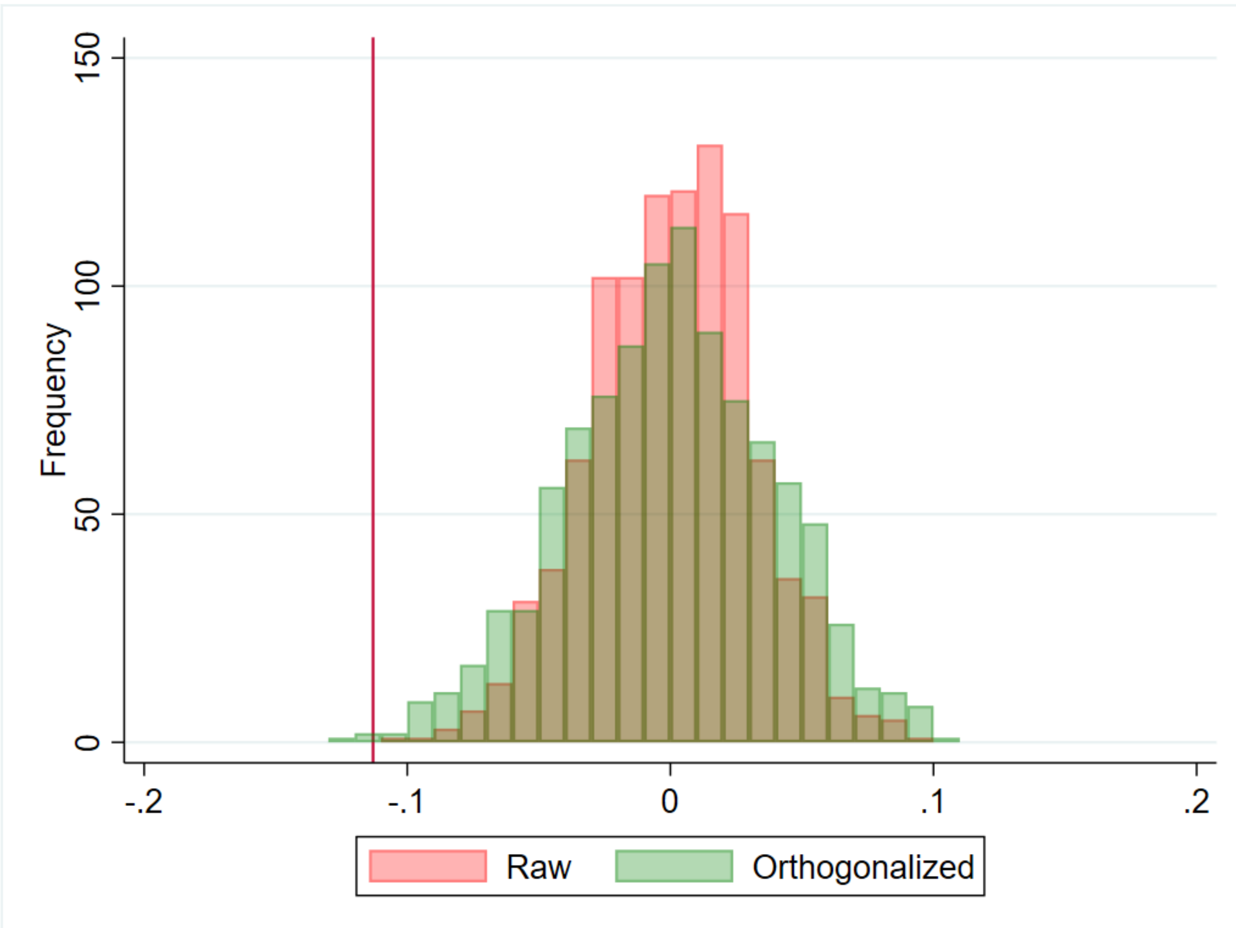
Outcome variable mean is 0.580. Information treatment controls are fixed effects for treatment group. State outcome controls are the log of deaths per capita (columns 1-3) and employment decline (%) (columns 4-6). Initial beliefs are actual and benchmark beliefs of the respondent for the log of deaths per capita (columns 1-3) and employment decline (%) (columns 4-6). Demographic controls are fixed effects for individual sex, race, household income, education, a quadratic in age, and party ID interacted with the party of an individual's state governor. Robust standard errors. * p<0.10, ** p<0.05, *** p<0.01

