Which Beliefs? Behavior-Predictive Beliefs are Inconsistent with Information-Based Beliefs: Evidence from COVID-19

Ori Heffetz Guy Ishai*

This version: October 26, 2021
First draft: June 3, 2021

Abstract

We investigate the relationship between (a) official *information* on COVID-19 infection and death case counts; (b) *beliefs* about such case counts, at present and in the future; (c) *beliefs* about average infection chance—in principle, directly calculable from (b); and (d) self-reported health-protective *behavior*. We elicit (b), (c), and (d) with a daily online survey in the US from March to August 2020 ($N \approx 13,900$). Beliefs about future infection *cases* are closely related to official information, but are inconsistent with beliefs about infection *chances*—risk perceptions—which are better predictors of reported behavior. We discuss potential implications for public communication of health-risk information.

Keywords: information, expectations, beliefs, risk perceptions, survey elicitation, COVID-19

JEL Classification: D83, D84, D91, I12

---

*Heffetz: S.C. Johnson Graduate School of Management, Cornell University; Bogen Family Department of Economics and Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem; and NBER (e-mail: oh33@cornell.edu). Ishai: Bogen Family Department of Economics and Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem (e-mail: guy.ishai@mail.huji.ac.il). We thank Ned Augenblick for substantially contributing to this project in its early stages, Tal Asif and Lev Maresca for excellent research assistance, Steven Woloshin for helpful comments and data sharing, and Hunt Allcott, Michele Belot, Wändi Bruine de Bruin, Bnaya Dreyfuss, Ofer Glicksohn, Aharon Haver, Charles Manski, Ted O'Donoghue, Alex Rees-Jones, Germán Reyes, Paul Windschitl and participants in Cornell's BERG, HUJI, Purdue, TAU Coller, the 36th IECA Conference, and USC's CESR COVID-19 Work In Progress Conference for helpful comments. We are grateful to the Cornell Center for Social Sciences (CCSS) for financial support.

An online appendix is available at http://nber.org/~heffetz.

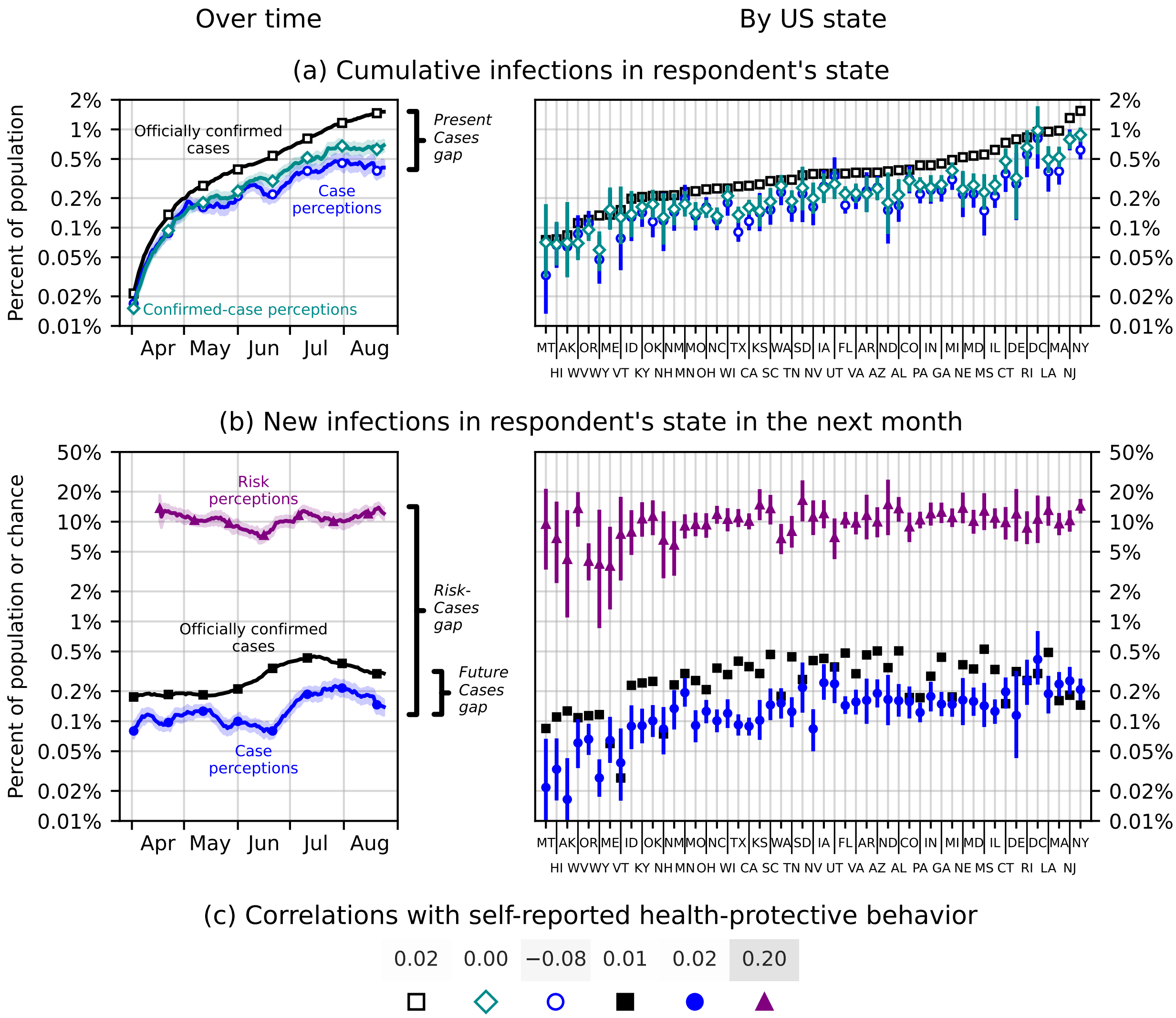
To what extent are beliefs under uncertainty based on available information? To what extent do they predict actions? These questions, always at the core of economics, become questions of life and death during pandemics. Contributing to the literature that studies beliefs by direct elicitation (recently reviewed by Manski 2018), we present new survey evidence that their answers may change dramatically depending on *which* beliefs are studied, i.e., on *how* they are elicited.

Our survey takes advantage of the unusual ubiquity and unique standardization, early in the pandemic, of communicated official information about the spread of COVID-19. For several months in 2020, confirmed local daily infection and death counts appeared saliently and frequently in the media and on official websites, and were closely monitored and discussed by the public. By comparing individuals' reported beliefs against such official information benchmarks, we investigate the extent to which beliefs are information-based; and by examining the correlations of these beliefs with reported behavior, we investigate the extent to which beliefs predict actions.

We have three main findings: (1) beliefs elicited as infection *case* counts are closely related to present and future official case-count information; however (2) beliefs elicited as *risk* perceptions—i.e., the chance to get infected—are inconsistent with those case-count beliefs, *even when mathematically, they should be identical*; notably, (3) it is the latter—the risk perceptions—that are significantly better predictors of reported behavior than the former. Together, these findings suggest that researchers and policymakers, who increasingly engage in direct elicitation and communication of numeric measures of uncertainty, may get very different outcomes, depending on *which* measures they use.

Section 1 describes our data. We use a daily online survey of US adults on Amazon MTurk, from March 24, 2020 to August 24, 2020 ($N = 13,880$), to elicit a set of COVID-related beliefs and reported behaviors. We merge the survey data with daily official state-level infection and death case counts. Figure 1 (next page) summarizes our main results, which we investigate in detail in Section 2. All quantities are reported as percentages on log scales.

Figure 1: Information, Beliefs and Behavior (March–August 2020)



Legend: (quantities defined at the survey-response level, referring to respondent's state population)

Official data:

□ Cumulative percent of population confirmed as infected as of today (the observation day).

■ Future percent of population newly confirmed as infected during the next month (30 days).

Survey data:

◇ Perceived cumulative percent of population confirmed as infected as of today.

○ Perceived cumulative percent of population (actually) infected as of today.

● Predicted percent of population to get infected during the next month.

▲ Predicted average chance of a person (from the state) to get infected during the next month.

Notes: Panels (a), (b): All quantities are represented on log-scale vertical axes as percent of state population or percent chance. Each observation is transformed using $\log\left(\left[1+\frac{x}{100}\cdot(\text{state pop.})\right] / [1+\text{state pop.}]\cdot 100\right)$, the transformed observations are averaged, and averages are exponentiated (Appendix D.1 shows that all results are robust to this log transformation). Over time: 10-day moving (weighted) averages. Light-colored areas and error bars: bootstrapped 95% confidence intervals. Panel (c): Pearson correlations between the log-transformed percentages (by color/shape) and the number (0–9) of reported health-protective private behaviors (e.g., washing hands more often).

We first investigate the relationship between beliefs about the number of COVID-19 cases and official case-count information. We elicit beliefs about current and future numbers of infections in each respondent's US-state of residence. Panel (a) compares beliefs about current cumulative cases (blue hollow circles) to the officially reported numbers (black hollow squares). Panel (b) compares predictions about *new* cases in the next month (blue solid circles) to official new-case numbers (black solid squares); these two new-case measures are constructed by subtracting, respectively, believed and official present cumulative cases from predicted and future-reported cumulative cases.[1]

In a perfect-measurement, perfect-information benchmark, circles and squares should coincide in panel (a) and furthermore, at the individual-observation level, be perfectly correlated. Assuming rational expectations, circles and squares should also coincide *on average* in panel (b), with no predictions regarding correlations.

Finding 1. *Case perceptions—perceptions and predictions of numbers of infection cases—generally follow official numbers, and somewhat understate them.*

While case perceptions in panels (a) and (b) understate (and possibly lag behind) the official numbers by 49 and 52 percent on average—henceforth, *Present Cases gap* and *Future Cases gap*, respectively—they remain in the same order of magnitude, and they generally follow the official-information time and state trends. Furthermore, case perceptions in panel (a) are moderately correlated with official numbers: $r = 0.41$.[2]

We show this finding's robustness across respondent subpopulations and (randomized) question order in Section 2.1. We also show there that elicited perceptions of current *officially confirmed* cases "reported by the authorities" (cyan diamonds in panel (a)) are close to the above perceptions of current *actual* cases (circles)—suggesting belief in neither under- nor

[1] We use the terms "beliefs" and "perceptions" interchangeably, and sometimes use the term "predictions" for beliefs that concern future outcomes.

[2] Figures for elicited and official *death* numbers, rather than *infection* numbers, are reported in Appendix C.1. They show an even tighter relation between official information and perceptions: the Present Cases and Future Cases gaps are, respectively, 32 and 14 percent understatements; the correlation between perceived current deaths and information is 0.54.

over-detection/reporting of COVID cases. Perceived cases (circles) are on average 17 percent lower than perceived confirmed cases (diamonds), a robustly small gap with unstable sign; the two are strongly correlated: $r = 0.68$.

Second, we investigate the relation between *case* perceptions and *risk* perceptions. We elicit a set of perceptions of medical and economic risk. To cleanly compare these risk perceptions—elicited as probabilities—and the above case perceptions—elicited as case counts—the set also includes risk perceptions that are in principle *mathematically equivalent* to the case perceptions represented by the solid circles in panel (b): the perceived state-average infection risk in the next 30 days (purple solid triangles). In an internally-consistent-beliefs benchmark, solid circles and triangles should be identical and, at the individual level, be perfectly correlated.[3]

Finding 2. *Risk perceptions are inconsistent with their in-principle-mathematically-equivalent case perceptions, and overstate them by orders of magnitude on average.*

Risk perceptions (triangles) are 79 times higher on average than case perceptions (circles)—henceforth, *Risk-Cases gap*. The correlation between triangles and circles, $r = 0.13$, is also surprisingly far below the perfect-correlation benchmark.

Notably, we find this Risk-Cases gap in spite of a survey design that arguably facilitates consistent reporting of case and risk perceptions. For example, the survey interface accepts case perceptions as either an absolute number (of people) or a percent (of the population), simultaneously translating between the two formats and displaying both as the respondent types in one or the other; risk perceptions are entered as a percent (chance) (for screenshots, see Figure 2 on page 12). This gap remains large both over time and across states; in Section 2.2 we find that demographic characteristics and question order can explain only a

[3]The above statement is true under the assumption that when asked about perceived/predicted cases (solid circles), respondents report *averages*, rather than other measures of central tendency. (If person i in a state with population N gets infected with probability p_i, then the state-average infection probability (triangles) is $\frac{1}{N}\sum_{i=1}^{N} p_i$. This expression equals the *average* realized fraction of infected people in the population.)

small fraction of it.[4]

Finally, to investigate the relation between beliefs and behavior, we elicit self-reports of risk-mitigating actions. Beginning in June 2020, we asked respondents about adopted health-protective behaviors, such as washing hands more frequently or avoiding crowds (for full details on the evolution of our survey design, see Section 1.1).

Finding 3. *Self-reported health-protective behavior is moderately positively correlated with risk perceptions and uncorrelated with case perceptions.*

Of all the perceptions elicited in our survey, the number of behaviors reported as adopted is best predicted by—i.e., it is most strongly correlated with—the above risk perceptions (triangles; $r = 0.20$), while having close to zero correlation with the (in principle mathematically identical) case perceptions (solid circles). Panel (c) reports these correlations (see Appendix C.8 for full correlation tables).

This finding also generally holds within respondent subpopulations and is little affected by controlling for demographic variables (including state and day fixed effects) and other specification variations. In terms of magnitudes, the bottom risk-perceptions decile report average perceived infection risk of 0.2 percent—which is rather close to official benchmarks—while engaging with 4.0 protective behaviors on average. In comparison, the top decile report perceived risk of 66 percent—which is wildly unrealistic—while engaging with 5.8 protective behaviors. This large difference in behavior, of 0.8 standard deviations, could make a big difference in a pandemic.

Our survey design and sample size allow us to conduct many other robustness checks, investigate subpopulations, and report a rich set of additional findings. We briefly summarize them throughout Section 2, with details relegated to the appendix.

Why do we find risk perceptions essentially unrelated, in both levels and correlations, to

[4] Together, our first and second main findings also imply that elicited risk perceptions (triangles in Figure 1) overstate the official information benchmark (solid squares) by orders of magnitude. In isolation, this gap could in principle be rationalized as reflecting a belief in massive under-detection/reporting of actual cases, but that would be inconsistent with the finding that actual and confirmed cases (hollow circles and diamonds) are perceived to be relatively close.

case perceptions (and to official benchmarks)? While theoretically equivalent, case and risk perceptions are elicited using different survey questions. Their differences must therefore be related to differences in elicitation details, which we explore in Section 3. First, exploiting our dual-format interface, we find that differences in response format (counts versus percentages) are likely important. Second, using an additional survey (conducted in early 2021, $N = 1,530$), we find that a case-perceptions question that we modify to be more similar to a risk-perceptions question in wording, response format, and structure, elicits perceptions that look strikingly similar to risk perceptions: unrelated to official benchmarks but correlated with behavior. We cautiously conclude that in our context, elicited beliefs may depend more on a question's wording, response format and structure—the question's "look and feel"—than on the underlying mathematical concept the question asks about.

We discuss related literature, implications of our findings, and open questions in Section 4. We first show that our first and second main findings appear consistent with a psychological literature investigating deviations between perceptions about probabilities and relative frequencies (e.g., Gigerenzer and Hoffrage 1995). We then relate our third main finding to a literature investigating the relation between health-risk beliefs and protective behavior (e.g., Brewer et al. 2007). To the best of our knowledge, we are the first to compare the association with behavior of beliefs elicited as probabilities versus as relative frequencies—two elicitation forms routinely used in economic studies and thus of special interest to economists. Moreover, we provide an "all-included" investigation, within a single study, of the very different associations these beliefs have both with information benchmarks—something that a vast psychological literature has documented—and with reported behavior—a novel contribution.

Finally, we discuss a potential implication for public communication of risk: depending on policy goals, policymakers may want to reconsider the case-count language that was so prominently used early in the pandemic. To demonstrate our point, we focus on one much-discussed application: partisan differences in beliefs and in risk-mitigating behaviors in the US during COVID-19. A growing body of recent work (e.g., Allcott et al. 2020, Barrios and Hochberg 2020, Bruine de Bruin et al. 2020, and Fan et al. 2020) finds that relative to Republicans, Democrats consume different news, perceive COVID as riskier, and engage in more social distancing—suggesting that information interventions may be effective in reducing such differences. We replicate past findings, but also find that *within* each political group, behavior is still much more strongly correlated with risk perceptions—which appear out of touch with reality—than with either case perceptions or official case information. Hence, to the extent that our correlations imply causation, policies improving *case-count* communication may have limited behavioral effects, while policies directly targeting risk perceptions, perhaps through directly communicating infection *chances* or *population percentages*, may be more effective.[5]

We conclude in Section 5, discussing broader implications as well as future directions. The disconnect we find between differently elicited beliefs calls into question researchers' ability to easily and reliably elicit beliefs using standard survey questions. However, viewing our findings as mainly demonstrating measurement issues in belief elicitation misses the bigger picture. Returning to the motivating questions we opened with, we ultimately find in our data a weak relation not only between differently elicited beliefs but, importantly, between observable objects: people's *information*—their "input"—and *behavior* (though self-reported)—their "output." Our study, which compares beliefs with both information and reported behavior in a single, real-world, high-stakes setting sheds light on *this* disconnect as well. Echoing vast literatures in psychology (reviewed in Section 4), our results may call into question the idea, still the standard among economists, that beliefs—the connecting link between information and behavior—should be modeled as a single object.

[5]However, our finding that those with highest risk perceptions, who appear to grossly overstate actual risk, engage in more protective behaviors, may create a dilemma for policymakers, because it may imply that such public panic can also have desired behavioral implications. Correcting risk perceptions may thus overall reduce protective behavior.

1 Data

We use an online survey to elicit perceptions and reported protective behaviors, and a public data source to retrieve counts of COVID-19 confirmed cases and deaths.

1.1 Online Survey

Survey Design. Table 1 shows shortened versions of the survey questions and summarizes details regarding question order and response format. (See Appendix A.1 for all screenshots and Appendix B for all response distributions.) The survey consists of six modules, A–F. They are preceded by an entry question that elicits current US state of residence (or DC), and are followed by a final screen that includes demographic and exit questions. The six modules' internal order is, with equal probabilities, *F*–A–(B↔C)–D–E or A–(B↔C)–*F*–D–E or A–(B↔C)–D–E–*F*, meaning that (i) the reported-behavior module F can be first, in the middle, or last; (ii) B–C and C–B are equally likely; and (iii) the other modules' order is fixed.[6] Section 2 and Appendix D.2 investigate order effects and find that no single order substantially changes our main results. In the rest of the paper we therefore pool the data across all survey orders.

As the table conveys, the survey structure generally encourages respondents to think about the elicited constructs as related to one another. As a consequence, our results should be viewed as an upper bound on the strength of the relations between information, different types of perceptions, and behavior.[7]

Finally, to aid readers of our paper in judging how compelling our evidence is, we provide information on the evolution of our survey design. We started collecting data early in the

[6] A small subsample was given an order A–(B↔C)–*E*–D (before F was added to the survey), to test the effect of the distance between modules B and E on the Risk-Cases gap (for results, see Section 2.2; for more details, see Appendix A.2; for randomization balance tests, see Appendix A.4).

[7] We do not incentivize respondents to report accurate beliefs. Rather, we ask them to "answer truthfully" and, when eliciting perceptions about publicly available information (in module A), to "answer without looking up the information." Respondents therefore have no incentive to "cheat" by looking up these numbers, and we find no evidence that they do (for example, only 13 percent of module-A responses are within 5 percent of the official counts). Importantly, such "cheating" would have only affected the Present-Cases-gap part of our first main finding, but neither its Future-Cases-gap part nor our second and third main findings.

Table 1: Survey Design

Module		Question		Timing	Format	Order	Comments
Confirmed-case perceptions	A	How many people in *[state]* have been **reported** by the authorities as [**infected** [A1] / **dead** [A2]] due to the coronavirus *[timing]*?		as of today	# or %	A1–A2 (fixed order)	Page 1.
Case perceptions	B + C	How many people in *[state]* will have [been **infected** [B]] / [**died** [C]] due to the coronavirus *[timing]*? (number may differ from the one reported by the authorities)	(1)	as of today	# or %	2×2: B–C or C–B, (1)–(2)–(3) or (3)–(2)–(1)	B (C) on page 2 and C (B) on page 3.
			(2)	a week from now	# or %		
			(3)	a month from now	# or %		
Risk perceptions + Predicted well-being	D1	What is the chance that **you or your immediate family** will suffer **bad medical outcomes** due to the coronavirus *[timing]*?		in the next month	%	D1–D2–D3 or D3–D1–D2	Page 4.
	D2	What is the chance that **you or your immediate family** will **lose your jobs or run out of money** due to the coronavirus *[timing]*?		in the next month	%		
	D3	What is the anticipated **well-being** of **you and your immediate family** *[timing]*?		in the next month	0–100		
	E1	What is the chance that **you** will get **infected** *[timing]*?		in the next month	%	E1–E2 or E2–E1	Page 5 (pages 4–5 in a complementary version). E1 and E2 asked since days 15, 25 respectively.
	E2	What is the **average chance** of a person in *[state]* to get **infected** *[timing]*?		in the next month	%		
Self-reported health-protective behavior	F	Which of the following have you done *[timing]* to **keep yourself safe from coronavirus**? - 9 private-domain questions about hand-washing frequency, cleaning habits and cautious touching / breathing habits. - 3 public-domain questions about going out and meeting others.	(1) – (12)	in the last week	Yes / No	F1–F12 (fixed order)	Page 6 (before page 1 or between 3 and 4 in complementary versions). Asked since day 88.

Notes: Design details of main survey modules. *[state]*: US state of residence (self-reported in survey intro). Modules A–C are answered using a dual-format (# or %) interface; see Figure 2. Question order is sometimes randomized within modules; see Order column. For full survey text and screenshots, including behavior and demographic questions, see Appendix A.1. For details about complementary survey versions, see Appendix A.2.

pandemic, using modules A–D, to investigate general relations between elicited beliefs and the newly ubiquitous, uniquely standardized, official COVID-19 case-count information. We immediately noticed a striking gap between respondents' general health-risk estimates in question D1 and predicted infection rates in module B. To investigate it, starting on the 15th day of data collection, we added module E (which always appears to respondents only *after* the original modules A–D). At first it contained only a personal-infection-risk question, E1. As the gap remained, ten days later we added a state-average-infection-risk question, E2, to rule out several potential explanations (e.g., private health-risk information). Finally, two months later we added module F, with self-reported-behavior questions, to investigate the relation between actions and the beliefs in modules A–E. The table's Comments column reports the timing of added modules and questions.

Response Format. Figure 2 reproduces the survey's perceptions-elicitation screens. Panel (a) shows that case perceptions (A–C) are elicited using a dual-format interface. Respondents are asked about absolute case numbers ("How many people..."), but can choose to enter either an absolute number or a percent of the state population. As they type in, their response is simultaneously translated into the other format and saliently displayed in both formats. Panel (b) shows that risk perceptions (D–E) are elicited using simple textboxes for entering percent chance. This combination of interfaces encourages respondents to recognize (i) the equivalence between number and percent in the case-perceptions questions and (ii) the link between case perceptions and risk perceptions, as both are displayed (and possibly also entered) in percent.[8]

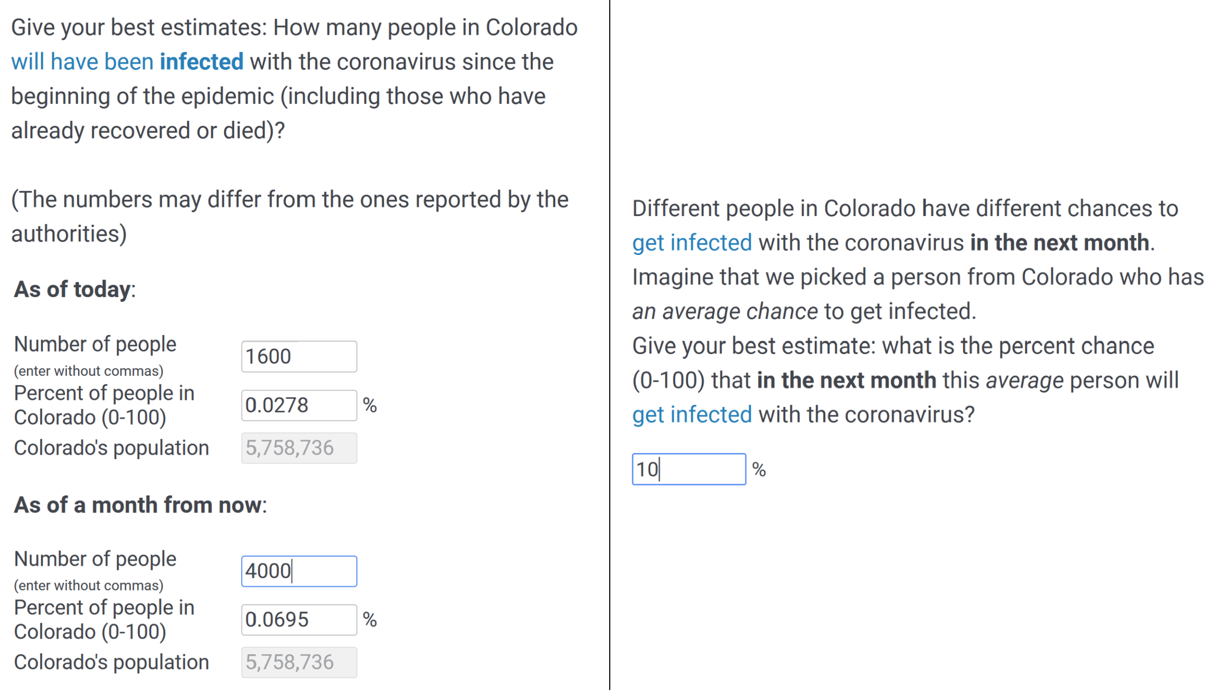
Respondent Population. Data were collected daily for 5 months, from March 24 to August 24, 2020 (154 days in total). The daily mean number of responses is 92 (SD = 26),

[8]For completeness, we mention that two other elicitation formats are used in the survey: anticipated well-being (D3) is elicited with a simple textbox for entering a number from 0 to 100; protective behaviors (F) are elicited using a set of Yes/No checkboxes.