Figure 1 – Believed Versus Actual Relative State Performance



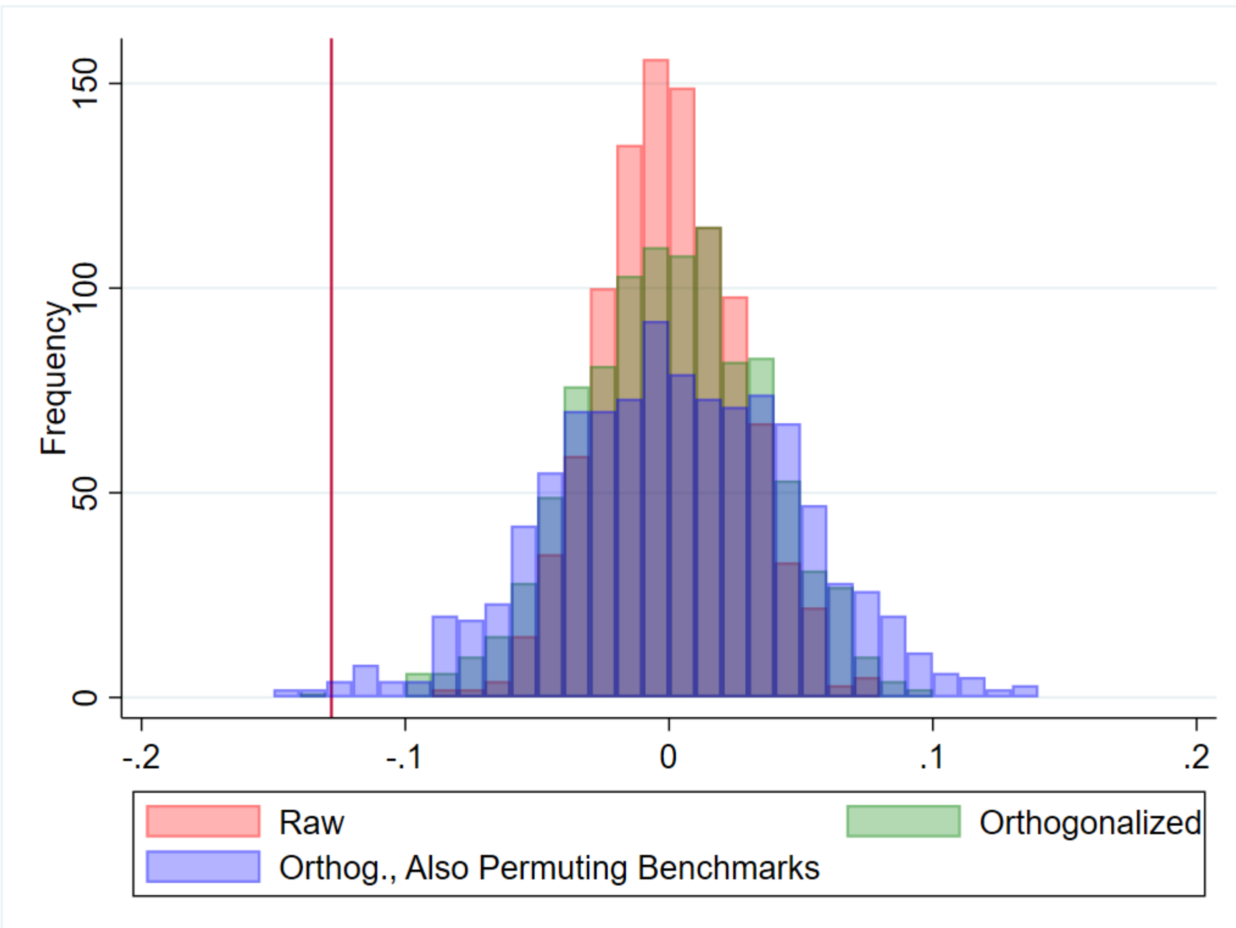
Note: This graph plots the relative frequency with which respondents guess each state has a higher death rate against the relative frequency with which each state truly has a higher death rate.