Figure 2: Screenshots of Perceptions Elicitation Interfaces

(a) Case-perceptions questions

(b) Risk-perceptions question



Notes: Panel (a): perceived cumulative number/percent of state-level COVID-19 infection cases as of today (question B1) and a month from now (B3); the difference (B3 − B1), new infections in the next month, is used to construct the case perceptions (solid circles) in Figure 1. A similar dual-format elicitation interface—simultaneously translating numbers to percentages and vice versa—is used throughout modules A–C of the survey. Panel (b): perceived state-average infection risk in the next month (E2); risk perceptions (triangles) in Figure 1. A similar elicitation format, using percent chance, is used throughout modules D–E.

and median survey duration is 5 minutes.[9] We recruited respondents on Amazon MTurk by posting on the platform, each day typically around noon ET, a task paying $0.70. We set a minimum MTurk-experience criterion and screened out those from outside the US, following the protocol suggested by Kennedy et al. (2018), to minimize low-quality responses. For more details, see Appendix A.3.

From March to June 1, 2020, 6,327 (unique) respondents completed the survey. As we observed sign-up slowing down, starting on June 2 we allowed past respondents (as of June

[9] March 30 and June 12 had very few responses due to a human error in publishing the survey. Excluding those dates, the minimum number of observations per day is 48 and the maximum is 241. For a histogram and distribution of daily responses, see Appendix A.4.

2) to participate one more time; we collected 7,840 additional responses. In Section 2 we use the resulting partial panel to further show the robustness of our findings.[10]

The sample is not US-representative; it is younger, more educated, and more liberal-leaning. However, it is fairly broad, heterogeneous, and representative of all US states. For descriptive statistics, see Appendix B.

Raw, Full and Main Samples. Our raw sample includes 14,167 full survey responses. We exclude observations that (1) managed to bypass the single-participation restriction and respond more than once before or after the restriction reset (210 observations, 1.5 percent), (2) report percent values below 0 or above 100 in at least one question (65 obs., 0.5 perc.), or (3) report an age below 18 (14 obs., 0.1 perc.). This generates our *full sample* of 13,880 observations. Additionally, since our main analysis focuses on respondents' predictions of *new* infections in the next month—constructed as the difference between their predicted *cumulative* infections a month from now (question B3) and today (B1)—we exclude 724 responses (5.2 percent of the full sample) with a negative difference. To verify that this exclusion does not drive our results, we also conduct several versions of our main analysis on the full sample and, as we report in Appendix D.3, none of the results we examine is meaningfully impacted. The resulting *main sample* consists of 13,156 responses by 10,538 unique respondents, of whom 2,618 (25 percent) responded twice (once before and once after June 2, 2020).

[10]Difficulty to reach a respondent is an often unobserved characteristic that may affect survey results (Heffetz and Rabin 2013). In our context, respondents who participate twice (a quarter of all participants) may be easier to reach than those who respond only once; they may also be more experienced in (or bored by) answering our survey in their second time. However, we find that our main results are similar across one-time and two-time participants both prior to and after June 2 (for details, see Appendix C.10).

1.2 Official Case-Count Data

We use publicly available data from *The New York Times*.[11] Each state$\times$date record includes the cumulative numbers of officially confirmed COVID-19 infections and deaths announced by that day midnight ET. We match each survey response with the infection and death records from the relevant state and date, ET.[12] Past published numbers are sometimes updated later on, but we only use the originally published numbers, since they are the ones that were available when our respondents answered the survey.[13]

2 Results

Our main results, reported in Figure 1, were summarized in the introduction. This section analyzes them in more detail, and summarizes additional results and robustness checks fully reported in the appendix.

2.1 Relation Between Information and Perceptions

Infection and death cases, month- and week-forward predictions. Our first main finding is that respondents' case perceptions are closely related to official numbers. Figure 1 shows that perceptions of confirmed infections at present (hollow circles), and predictions of new infections in the next 30 days (solid circles), are generally close to official figures, understating them by 49 percent (Present Cases gap) and 52 percent (Future Cases gap) on average, respectively. These results are not unique to the perceptions plotted in Figure 1. Appendix C.1 shows that predictions of new infections in the next 7 days are understated by 42 percent; and that perceptions of cumulative confirmed deaths and predictions of new

[11] https://github.com/nytimes/covid-19-data. (In earlier versions of our analysis we got essentially the same results using data provided by The COVID Tracking Project at The Atlantic, at https://covidtracking.com/, but the Project stopped reporting data in March 2021.)

[12] Appendix D.4 shows that whether survey responses are matched with previous-, same-, or next-day official reports makes little difference.

[13] The public dataset was first published on March 28, 2020, hence data for our first four survey dates, March 24–27, may include some ex-post updates applied by March 28.

deaths in the next 7 and 30 days are less understated, by 32, 6, and 14 percent respectively.

Demographics and general platform experience. We explore possible drivers of the Present and Future Cases gaps (see Appendix C.2). We find that they are generally stable not only across state and time, but also across demographic groups. We also find little effect of respondents' previous experience on the MTurk platform, using an additional, time-limited sample ($N = 255$) of Workers with less experience than the baseline sample (described in Appendix A.2).

Module and question order. We find that different randomized orders of the survey modules and questions (see Table 1 and surrounding text) have some meaningful quantitative, but no qualitative effect on the Present and Future Cases .gaps. Both gaps are consistently negative and, importantly, orders-of-magnitude smaller than the Risk-Cases gap (Appendix D.2).

Perceived cases vs. perceived confirmed cases. Perceived (actual) cases and perceived confirmed cases (hollow circles and diamonds in Figure 1) are close not only on average, but also at the individual level: they are identical in 37 percent of responses, and the distribution of differences when they are not is concentrated around zero (for example, in 76 percent of responses, one is at most twice the other). On average, perceived actual cases are 17 percent lower than perceived confirmed cases, suggesting an average belief in over-detection or over-reporting of cases (or both)—a belief that, intuitively, appears to have the wrong sign. But this small average difference, and its sign, are sensitive to randomly assigned order of the case-perceptions questions. In particular, the half of respondents who are asked to predict cases in a time-horizon order of [today]–[in a week]–[in a month] (see Table 1, modules B and C) perceive the above difference to be −9 percent on average, consistent with an average belief in little *under*-detection/reporting; the other half, who are asked in reverse time-horizon order ([in a month]–[in a week]–[today]), perceive the difference to be 37 percent on average. For more details and analysis of all order effects, see Appendices C.3 and D.2.

Importantly for our first and second main findings, perceptions of cases—whether confirmed or actual—are in the same order of magnitude as official case-count information, while risk perceptions are orders-of-magnitude larger than such perceptions.

2.2 Relation Between Case Perceptions and Risk Perceptions

Outliers. Our second main finding is a large Risk-Cases gap, between perceived average infection risk (risk perceptions, solid triangles in Figure 1) and its in-principle mathematical equivalent, percent of the population predicted to be newly infected (case perceptions, solid circles). The large gap is not driven by outliers. Its distribution is symmetric and bell-shaped, and 96 percent of the sample have a positive gap (Appendix C.4).

Demographics and general platform experience. As in Section 2.1 above, we explore (in Appendix C.2) possible drivers of the Risk-Cases gap and find that it varies little across state, time and demographic groups. Among less-experienced MTurk Workers, the gap is in fact 2.1 times larger, on average, than in our main sample (with a 95% confidence interval between 1.4 and 2.8).

Module and question order. We find (in Appendix D.2) that survey order has some effect on the Risk-Cases gap, but it is rather limited. Even when the question about infection cases in 30 days appears immediately before the question about perceived state-average infection risk in the next 30 days, the Risk-Cases gap only shrinks to 0.7 of the average gap (with a 95% confidence interval between 0.3 and 1.0).

2.3 Relations Between Perceptions and Self-Reported Behavior

Controlled regressions. Our third main finding is that behavior is much more strongly correlated with risk perceptions ($r = 0.20$ in Figure 1) than with case perceptions ($r = 0.02$).

Table 2 reports OLS regressions of self-reported health-protective behavior on: perceived state-average infection risk ("Risk perceptions" row); percent of the population predicted to be newly infected in the next 30 days ("Case perceptions"); their interactions with the survey position of the behavior module F, which appeared in the beginning ("Behavior first"), middle ("Behavior middle"), or end of the survey (omitted category); demographic controls;[14] and state and day fixed effects. The baseline dependent variable is defined as the sum of behaviors reported as adopted out of a list of nine private behaviors, i.e., excluding three public behaviors, which may be affected by state regulations.[15] Controlling for state and day fixed effects, an alternative dependent variable in column (6) includes all twelve behaviors, private and public.[16]

The regression results reinforce the main message of Figure 1's panel (c). Under all specifications, the perceived state-average infection risk is dramatically more strongly associated with behavior than the percent of population predicted to be newly infected, both economically (coefficient range = 0.19–0.25 vs. −0.00–0.04) and statistically, although the two elicit, in theory, a mathematically identical concept. A coefficient of, e.g., 0.22 on log perceived state-average infection risk (columns 1 and 3) means that an e-fold increase in perceived risk, or 0.48 standard deviations, is associated with an increase of 0.22 in the number of behaviors adopted, or 0.10 standard deviations. Adding order effects of the behavior module F, demographics and state and day fixed effects, and the three public behaviors (columns 4–6) makes

[14] These include: number of people in household; number of people above 18; gender (male/female/other); Hispanic origin (yes/no); race (White/Black/Asian/Native/other); year born; education (10 categories); marital status (6); employment status (6); economic attitudes (7-point scale from very liberal to very conservative); social attitudes (same 7-point scale); political self-identification (Republican/Democrat/Independent/other/none); combined household income (8 brackets for \$0–\$200,000; or above \$200,000); medical insurance coverage (bad/fair/good); have been infected with COVID-19 (yes/no/prefer not to answer); someone from immediate family has been infected (same options).

[15] Private behaviors: increasing hand-washing frequency relative to pre-pandemic habits by at least 5/10/15 times per day (three separate questions); cleaning or sanitizing incoming mail and deliveries; cleaning or sanitizing groceries; cleaning or sanitizing furniture or frequently touched items; avoiding touching own face; stopping breath when passing near others; coughing into elbow rather than palm. Public behaviors: avoiding contact with people from a high-risk group; avoiding meeting family and friends; avoiding public spaces, gatherings and crowds.

[16] All regressions report Driscoll and Kraay (1998) standard errors, which assumes a heteroskedastic error structure, possible cross-individual correlations, autocorrelation up to some time lag—chosen to be 4 days according to a formula provided by Hoechle (2007)—and using the standard Bartlett kernel.

Table 2: Perceptions and Behavior

Dependent variable: Self-reported protective behavior

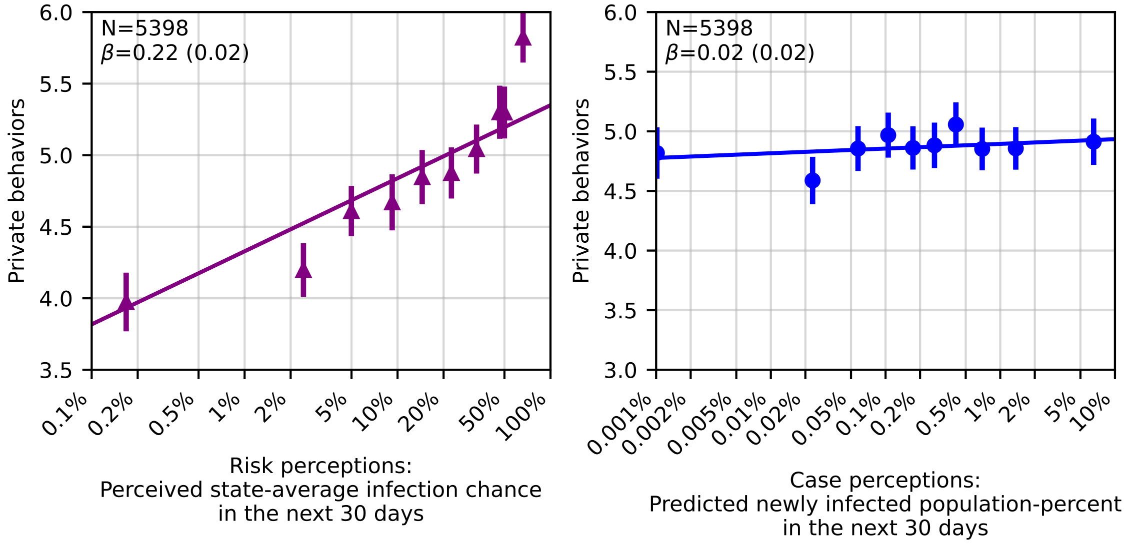
	Only 9 private behaviors					All 12 behaviors
	(1)	(2)	(3)	(4)	(5)	(6)
Risk perceptions	0.22		0.22	0.25	0.19	0.24
	(0.02)		(0.02)	(0.02)	(0.02)	(0.03)
Case perceptions		0.02	−0.00	0.01	0.02	0.04
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Behavior first				−0.19	−0.17	−0.27
				(0.06)	(0.06)	(0.08)
Behavior first × Risk perceptions				−0.07	−0.06	−0.07
				(0.02)	(0.03)	(0.03)
Behavior first × Case perceptions				−0.02	−0.03	−0.01
				(0.04)	(0.03)	(0.04)
Behavior middle				−0.17	−0.18	−0.22
				(0.05)	(0.05)	(0.07)
Behavior middle × Risk perceptions				−0.01	−0.01	−0.02
				(0.04)	(0.04)	(0.05)
Behavior middle × Case perceptions				−0.02	−0.02	−0.03
				(0.03)	(0.03)	(0.03)
Constant	4.33	4.89	4.32	4.39		
	(0.06)	(0.05)	(0.07)	(0.08)		
Demographics	No	No	No	No	Yes	Yes
State fixed effects	No	No	No	No	Yes	Yes
Day fixed effects	No	No	No	No	Yes	Yes
N obs.	5398	5398	5398	5398	5398	5397
R^2	0.04	0.00	0.04	0.04	0.14	0.15

Notes: OLS regressions. Dependent variable: number of self-reported health-protective behaviors, out of nine private behaviors (columns 1–5) or twelve private and public behaviors (column 6). Independent variables: Risk perceptions: (log) perceived state-average infection risk in the next 30 days; Case perceptions: (log) percent of population predicted to be newly infected in the next 30 days; Behavior first/middle: survey position of the behavior module (F).
The non-binary interacted variables (Risk perceptions and Case perceptions) are centered around their means (to estimate the uninteracted order effects at the mean perception values). In parentheses: Driscoll-Kraay standard errors using Bartlett's kernel and a bandwidth of 4 days.

little difference. Comparing $R^2 = 0.04$ in column (1) with $R^2 = 0.14$ in column (5)—where all demographics, fixed effects, and order indicators are included—provides another indication that the perceived state-average infection risk has a *relatively* strong predictive power. *Absolutely*, however, these R^2 values show that the bulk of variation in behavior remains unexplained.

Disaggregated behaviors. Appendix C.5 shows that the above results do not depend on the specific way the twelve behaviors are aggregated: they generally hold separately for each

Figure 3: Non-parametric relations between perceptions and behavior



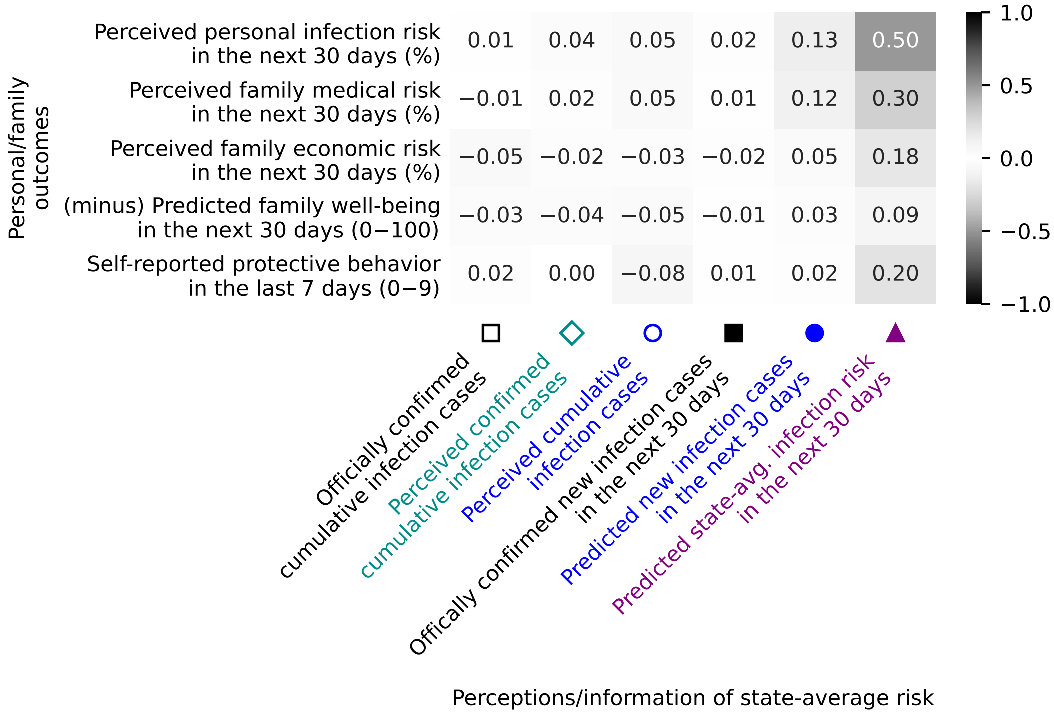
Notes: Triangles and circles: mean number of private behaviors by decile (their horizontal location is the exponentiated mean log percent perceptions within each decile). Error bars: bootstrapped 95% confidence intervals. Lines: OLS-regression estimates from columns (1) and (2) of Table 2. β's: regression coefficients (SEs).

behavior.

Non-parametric relation estimates. Figure 3 shows the regression lines from columns (1) and (2) of Table 2, as well as private-behavior averages by perception deciles. We find no evidence for non-monotonic relations, and the linear correlations and regressions above seem to summarize the (non-parametric) relations reasonably well. In the left panel, the bottom risk-perceptions decile (who perceive on average a 0.2 percent state-average infection risk) report on average 4.0 protective behaviors, while the top decile (66.0 percent) report 5.8—an increase of 0.8 standard deviations. In the right panel the relation is flat.

Behavior/perceptions order effects. While our correlational data do not allow us to identify a causal effect of risk perceptions on protective behavior, our findings can at least rule out the possibility that elicited beliefs are generated ad-hoc to merely match the protective behavior subjects have just reported (e.g., in order to appear consistent or avoid cognitive dissonance). Indeed, columns (4)–(6) in Table 2 show that when the behavior module F is

Figure 4: Correlations of Perceptions and Different Outcomes



Notes: Correlations of officially confirmed cases, case perceptions and risk perceptions shown in Figure 1 with additional outcome variables, listed on the vertical axis. For full correlation tables see Appendix C.8.

presented first rather than last (the baseline), the coefficient on risk perceptions *decreases* (by 0.06–0.07 (SEs 0.02–0.03)). More generally, that these order effects are so much smaller than the baseline risk-perceptions coefficient (0.19–0.25 (all SEs 0.02)) is reassuring.

Correlations of perceptions with additional personal and family outcomes. Figure 4 extends the correlations bar from Figure 1's panel (c)—replicated as Figure 4's bottom row—to four additional outcome variables: three of personal and family risk perceptions and one of predicted family well-being. The original finding (bottom row), that perceived state-average infection risk is a dramatically stronger predictor of reported personal behavior than case perceptions and official reports, extends to these additional personal/family outcomes (top four rows).

Within-respondents correlations. Appendix C.6 further investigates the relations between perceptions and these four additional outcomes using only *within-respondent* variation.[17] It uses the subsample of respondents who completed the survey twice—once before and once on or after June 2 (N = 5,236 responses by 2,618 respondents). We find that controlling for individual fixed effects, Risk perceptions (as in Table 2) remain a stronger predictor than Case perceptions both when the dependent variable is perceived personal infection risk (the coefficients are 0.72 (SE 0.08) and 0.07 (0.04), respectively) and when it is family medical risk (0.34 (0.08) and 0.03 (0.04)). Risk and Case perceptions are both weak predictors when the dependent variable is family economic risk (0.16 (0.07) and −0.05 (0.08)) or (minus) predicted well-being (0.53 (0.39) vs. 0.19 (0.36)).

Explaining behavior-predictive perceptions with other variables. Can risk perceptions, which predict reported behavior as well as the additional personal and family outcomes in Figure 4, themselves be predicted from other variables? We find in Appendix C.7 that not only are they very far on average from case perceptions and official cases, risk perceptions are also hard to explain using other observables. Using all relevant available variables—officially confirmed cases (at present and newly added in the next 30 days), demographic characteristics, and state and day fixed effects—an $\mathrm{R}^2 = 0.08$ suggests that little of the variance in risk perceptions is explained. At the same time, within the subsample of respondents who completed the survey twice, adding individual fixed effects increases the explained variance to $\mathrm{R}^2 = 0.81$ (from a baseline $\mathrm{R}^2 = 0.14$ in this subsample).[18] This finding suggests that the main determinant of risk perceptions in our data is a stable personal characteristic, consistent with, e.g., Giglio et al.'s (2021) findings in a stock-market context (see Manski 2018 for more findings on intra-personally stable characteristics of beliefs).

[17] We cannot investigate within-respondent variation in self-reported behavior because it was only collected after respondents were allowed to re-participate in the survey.

[18] Case perceptions are also poorly explained by these observables (R^2 = 0.11, 0.17 and 0.70, respectively, in the main sample, the responding-twice subsample, and that latter subsample including individual fixed effects); however, recall that case perceptions are on average dramatically closer to official cases at the state and day level. In comparison, perceptions of cumulative infections at present (hollow circles in Figure 1) are explained better (R^2 = 0.22, 0.33 and 0.75, respectively).

3 The Role of Elicitation Details

We study the role, in our three main findings, of several elicitation details, including response format and question language. We use our original survey and an additional survey conducted in early 2021.

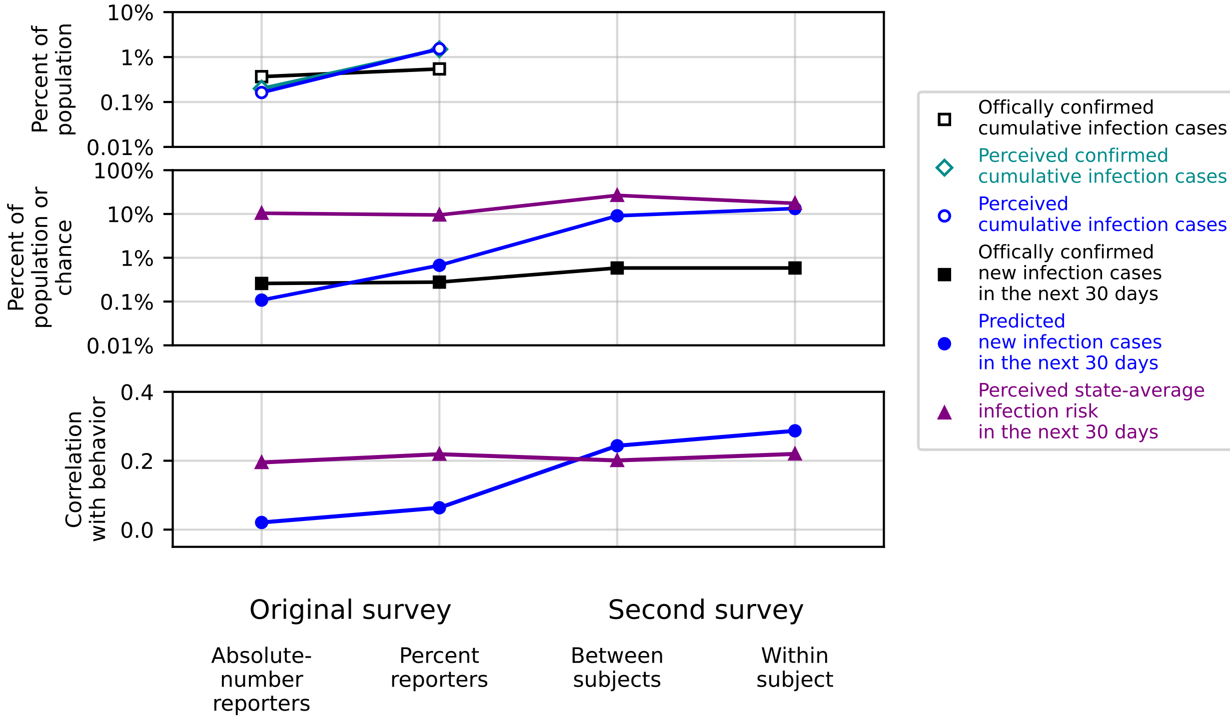
Role of response format: non-causal evidence. Respondents in our survey report risk perceptions using percent chance (Figure 2b on page 12). However, they can choose to report case perceptions using either percent of population or absolute number of cases (Figure 2a). Respondents who choose to report absolute numbers—92 percent of all responses—may think about cases differently than those who choose to report percentages—the remaining 8 percent.[19] In particular, the latter actively use the same response format (percent) for reporting their case perceptions as all respondents use for reporting risk perceptions. Do they also report case perceptions that look more like risk perceptions?