Figure 2 – Permutation Test of Baseline Specification: Distribution of Placebo Coefficients



Note: This plot displays the distribution of placebo coefficients resulting from a permutation test of the specification in Column (2) of Table 4. The red distribution is the result of regressing governor COVID-19 approval on (placebo) erroneous beliefs about deaths. The green distribution results from regressing governor COVID-19 approval on the placebo beliefs about deaths - orthogonalized against actual deaths. The vertical line represents the true estimate. The implied p-value is less than 0.01 in both cases.

Figure 3 – Permutation Test of Saturated Specification: Distribution of Placebo Coefficients



Note: This plot displays the distribution of placebo coefficients resulting from a permutation test of the specification in Column (5) of Table 4. The red distribution is the result of regressing governor COVID-19 approval on (placebo) erroneous beliefs about deaths. The green distribution results from regressing governor COVID-19 approval on the placebo beliefs about deaths - orthogonalized against actual deaths. The blue distribution also orthogonalizes the benchmark variable against actual deaths. The vertical line represents the true estimate. The implied p-value is less than 0.01 in all cases.

B Proofs and Derivations

B.1 OLS Estimator Attenuation

Let $y = X^*\beta + u$, where X is measured with error such that $X = X^* + \xi$, and ξ, u, X^* are jointly independent. Accordingly, $y = (X - \xi)\beta + u$.

Consider the properties of the OLS estimator when regressing y on X in this circumstance.

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}(X'y) \\
&= (X'X)^{-1}X'(X\beta + u - \xi\beta) \\
&= \beta + (X'X)^{-1}X'u - (X'X)^{-1}X'\xi\beta \\
&= \beta + (X'X)^{-1}X'u - (X'X)^{-1}(X^* + \xi)'\xi\beta \\
&= \beta + (X'X)^{-1}X'u - (X'X)^{-1}X^{*\prime}\xi\beta - (X'X)^{-1}\xi'\xi\beta
\end{aligned}$$

Specifically considering the case where only x_j is mismeasured, and taking probability limits,

$$\begin{aligned}
\underset{n\to\infty}{p\text{-lim}}\,\hat{\beta} &= \beta + \underset{n\to\infty}{p\text{-lim}}\left(\left(\frac{X'X}{n}\right)^{-1}\right)\underset{n\to\infty}{p\text{-lim}}\left(\frac{X'u}{n}\right) - \underset{n\to\infty}{p\text{-lim}}\left(\left(\frac{X'X}{n}\right)^{-1}\right)\underset{n\to\infty}{p\text{-lim}}\left(\frac{X^{*\prime}\xi}{n}\right)\beta \\
&\quad - \underset{n\to\infty}{p\text{-lim}}\left(\left(\frac{X'X}{n}\right)^{-1}\right)\underset{n\to\infty}{p\text{-lim}}\left(\frac{\xi'\xi}{n}\right)\beta \\
&= \beta - (Var(X))^{-1}\begin{pmatrix} 0 & \dots & & 0 \\ \vdots & \ddots & & \vdots \\ & & \sigma^2_{\xi_j} & \\ 0 & \dots & & 0 \end{pmatrix}\beta \\
&= \beta - (Var(X))^{-1}\begin{pmatrix} 0 & \dots & \beta_j\sigma^2_{\xi_j} & \dots & 0 \end{pmatrix}' \\
\Rightarrow \underset{n\to\infty}{p\text{-lim}}\,\hat{\beta}_j &= \beta_j - (Var(\tilde{x}_j))^{-1}\beta_j\sigma^2_{\xi_j} \\
&= \beta_j\left(1 - \frac{\sigma^2_{\xi_j}}{\sigma^2_{\tilde{x}_j}}\right)
\end{aligned}$$