Figure 5's two leftmost columns show that they generally do. The two columns compare the perceptions from Figure 1 and their correlations with behavior between absolute-number reporters and percent reporters. The middle panel shows that percent reporters indeed have a smaller Risk-Cases gap: their perceived state-average infection risk (triangles) is 14 times larger than predicted newly infected population percent (circles), compared with 92 times larger for absolute-number reporters. Not reported in the figure, percent reporters' case and risk perceptions are also more strongly correlated (0.25) than absolute-number reporters'

[19] Recall, as Figure 2a shows, our dual-format interface provides automatic, real-time, on-screen translation of one format into the other; we use rounding patterns to distinguish responses typed in using either format (see Appendix C.9 for details and robustness).

That so many of our respondents choose to use the absolute-number format is not surprising: in addition to appearing first on the screen (potentially making it a natural default), at the time of our study the absolute-number format was prominently used both by official sources and in the popular media to communicate information about COVID-19 cases. For example, as of August 19, 2020, absolute numbers were the default data (a) on a graph provided by Google following a Google search of the phrases "COVID 19 new cases," "COVID statistics," or similar phrases; (b) on data websites such as www.cdc.gov, www.worldometers.info, and coronavirus.jhu.edu/map.html; and (c) on the most popular news websites, including the following four out of the five most visited websites (according to ebizmba.com, ordered by popularity): Yahoo!, Google News, HuffPost, and New York Times (which reported both absolute numbers and proportions). (The fifth, ranked 4 on ebizmba.com, is CNN. Its default presentation is proportion out of 100,000 people.)

Figure 5: Main Results' Dependence on Elicitation Details



Notes: Two left columns: Figure 1's results, averaged over all states and days, split by respondents' self-selection to report case perceptions using absolute numbers ($N = 12,142$) vs. using percentages ($N = 1,014$) in the original survey. Two right columns: Figure 1's results from a second survey ($N = 1,530$), in which both case perceptions and risk perceptions are reported using percentages and are asked using similar wording. Top and middle panels: average officially confirmed cases, case perceptions and risk perceptions, listed in the legend. Bottom panel: correlation coefficients with self-reported (private) protective behavior. Error bars (top and middle panels; sufficiently narrow that they are hidden behind the markers) indicate bootstrapped 95% confidence intervals.

(0.12). At the same time, risk perceptions (triangles) are essentially identical across the two groups, and the smaller Risk-Cases gap is almost entirely explained by case perceptions (circles), which are higher among percent reporters than among absolute-number reporters. Relatedly, percent reporters overstate, rather than understate, present and future official counts (compare, across the two columns, hollow circles vs. squares in the top panel, and solid circles vs. squares in the middle panel), yielding positive Present Cases and Future Cases gaps (2.8- and 2.4-fold overstatement, respectively, compared with 56 and 58 percent understatement among absolute-number reporters). The bottom panel shows that the correlations between risk perceptions and behavior are 0.22 and 0.19 respectively; between case perceptions and behavior they are 0.06 and 0.02.

These results are qualitatively unchanged when controlling for demographics and fixed effects (see Appendix C.9), implying that the chosen response format is an important explanatory variable, only weakly explained by our other variables. We conclude that while this evidence is correlational—respondents self-select into the absolute-number- and percent-reporting groups—it points to a potential central role for response-format differences in the risk-cases inconsistency.

Joint role of response format and other elicitation details: evidence from a second survey. Due to self selection, causality in the above analysis could run in both directions. In addition, other elicitation differences in our survey between case and risk perceptions may have an even larger role in our main findings. These include, e.g., question wording and style, user interface, case perceptions being based on two questions rather than one, and the salient display of a state's population only in case-perceptions questions. To investigate the overall effect of these elicitation details jointly, we conducted another short survey. It includes only three questions, on three different pages, randomly ordered: case perceptions, risk perceptions, and self-reported behavior (see Appendix A.5 for screenshots).[20] Importantly, case perceptions are elicited more similarly to risk perceptions. Modifications include, e.g., replacing "How many people" with "What percent of the population"; using a simple, single-format (percent) textbox as in Figure 2b for responses; and not displaying total state populations.[21]

Figure 5's two rightmost columns show that in this second survey, case perceptions be-

[20] *Case perceptions question*: Give your best estimate: what percent (0-100) of the population in *[state]* will get infected with the coronavirus during the next month?

Risk perceptions question: Different people in *[state]* have different chances to get infected with the coronavirus during the next month. These chances depend on many things, such as personal circumstances, lifestyle, and behavior. Give your best estimate: what is the average chance (0-100 percent) for a person in *[state]* to get infected with the coronavirus during the next month?

Private behaviors: washing hands often, sanitizing groceries, avoiding touching face; *public behavior*: avoiding public spaces and crowds. (We chose this subset because it is diverse and has $R^2 = 0.96$ in a regression explaining the sum of Yes answers to all twelve behaviors in our main survey.)

[21] The questions were incorporated as a last module in a larger survey on a different topic (consumer expenditures), conducted between February 8, 2021 and March 10, 2021, on an online US sample that matches the adult population on several key demographics ($N = 1{,}530$). See Appendix A.5.

come qualitatively similar to risk perceptions, in both between-subjects ($N = 1,031$) and within-subject (N = 1,530) analyses.[22] In the middle panel, the Risk-Cases gap (solid triangles vs. circles) dramatically shrinks, from a 79-fold overestimation in the original survey's main sample to mere 2.9- and 1.3-fold overestimation, respectively, between and within subjects, while both case and risk perceptions dramatically overstate official numbers, with a Future Cases gap (circles vs. squares) growing to 16 and 23-fold *over*estimation. In the bottom panel, case perceptions become a bit *more* predictive of behavior than risk perceptions (correlation of 0.24 vs. 0.20 between- and 0.29 vs. 0.22 within-subjects), which remain as predictive of behavior as in the main sample (correlation of 0.20). (Not reported in the figure, case and risk perceptions themselves become highly correlated, at 0.76.)

In summary, our purposefully designed "case-less" case-perceptions question, or "cases-as-percent"-perceptions question, behaves qualitatively like a risk-perceptions question: it elicits gross overestimates of case numbers yet it is correlated with reported behavior. We draw two conclusions, both in line with the main message of this paper. First, highlighting the importance of the "Which Beliefs?" question in the paper's title, elicited beliefs crucially depend on elicitation details that standard economic theory is agnostic about. In our case, we find the differences between case and risk perceptions to be a function of question structure, wording and response format more than of the specific mathematical object the question appears to target. Second, consistent with the rest of the paper's title, the perceptions we elicit in our surveys are related to either information or behavior, but not both.

4 Discussion

We now relate our three main findings to existing evidence from psychology and economics, focusing on health contexts and in particular on COVID-19. We then discuss potential implications for public communication of health-risk information.

[22] For case and risk perceptions, respectively, the between-subjects analysis includes data from the first two pages only, for respondents who saw the survey orders [Case↔Behavior]–Risk and [Risk↔Behavior]–Case.

4.1 Existing Evidence of Inconsistent Beliefs

Our first finding that case perceptions are not far from official information benchmarks—i.e., limited Present and Future Cases gaps—and our second finding of an inconsistency between mathematically equivalent risk and case perceptions—i.e., a large Risk-Cases gap—are related to previously documented biases in people's elicited beliefs. One strand of literature finds that beliefs depend on whether they are elicited as probabilities (e.g., 10 percent or 1 in 10 probability) or as relative frequencies, i.e., as absolute counts in an imaginary sample (e.g., 100 out of 1,000 people). Gigerenzer and Hoffrage (1995) find that some well-documented probability-reasoning biases such as base-rate neglect or the conjunction fallacy significantly shrink in magnitude when both the questions and the responses are communicated using relative frequencies rather than probabilities. Slovic et al. (2000) find that psychiatrists' estimates of the risk that discharged patients will be violent are lower when they use relative frequencies rather than probabilities.

Another strand of literature finds that when asked to explicitly relate probabilities to their corresponding relative frequencies, a large proportion of people do not give the expected answer. For example, Galesic and Garcia-Retamero (2010) find that only 57.7% of a US-representative sample give the answer "10" to the question: "In the Bingo Lottery, the chance of winning a $10 prize is 1%. What is your best guess about how many people will win a $10 prize if 1,000 people each buy a single ticket for Bingo Lottery? ____ person(s) out of 1,000." See also Woloshin and Schwartz (2011).

Our evidence appears generally consistent with the findings of both strands of literature. We find that respondents (a) report more accurate perceptions and predictions using relative frequencies (out of their state's population); and (b) fail to relate population percentages to average probabilities, even when the questions are adjacent. Section 3 highlights that the main drivers of such results may be subtle "look and feel" differences between elicitation questions, e.g., in response format or wording, rather than the *conceptual* difference between probabilities and relative frequencies. Supportive of this view, Bordalo et al.'s (2020) respondents predict that around 5 percent of the population will get infected with COVID-19—getting closer to our respondents' high risk-perceptions levels—even though they are asked about infection *frequencies* (i.e., cases out of 1,000 people) and not about *probabilities.*[23]

4.2 Existing Evidence on the Relationship Between Beliefs and Protective Behavior

Our third main finding—that risk perceptions predict behavior better than case perceptions—contributes to a literature comparing belief-elicitation questions by their power to predict behavior, e.g., Windschitl and Wells (1996), Weinstein et al. (2007), and Dillard et al. (2012). While we only explore *numeric* response scales, that literature also explores response scales with *verbal* descriptions of uncertainty levels, e.g., asking "Without a flu shot, do you think you're likely to get the flu this year?" with six response options: "extremely likely," "very likely," "somewhat likely," "somewhat unlikely," etc. Two main findings of that literature are that eliciting beliefs using such verbal scales often predicts behavior better than using numeric scales, and that questions that involve a language of feelings predict better than questions with more objective language. Within the domain of numeric scales, we are, to the best of our knowledge, the first to compare the predictive power of behavior of relative-frequency perceptions (cases out of a given sample) with that of percent-chance perceptions. Because we compare two objective-language, numeric questions that can be directly interpreted as probabilities, our contribution is especially relevant for economic studies eliciting beliefs.

Another aspect of our third main finding is that the correlation between perceived risk and

[23] Their elicited frequencies are of future infection cases in the next 9 weeks among US subpopulations in May, 2020. Furthermore, their elicitation interface shows respondents, after they responded, their reported relative frequencies in terms of percentages and proportions out of 100,000 people, and allows them to revise their answers, somewhat similarly to our dual-format interface. This evidence may in addition suggest that our dual-format interface is also not the main driver of the relatively small Present and Future Cases gaps we observe.

protective behavior is positive. Theoretically, it could have either sign, change magnitude in different ranges, and even be nonmonotonic.[24] Empirically, that we find a positive correlation is consistent with previous findings in the literature on health-risk perceptions and risk-mitigating behavior, including findings in the COVID-19 context. Among studies similar to ours in their questions and analysis, Brewer et al. (2007) conduct a meta-analysis of twelve studies, mostly longitudinal, and report a pooled positive correlation of 0.26 between perceived flu infection risk (conditional on not being vaccinated) and getting vaccinated; and Allcott et al. (2020) find a positive standardized effect of 0.32 of perceived COVID-19 personal infection chances on self-reported social distancing, where beliefs are conditioned on a hypothetical scenario in which the person maintains a pre-COVID routine, controlling for individual characteristics and state fixed effects.[25] In contrast, Akesson et al. (2020) and Papageorge et al. (2020) find *negative* correlations between risk perceptions and protective behaviors.[26]

Other studies of COVID-19 are less similar to ours, but still find positive relations between risk perceptions and protective behaviors. These include, e.g., Fan et al. (2020), who study only differences between demographic groups, Wise et al. (2020), who only control for age in

[24] In a standard expected-utility model, holding behavior cost and personal characteristics constant, the sign of the protective-behavior response to increased risk depends on the returns to protective behavior. If washing hands, for example, reduces infection risk by a constant multiplying factor, then the higher the baseline risk, the higher the returns to washing hands, generating a positive sign. In contrast, if washing hands is less effective at high infection risk, the sign could be negative (or be first positive and then negative, for example).