B.2 Empirical Estimate of Theoretical OLS Attenuation

From above we have that

$$\underset{n\to\infty}{p\text{-lim}}\,\hat{\beta}_j^{\,OLS} = \beta_j \left(1 - \frac{\sigma^2_{\xi_j}}{\sigma^2_{\tilde{x}_j}}\right)$$

We wish to obtain an empirical estimate of this theoretical attenuation. Typically this is impossible because $\sigma^2_{\xi_j}$ is unknown. Here, however, the measurement errors ξ_j are OLS sampling errors of the FE obtained in the first-stage regressions, $\xi_j = \hat{\gamma}_j - \gamma_j$. While the true errors, and thus the actual empirical value of $\sigma^2_{\xi_j}$ still cannot be observed, we can exploit the properties of the OLS estimates to construct a theoretical estimate of of $\sigma^2_{\xi_j}$. Specifically, the OLS estimate sampling variance $\widehat{Var}(\hat{\gamma}_j)$ is an unbiased estimate of the true parameter uncertainty, from which we obtain unbiased estimates of ξ_j^2 for each j (and wave t). Since $\mathbb{E}(\xi_j) = 0\ \forall j$, it follows that $\sigma^2_{\xi_j} = \mathbb{E}[(\xi_j)^2]$, and therefore we can obtain an unbiased estimate of $\sigma^2_{\xi_j}$ by constructing the sample analogue of $\mathbb{E}[(\xi_j)^2]$, namely $\hat{\sigma}^2_{\xi} = \frac{1}{k}\sum_t \left(\frac{1}{n}\sum_j \widehat{Var}(\hat{\gamma}_{jt})\right)$.

An important distinction in practice, however, is that estimation of fixed effects omits one group (here, in each wave), and the remaining coefficients are estimated relative to the omitted group. Suppose the 'true' state fixed effects are γ_j for state j. Then, without loss of generality, suppose group 1 is omitted so that $\delta_j = \gamma_j - \gamma_1$ are actually obtained. Then $\widehat{Var}(\hat{\delta}_j) = \widehat{Var}(\hat{\gamma}_j - \hat{\gamma}_1)\ \forall j \neq 1$. In other words, the sample variance of the fixed effects estimates depends on the arbitrary choice of which group is treated as the baseline. Clearly, an arbitrary normalisation does not correctly characterise the true degree of measurement error.[20]

Accordingly, we proceed as follows. Since our interest is in relative beliefs of which states had relatively few or more deaths, a natural choice is to have the fixed effects normalised such that the sample mean fixed effect $\hat{\gamma}_j$ is zero in each wave (weighting state-wave pairs by how frequently they occur in the data in wave t, w_{lt} with $\sum_l w_{lt} = 1\ \forall t$). In fact, the absolute level of relative beliefs have no sensible interpretation, and thus some mean zero normalisation is the only credible option. To achieve this, we construct $\hat{\delta}_{jt} - \sum_l w_{lt}\hat{\delta}_{lt} = \hat{\gamma}_{jt} - \sum_j w_{lt}\hat{\gamma}_{lt} - (\hat{\gamma}_{1t} - \sum_j w_{lt}\hat{\gamma}_{1t}) =$

[20] For example, if the omitted group has a very small sample size pertaining to it, this will amplify $\widehat{Var}(\hat{\delta}_{jt})\ \forall j \neq 1$. More generally, if sampling errors are appropxímately orthogonal, then the difference between two state effects will have larger sampling variation than each individual state's fixed effect does.