[25] Notably, they also find a much weaker positive effect, of 0.07, of predicted future number of US COVID-19 cases on self-reported social distancing—strikingly similar to our relations between risk perceptions, case perceptions and behavior. Also like our findings, their case perceptions underpredict officially confirmed infections—though by a larger factor of 4.4. They use their case- and risk-perceptions measures to study partisan-differences, and do not investigate the inconsistencies between those perceptions.

[26] Studies that are most comparable to ours (a) elicit perceptions of risk that is exogenous to behavior and (b) attempt to identify within-individual correlations.

Criterion (a) is relevant due to a theoretical result, based on a standard expected-utility model, that any correlation sign between *endogenous* risk and behavior is consistent with our result of a positive correlation between *exogenous* risk and behavior (see Appendix E). That result renders comparisons with correlations of the former kind less informative. Criterion (b) is relevant because our analysis attempts to identify, to the extent possible given our data, within-individual correlations: we control for personal characteristics and external costs of protective behavior (by using either private behaviors or controlling for state and day fixed effects, e.g., in Table 2), and we report panel results (with outcomes different from behavior; see Section 2.3). Comparing our study to studies that do not include a rich set of controls, or that do not focus on private behaviors, is therefore less informative.

their analysis, and Bruine de Bruin and Bennett 2020 and Dryhurst et al. 2020, who elicit personal infection risk, which is endogenous to own behavior (in contrast to our exogenous state-average infection risk).

4.3 Implications for Public Communication of Health-Risk Information

The COVID-19 pandemic brought back important policy questions including how to effectively inform the public about risks and, separately, whether and how to communicate risk in order to induce behavioral change (especially when facing strong externalities). Our findings suggest that depending on policy goals, policymakers may want to explore alternatives to the case-count language that was so prominently used early in the pandemic.

The effective-communication question has become especially pronounced in light of large documented differences in protective behavior across demographic groups. In the US, for example, Allcott et al. (2020), Barrios and Hochberg (2020), Bruine de Bruin et al. 2020 and Fan et al. (2020) all find that people who consider themselves Democrats engage in more COVID-protective behavior than those who consider themselves Republicans. They also find that the behavior differences can be partially explained by differences in perceptions and media consumption. Consistent with these studies, we document in Appendix C.2 that (self-identified) Republicans' case perceptions are understated more than Democrats' (Future Cases gaps of 70 and 39 percent underestimation, respectively), and that Republicans perceive lower risks and engage less in protective behavior (4.7 private behaviors compared with Democrats' 5.2).

Such findings may lead policymakers to speculate, for example, that a national campaign providing the public with accurate facts—such as infection and death case counts—may help close such behavioral gaps. However, our second and third main findings, which continue to hold within each partisan group, may suggest otherwise: both Republicans and Democrats show huge Risk-Cases gaps (factors of 92 and 80); and their protective behavior is essentially uncorrelated with case perceptions (−0.04 and 0.00, respectively) while being moderately correlated with risk perceptions (even more so among Republicans: 0.25 and 0.15, respectively). Our findings therefore suggest that communicating accurate official counts more prominently may not be an effective way to close such cross-group gaps and, more generally, to affect behavior: in our data, both official count information and its perceptions are only weakly related to reported behavior.

At the same time, many studies do find *certain* risk-related information-provision policies to be rather effective. Shermohammed et al. (2021), for example, find that informing high-risk influenza patients about their high-risk status increases their likelihood to get vaccinated for flu by 5.7 percent. Outside the health-risk context, Abito and Salant (2019), for example, find that the demand for expensive extended warranties, largely related to exaggerated perceived failure-risk probabilities, decreases when these perceived probabilities are corrected. Such results may, again, depend on *how* information is communicated. Our finding that language matters is consistent with the possibility that, for example, communicating risk referring to *chances* or *population fractions* rather than *cases* may shift people's risk perceptions and behaviors—a possibility that should be further investigated. At the same time, our findings also suggest that to the extent that such communication is effective, it may in fact *reduce* protective behavior. As Figure 3 above shows, in our sample, respondents in the top risk-perceptions deciles, who engage the most in protective behaviors, grossly overpredict (30-day state-average) infection risk at 50 percent or above—arguably a public-panic level. To the extent that our correlations imply causation, correcting these respondents' risk perceptions may reduce their protective behavior and therefore worsen social health outcomes.

Finally, language and framing could also have other effects on behavior that should be considered and further investigated. Freeman et al. (2020) find that people rate the risk of dying of COVID-19 (conditional on infection) higher (on a scale from "very low risk" to "very high risk") when they are informed about the probability using spelled-out-fraction language that is similar to relative frequencies (e.g., a "120 out of 1,000" chance) than when using percent chances (e.g., a "12%" chance). In a different health domain, the Slovic et al. (2000) study mentioned above similarly finds that psychiatrists rate discharged patients' violence risk higher (on a "low"/"medium"/"high" scale) when they are informed about this risk using relative frequencies (e.g., "10 out of 100" patients) than when using probabilities expressed as percent chances (e.g., a "10%" chance). In our context, official communication of COVID-19 risk that uses case counts (i.e., frequencies) may therefore be a contributing *cause* of our finding of the public's exaggerated risk perceptions.

In summary, our evidence, while correlational, suggests that public-communication campaigns aimed at affecting protective behavior may benefit from exploring risk communication that emphasizes chances or population fractions, rather than case counts. Past evidence suggests in addition that public campagins aimed at correcting risk perceptions—irrespective of the potential protective-behavior consequences—may independently benefit from exploring alternatives to case counts. Yet, at the time of this writing—more than a year and a half into the pandemic—the case-count language that became the default way of public communication early in the pandemic appears to remain an overwhelmingly common standard.

5 Conclusion

In this study we elicit two types of forward-looking beliefs: about a population's future infection *case counts* and about the population's average future infection *risk*. While the two are mathematically equivalent, we find that they are reported at levels that are orders-of-magnitude apart and are only weakly correlated across respondents. Furthermore, we find this gap between beliefs to be closely related to an economically meaningful gap between two observable objects: information and protective behavior. Indeed, the correlation bar in Figure 1 shows that it is not only *beliefs* about COVID-19 infection case rates that are essentially unrelated to behavior, but also the *official* infection rate, i.e., the objective information. In our second survey, when replacing case language and numbers with caseless language and percentages, elicited beliefs move unrealistically away from information benchmarks but become correlated with behavior.

One implication of our findings is that merely communicating accurate factual information to people (e.g., about case counts) may be insufficient for affecting behavior even in life-and-death situations. Moreover, in our data, we do not find that such information is grossly misperceived or misremembered—indeed, our respondents report and predict case information with relatively limited (downward) bias. However, this information, while apparently successfully communicated and understood, does not seem to affect behavior, perhaps because it does not sufficiently shape behavior-relevant perceptions (e.g., risk perceptions). This distinction, between information that is merely understood (and remembered) and information that "sinks in" and affects behavior, should be further studied.[27]

Other implications concern researchers' ability to reliably elicit beliefs using standard survey questions. Of course, our findings may merely highlight the imperfection of existing belief-elicitation questions. Other questions that we have not tried may be related to both information and behavior (and, of course, behavior itself may not be perfectly captured by the *self-reported* behavior we elicit). But our findings may reflect deeper issues with elicited beliefs. That our modified case-perceptions question elicits beliefs that so easily move away from those elicited with our original case-perceptions question may support the idea that people report ad-hoc beliefs formed in response to the elicitation context and framing (e.g., Windschitl 2002; Benjamin 2019 reviews some other context dependencies). That our elicited risk perceptions have unrealistic levels and yet they are correlated with protective behavior may support the idea that elicited beliefs sometimes confound probabilities with preferences (see Manski 2018).

Finally, that the belief inconsistency we find is related to an apparent economically

[27] See Heffetz (2021) for a discussion of "sunk-in" beliefs in the context of expectations-based reference-dependent preferences. There, lab participants understand and remember objective probabilities; yet that demonstrably successful communication of information does not seem enough for those probabilities to become participants' reference point and affect behavior.

meaningful disconnect between information and behavior may have theoretical implications that go far beyond belief-elicitation measurement issues. Economic models routinely assume that information affects beliefs, and that beliefs affect behavior. Our findings highlight the possibility that information may affect *some* beliefs, while *some other* beliefs may affect behavior. Our findings may thus be consistent with work that questions the notion, standard in economics, that beliefs should be modeled as a *single* probability distribution over relevant outcomes. Dual process models, such as in Loewenstein et al. (2015), take a step in this direction by allowing beliefs to shape behavior in two different ways, associated with different processes. For example, choice may maximize a mix of "deliberate" utility, based on rational beliefs, and "affective" utility, based on probability-weighted beliefs. To explain our findings using such models, future research would need to investigate the relationships between different theoretically defined cognitive processes and different empirically elicited beliefs.

References

Abito, J. M. and Salant, Y. (2019). The effect of product misperception on economic outcomes: Evidence from the extended warranty market. *The Review of Economic Studies*, 86(6):2285–2318.

Akesson, J., Ashworth-Hayes, S., Hahn, R., Metcalfe, R. D., and Rasooly, I. (2020). Fatalism, beliefs, and behaviors during the COVID-19 pandemic. *NBER Working Paper*, w27245.

Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., and Yang, D. (2020). Polarization and public health: Partisan differences in social distancing during the Coronavirus pandemic. *Journal of Public Economics*, 191:104254.

Barrios, J. M. and Hochberg, Y. (2020). Risk perception through the lens of politics in the time of the COVID-19 pandemic. *NBER Working Paper*, w27008.

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1*, 2:69–186.

Bordalo, P., Coffman, K. B., Gennaioli, N., and Shleifer, A. (2020). Older people are less pessimistic about the health risks of COVID-19. *NBER Working Paper*, w27494.

Brewer, N. T., Chapman, G. B., Gibbons, F. X., Gerrard, M., McCaul, K. D., and Weinstein, N. D. (2007). Meta-analysis of the relationship between risk perception and health behavior: the example of vaccination. *Health Psychology*, 26(2):136–145.

Bruine de Bruin, W. and Bennett, D. (2020). Relationships between initial COVID-19 risk perceptions and protective health behaviors: A national survey. *American Journal of Preventive Medicine*, 59(2):157–167.

Bruine de Bruin, W., Saw, H.-W., and Goldman, D. P. (2020). Political polarization in us residents' COVID-19 risk perceptions, policy preferences, and protective behaviors. *Journal of Risk and Uncertainty*, 61(2):177–194.

Dillard, A. J., Ferrer, R. A., Ubel, P. A., and Fagerlin, A. (2012). Risk perception measures' associations with behavior intentions, affect, and cognition following colon cancer screening messages. *Health Psychology*, 31(1):106–113.

Driscoll, J. C. and Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics*, 80(4):549–560.

Dryhurst, S., Schneider, C. R., Kerr, J., Freeman, A. L., Recchia, G., van der Bles, A. M., Spiegelhalter, D., and van der Linden, S. (2020). Risk perceptions of COVID-19 around the world. *Journal of Risk Research*, 23(7–8):994–1006.

Fan, Y., Orhun, A. Y., and Turjeman, D. (2020). Heterogeneous actions, beliefs, constraints and risk tolerance during the COVID-19 pandemic. *NBER Working Paper*, w27211.

Feldman, N. and Heffetz, O. (2021). A grant to every citizen: Survey evidence of the impact of a direct government payment in Israel. *NBER Working Paper*, w28312.

Freeman, A. L., Kerr, J., Recchia, G., Schneider, C., Lawrence, A. C., Finikarides, L., Luoni, G., Dryhurst, S., and Spiegelhalter, D. J. (2020). Communicating personalised risks from COVID-19: Guidelines from an empirical study. *medRxiv*.

Galesic, M. and Garcia-Retamero, R. (2010). Statistical numeracy for health: a cross-cultural comparison with probabilistic national samples. *Archives of Internal Medicine*, 170(5):462–468.

Gigerenzer, G. and Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102(4):684–704.

Giglio, S., Maggiori, M., Stroebel, J., and Utkus, S. (2021). Five facts about beliefs and portfolios. *American Economic Review*, 111(5):1481–1522.

Heffetz, O. (2021). Are reference points merely lagged beliefs over probabilities? *Journal of Economic Behavior & Organization*, 181:252–269.

Heffetz, O. and Rabin, M. (2013). Conclusions regarding cross-group differences in happiness depend on difficulty of reaching respondents. *American Economic Review*, 103(7):3001–21.

Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. *The Stata Journal*, 7(3):281–312.

Kennedy, R., Clifford, S., Burleigh, T., Jewell, R., and Waggoner, P. (2018). The shape of and solutions to the MTurk quality crisis. *Available at SSRN*, `https://papers.ssrn.com/abstract=3272468`.

Loewenstein, G., O'Donoghue, T., and Bhatia, S. (2015). Modeling the interplay between affect and deliberation. *Decision*, 2(2):55–81.

Manski, C. F. (2018). Survey measurement of probabilistic macroeconomic expectations: Progress and promise. *NBER Macroeconomics Annual*, 32(1):411–471.

Papageorge, N. W., Zahn, M. V., Belot, M., van den Broek-Altenburg, E., Choi, S., Jamison, J. C., Tripodi, E., et al. (2020). Socio-demographic factors associated with self-protecting behavior during the COVID-19 pandemic. *Institute of Labor Economics (IZA)*.

Robinson, J., Rosenzweig, C., Moss, A. J., and Litman, L. (2019). Tapped out or barely tapped? recommendations for how to harness the vast and largely unused potential of the Mechanical Turk participant pool. *PloS One*, 14(12):e0226394.

Shermohammed, M., Goren, A., Lanyado, A., Yesharim, R., Wolk, D. M., Doyle, J., Meyer, M. N., and Chabris, C. F. (2021). Informing patients that they are at high risk for serious complications of viral infection increases vaccination rates. *medRxiv*.

Slovic, P., Monahan, J., and MacGregor, D. G. (2000). Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and Human Behavior*, 24(3):271–296.

Weinstein, N. D., Kwitel, A., McCaul, K. D., Magnan, R. E., Gerrard, M., and Gibbons, F. X. (2007). Risk perceptions: Assessment and relationship to influenza vaccination. *Health Psychology*, 26(2):146–151.

Windschitl, P. D. (2002). Judging the accuracy of a likelihood judgment: The case of smoking risk. *Journal of Behavioral Decision Making*, 15(1):19–35.

Windschitl, P. D. and Wells, G. L. (1996). Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology: Applied*, 2(4):343–364.

Wise, T., Zbozinek, T. D., Michelini, G., Hagan, C. C., et al. (2020). Changes in risk perception and protective behavior during the first week of the covid-19 pandemic in the United States. *PsyArXiv*.

Woloshin, S. and Schwartz, L. M. (2011). Communicating data about the benefits and harms of treatment: a randomized trial. *Annals of Internal Medicine*, 155(2):87–96.

Web Appendix for

Which Beliefs? Behavior-Predictive Beliefs are Inconsistent with Information-Based Beliefs: Evidence from COVID-19

Ori Heffetz

Guy Ishai

October 26, 2021

A Data

A.1 Full Survey Questions and Data Coding

Self-reported behavior questions (module F). Health-protective behaviors elicited in module F, which are only briefly described in Table 1, include: 9 private-domain behaviors of increasing the hand-washing frequency relative to a pre-epidemic habit (by at least 5, 10 or 15 more times per day) (F1–F3); cleaning or sanitizing incoming mail and deliveries (F4); cleaning or sanitizing groceries (F5); cleaning or sanitizing furniture or frequently touched items (F6); avoiding touching own face (F7); stopping breath when passing near others (F8); coughing into elbow rather than palm (F9); and 3 public-domain behaviors of avoiding contact with people from a high-risk group (F10); avoiding meeting family and friends (F11); avoiding public spaces, gatherings and crowds (F12).

5 responses do not have answers to all 9 private-behavior questions, and 6 responses do not have answers to all 12 questions. These responses are excluded from behavior-relevant analyses.

Demographic Questions. The demographic questions in the end of the survey include: number of people in household (analyzed using 4 categories: 1/2/3/4 and above); number of people above 18 (same categories); gender (male/female/other); Hispanic origin (yes/no); race (White/Black/Asian/Native/other); age (6 categories: 18 (inclusive)–30 (not inclusive)/30–40/40–50/50–60/60–70/70 and above); education (10 categories); marital status (6); employment status (6); economic attitudes (7-point scale from very liberal to very conservative); social attitudes (same 7-point scale); political self-identification (Republican/Democrat/Independent/other/none); combined household income (8 brackets for \$0–\$200,000; or above \$200,000); medical insurance coverage (bad/fair/good); have been infected with COVID-19 (yes/no/prefer not to answer); someone from immediate family has been infected (same options).

Some responses seem to have mistakes or typos in reported birth years. We correct negative reported birth years to be positive; after this correction, birth years indicating an age below 18 or above 120 are marked as missing values (122 birth-year responses; 0.9 percent of the main sample). In other responses, not all demographic questions are answered. On total, 345 responses (2.5 percent of the main sample) have at least one missing value to a demographic question. Such missing values are treated as a separate category.

In some of the appendix analyses, demographics are coded binarily to investigate and present their effects in an easier fashion. The variables are coded binarily using the category(ies) in parentheses and its (their) negation: gender (Female); Hispanic origin (yes); race (not White); age (40 and above); education (less than 4-year college); marital status (not married); employment status (not working); economic attitudes (non-liberal); social attitudes (non-liberal); political self-identification (Republican); combined household income (less than 60K); medical insurance coverage (fair or bad); have been infected with COVID-19 (yes); someone from immediate family has been infected (yes). Missing values are pooled together with the negation of these definitions.

Other questions. The survey ends with an open question for general feedback. Other open questions were added for limited amounts of days just before the last question, to investigate some findings during the study. Since they were only added in the end of the survey in chunks of one or two questions at a time, we consider their effect on the whole survey as minimal and treat responses with answers to such questions similarly to other responses.

Screenshots. The following figures show screenshots of the entire survey.

Figure A.1: First page: IP screening

0% 100%

Warning!

This survey uses a protocol to check that you are responding from inside the U.S. and not using a Virtual Private Server (VPS), Virtual Private Network (VPN), or proxy to hide your country. In order to take this survey, please **turn off your VPS/VPN/proxy** if you are using one and also any ad blocking applications. Failure to do this might prevent you from completing the HIT.

For further information on why we are requesting this, see this post from TurkPrime (https://goo.gl/WD6QD4).

Next →

Figure A.2: Second page: Consent, state of residence and CAPTCHA screening

0% 100%

Thank you for participating in this survey! The survey asks about coronavirus outcomes. Your participation is voluntary, and is greatly appreciated. We do not anticipate any risks from participating in this survey. We anticipate that your participation in this survey presents no greater risk than everyday use of the internet. We are committed to protecting your privacy. After validating that your response qualifies for payment, your answers will remain anonymous and will take part in an aggregate statistical analysis only. You may withdraw from the survey at any time or skip questions you feel uncomfortable answering.

We anticipate that this survey will take you about 5 minutes to complete.

It is very important that you complete this survey on your own and answer truthfully, as we hope it will help understanding the virus situation better. **Responding without adequate effort may result in your responses being flagged for low quality.**

Additional Information

The main researcher conducting this study is Ori Heffetz, a professor at Cornell University. Please ask any questions you have at oh33@cornell.edu. If you have any questions or concerns regarding your rights as a subject in this study, you may contact the Institutional Review Board (IRB) for Human Participants at 607-255-5138 or access their website at http://www.irb.cornell.edu. You may also report your concerns or complaints anonymously through Ethicspoint online at www.hotline.cornell.edu or by calling toll free at 1-866-293-3077. Ethicspoint is an independent organization that serves as a liaison between the University and the person bringing the complaint so that anonymity can be ensured.

Please enter your MTurk Worker ID:

In which state do you currently reside?

To proceed, verify that you are a human:

I'm not a robot

reCAPTCHA

Privacy - Terms

Next →

Figure A.3: Module A

0% 100%

Answer **without** looking up the information: **as of today**, how many people in Colorado have been **reported by the authorities** as **infected** with the coronavirus since the beginning of the epidemic (including those who have already recovered or died)?

In this question and the next ones, enter a number of people or a percentage of the population, as you prefer.

Number of people
(enter without commas)

Percent of people in Colorado (0-100) %

Colorado's population 5,758,736

Answer **without** looking up the information: **as of today**, how many people in Colorado have been **reported by the authorities** as **dead** because of the coronavirus since the beginning of the epidemic?

Number of people
(enter without commas)

Percent of people in Colorado (0-100) %

Colorado's population 5,758,736

Next →

Figure A.4: Module B

0% 100%

Give your best estimates: How many people in Colorado will have been **infected** with the coronavirus since the beginning of the epidemic (including those who have already recovered or died)?

(The numbers may differ from the ones reported by the authorities)

As of today:

Number of people
(enter without commas)
Percent of people in Colorado (0-100) %
Colorado's population 5,758,736

As of a week from now:

Number of people
(enter without commas)
Percent of people in Colorado (0-100) %
Colorado's population 5,758,736

As of a month from now:

Number of people
(enter without commas)
Percent of people in Colorado (0-100) %
Colorado's population 5,758,736

Next →

Figure A.5: Module C

0% 100%

Give your best estimates: How many people in Colorado will have **died** because of the coronavirus since the beginning of the epidemic?

(The numbers may differ from the ones reported by the authorities)

As of today:

Number of people
(enter without commas)

Percent of people in Colorado (0-100) %

Colorado's population 5,758,736

As of a week from now:

Number of people
(enter without commas)

Percent of people in Colorado (0-100) %

Colorado's population 5,758,736

As of a month from now:

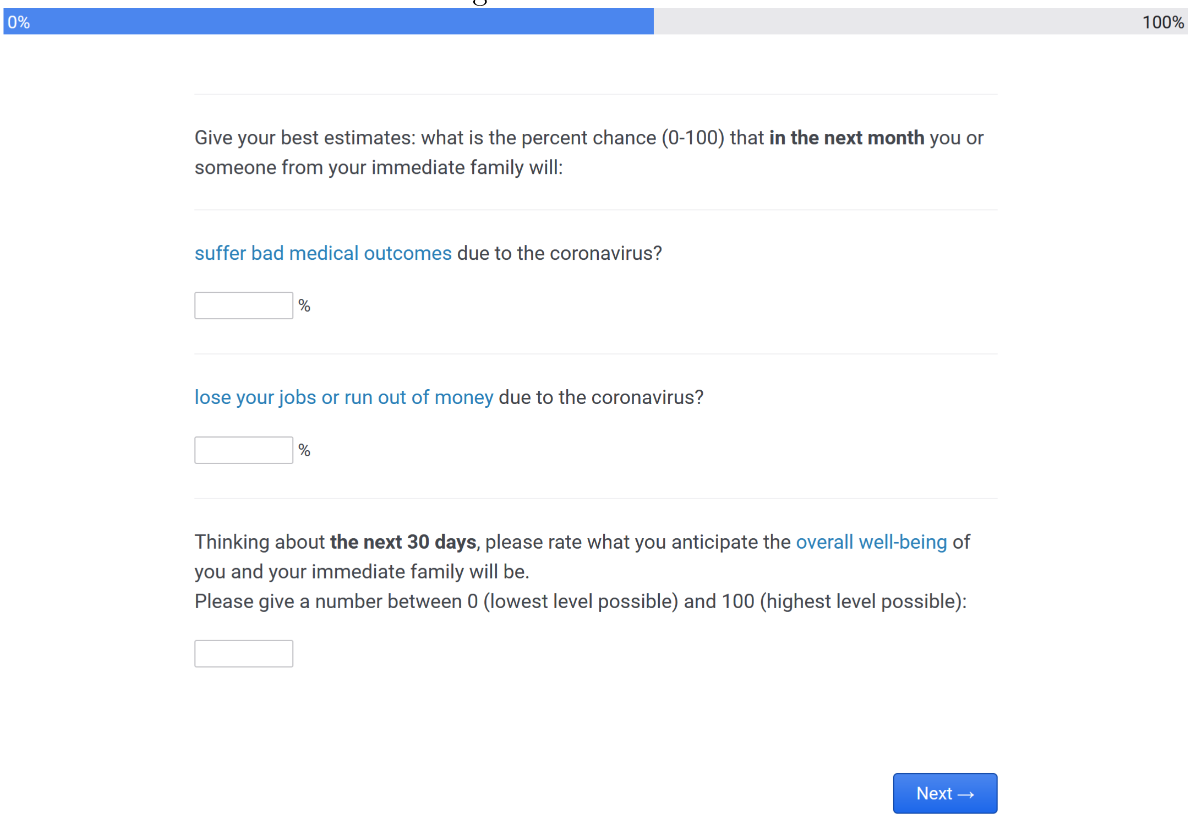
Number of people
(enter without commas)

Percent of people in Colorado (0-100) %

Colorado's population 5,758,736

Next →

Figure A.6: Module D



0% 100%

Give your best estimates: what is the percent chance (0-100) that **in the next month** you or someone from your immediate family will:

suffer bad medical outcomes due to the coronavirus?