$\hat{\gamma}_{jt} - \sum_j w_{jt}\hat{\gamma}_{jt}$, which is mean zero by design, and analogously obtain $\widehat{Var}(\hat{\gamma}_{jt} - \sum_j w_{jt}\hat{\gamma}_{jt})$ by manipulating the OLS variance-covariance matrix of $\hat{\delta}$. We take these as unbiased estimates of $(\xi_j)^2$ and input them into our expression for $\hat{\sigma}^2_{\xi}$. We calculate $\hat{\sigma}^2_{\xi} = 0.0126$ (i.e. a measurement error with a standard deviation of approximately 11 log points).

Obtaining $\sigma^2_{\tilde{x}_j}$ is simpler.[21] For each of the specifications of interest (corresponding to columns 1-3 in Table 7), we regress our beliefs variable on the other appropriate covariates, construct the residual sum of squares and divide by n. We calculate $\frac{1}{n}\sum_i \tilde{x}^2_{ji}$ of 0.0777 in the base specification, 0.0335 once controlling for benchmark beliefs, and 0.0256 once using our full set of controls. As expected, the orthogonal variation in beliefs reduces as more controls are added.

Taking these numbers, we estimate that the probability limit of the OLS estimator is attenuated by 16.2% in the baseline specification in Column 1, 37.7% in Column 2 which controls for benchmark beliefs, and 49.4% once the full set of controls are added in Column 3 (of Table 7).

C Survey Details

Amazon Mechanical Turk is an online platform on which users can opt-in to completing various tasks in exchange for monetary compensation. Participants in our survey were recruited through mTurk. We posted a brief, one-line advertisement on mTurk which summarized the survey, stating it involved making guesses about COVID death rates across states. The advertisement was visible only to mTurk users in the United States, given that our topic focuses on political approval of U.S. governors. Clicking on the posting was voluntary for all mTurk users. Would-be participants who clicked on the posting were first provided with an information sheet describing the survey and the compensation they would receive for participating, letting them know that its purpose was for a research study, describing the purpose of the research, telling them the source of funding, and reminding them that they were free to withdraw at any time if they so chose. They were provided the name and contact information of the researchers and of the institutional review board that reviewed the study. They were told that, to signal their consent, they should click the button to proceed with the survey and, if they did not consent, they could click the back button or exit

[21] From Appendix B.1 see that the actual relevant term in finite sample is $\frac{1}{n}\sum_i \tilde{x}^2_{ji}$, which converges to $\sigma^2_{\tilde{x}_j}$ in probability as $n \to \infty$. We have the empirical sum of squares, so do not need to estimate the population variance.

the page. All respondents were compensated at least $1.50. Average compensation was closer to $2.

Our initial survey was conducted in two waves. The first wave occurred between July 22nd and August 10th, 2020 and involved approximately 400 respondents. Our sample was limited to "mTurk Masters," mTurk workers specifically designated by Amazon as top performers due to consistent high-quality answers. Respondents were also required to be US residents. Generally speaking, mTurk workers skew younger than the general population, but this is somewhat less true of mTurk Masters. We compensated respondents with a base rate of $1.50, topped up with an incentive bonus of up to $0.50 for accuracy.[22]

The second wave of our initial survey was conducted on October 14th and October 15th, 2020. Since this largely conducted for validation purposes, we recruited a smaller sample of approximately 200 respondents. For this survey, instead of restricting participation to U.S. mTurk Masters, we restricted to U.S. mTurk workers who had completed at least 500 tasks with a success rate of at least 99%.[23] Compensation was again $1.50, with an incentive bonus of up to $0.50. The questionnaire for our survey can be found in full in the Online Appendix.

Finally, our information-revelation experiment was conducted on December 21st and December 22nd, 2020 with a sample size of approximately 600 respondents. We again restricted our sample to U.S. mTurk workers who had completed at least 500 tasks with a success rate of at least 99%. We compensated respondents with a base rate of $0.80, topped up with an incentive bonus of up to $0.30. The questionnaire for our survey experiment can be found in our online materials.

[22]This compensation was later increased – ultimately to a base rate of $2.50 and an incentive bonus of up to $0.75 – in order to attract additional respondents.

[23]We had exhausted the supply of U.S. mTurk Masters who were willing to take our survey at the compensation we offered.