%

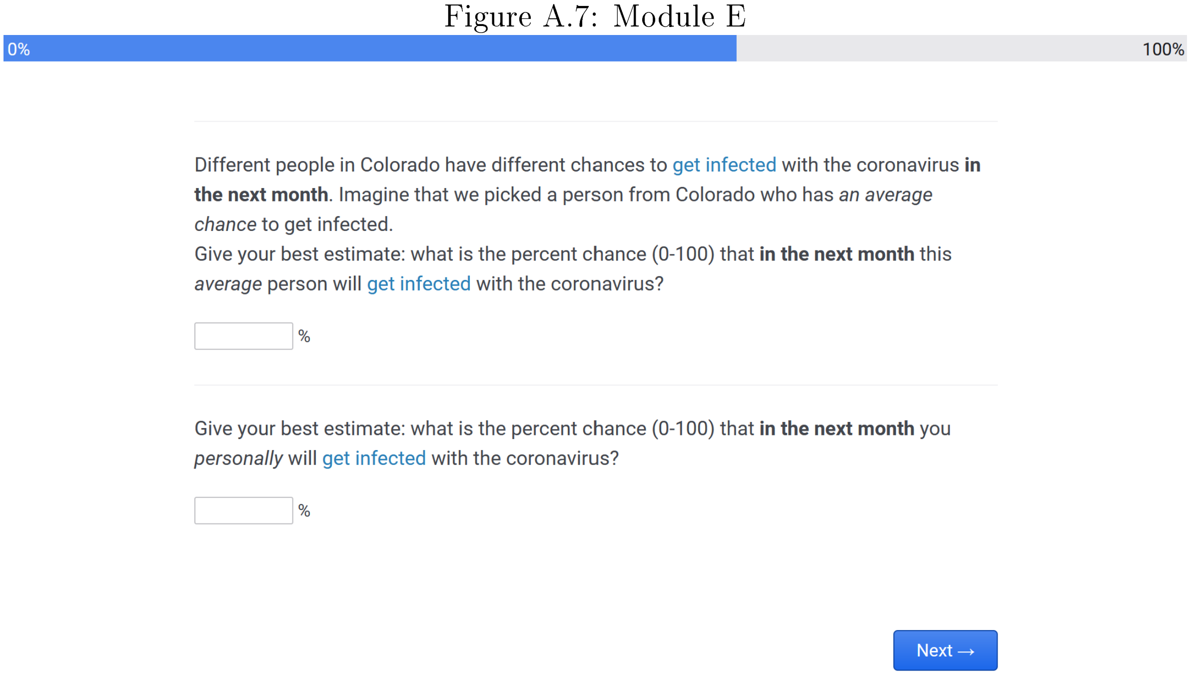
lose your jobs or run out of money due to the coronavirus?

%

Thinking about **the next 30 days**, please rate what you anticipate the overall well-being of you and your immediate family will be.
Please give a number between 0 (lowest level possible) and 100 (highest level possible):

Next →

Figure A.7: Module E



0% 100%

Different people in Colorado have different chances to get infected with the coronavirus **in the next month**. Imagine that we picked a person from Colorado who has *an average chance* to get infected.
Give your best estimate: what is the percent chance (0-100) that **in the next month** this *average* person will get infected with the coronavirus?

%

Give your best estimate: what is the percent chance (0-100) that **in the next month** you *personally* will get infected with the coronavirus?

%

Next →

Figure A.8: Module F

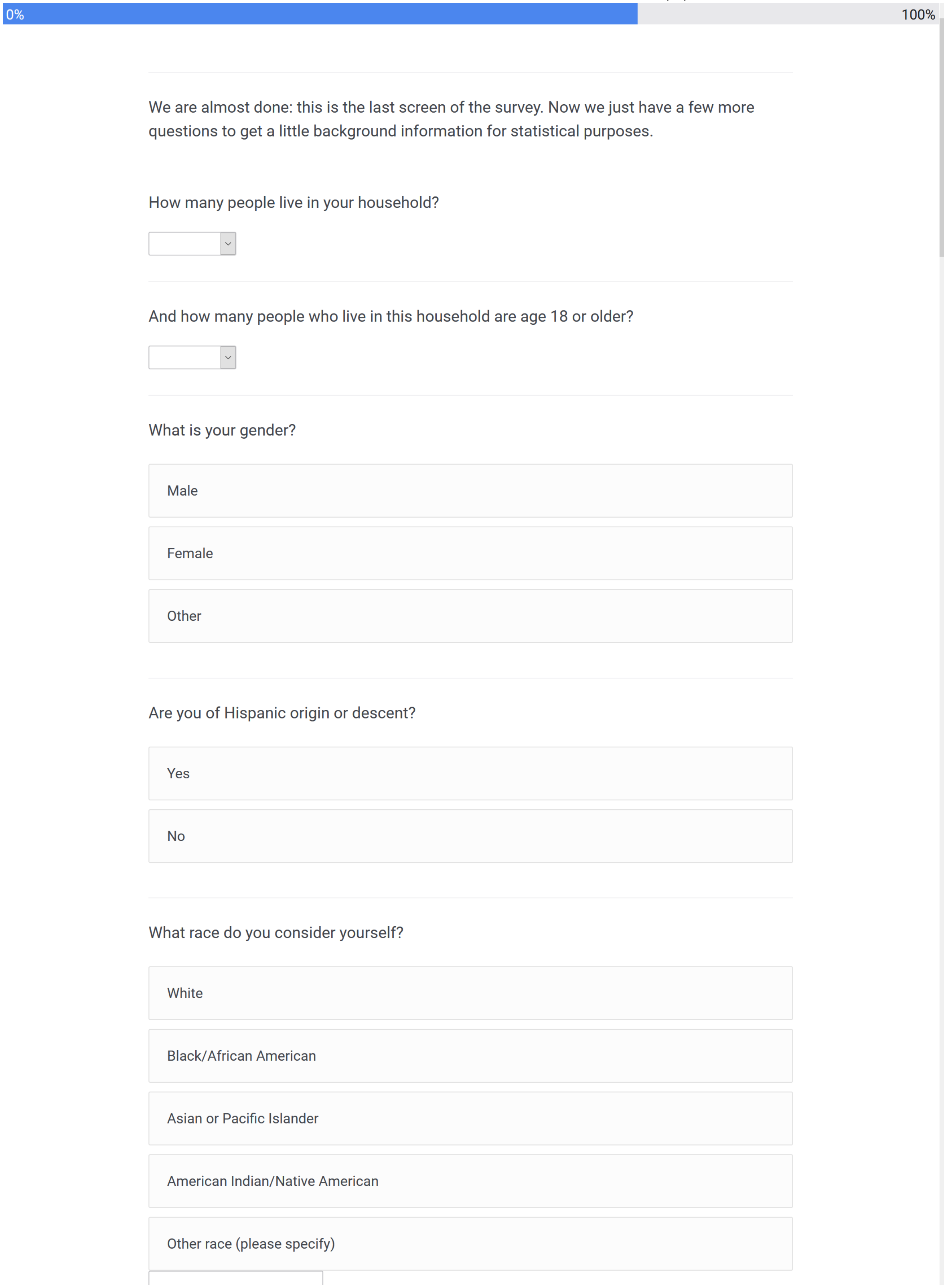
0% 100%

Which of the following have you done **in the last seven days** to keep yourself safe from coronavirus?
Only answer "Yes" for actions that you took or decisions that you made *personally*.

	No	Yes
Washed or sanitized your hands at least 5 **more** times per day **than your pre-epidemic habit**	○	○
Washed or sanitized your hands at least 10 **more** times per day **than your pre-epidemic habit**	○	○
Washed or sanitized your hands at least 15 **more** times per day **than your pre-epidemic habit**	○	○
Cleaned or sanitized incoming mail or deliveries	○	○
Cleaned or sanitized groceries	○	○
Cleaned or sanitized furniture or frequently-touched items such as: phone, wallet, keys, doorknobs, or steering wheel	○	○
Avoided touching your face	○	○
Coughed into your elbow rather than your palm	○	○
Stopped your breath when passing near other people	○	○
Avoided contact with people from a high-risk group (such as the elderly)	○	○
Avoided meeting family and friends	○	○
Avoided public spaces, gatherings or crowds	○	○

Next →

Figure A.9: Demographic questions (1)



0%

100%

We are almost done: this is the last screen of the survey. Now we just have a few more questions to get a little background information for statistical purposes.

How many people live in your household?

And how many people who live in this household are age 18 or older?

What is your gender?

- Male
- Female
- Other

Are you of Hispanic origin or descent?

- Yes
- No

What race do you consider yourself?

- White
- Black/African American
- Asian or Pacific Islander
- American Indian/Native American
- Other race (please specify)

Figure A.10: Demographic questions (2)

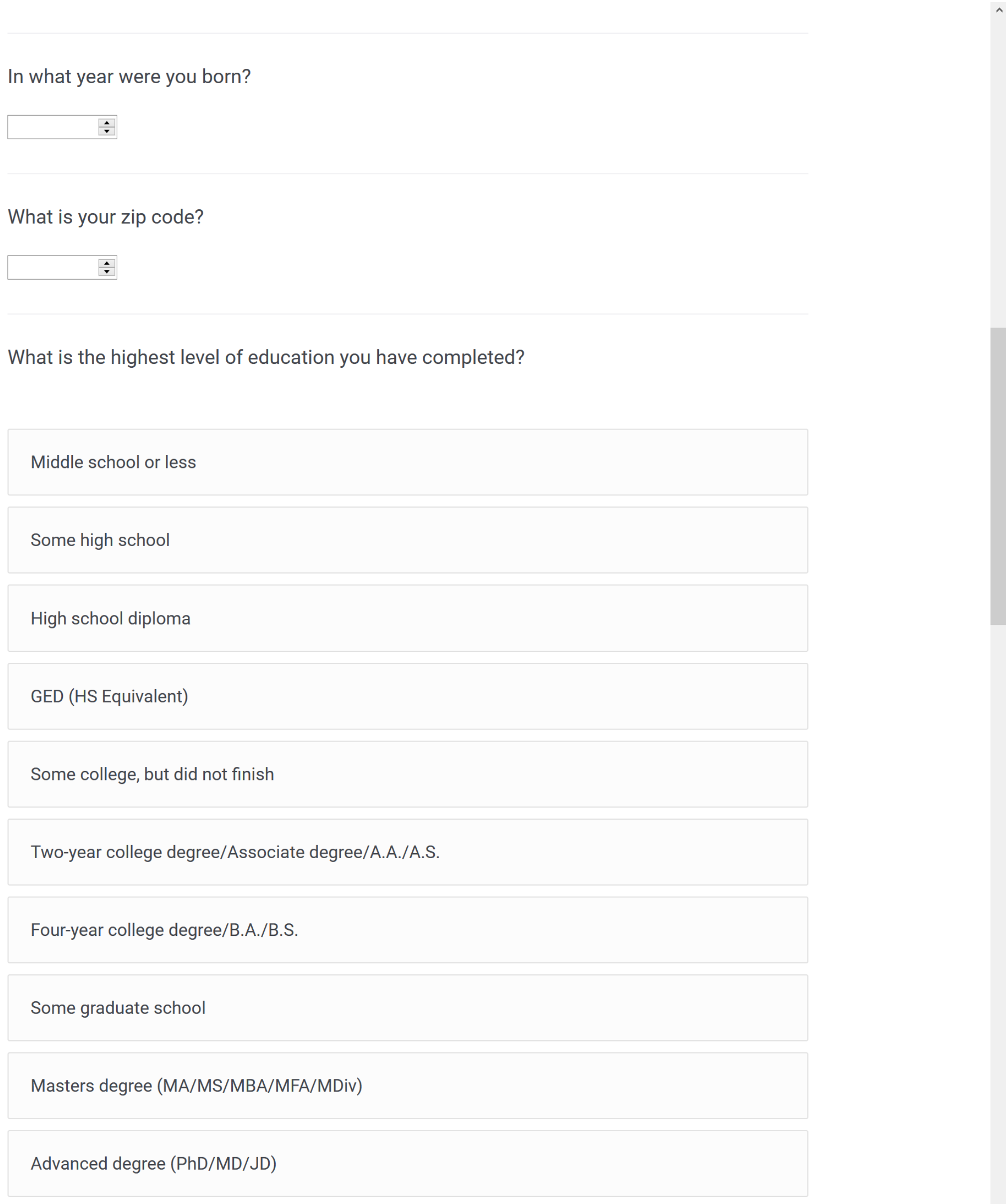


Figure A.11: Demographic questions (3)

What is your current marital status?

- Married
- Widowed
- Separated
- Divorced
- Single
- Living with a significant other

Which of the following best describes your main activity or your employment status?

- Working
- Unemployed (looking for work)
- Retired
- Keeping house
- Student
- Other

Figure A.12: Demographic questions (4)

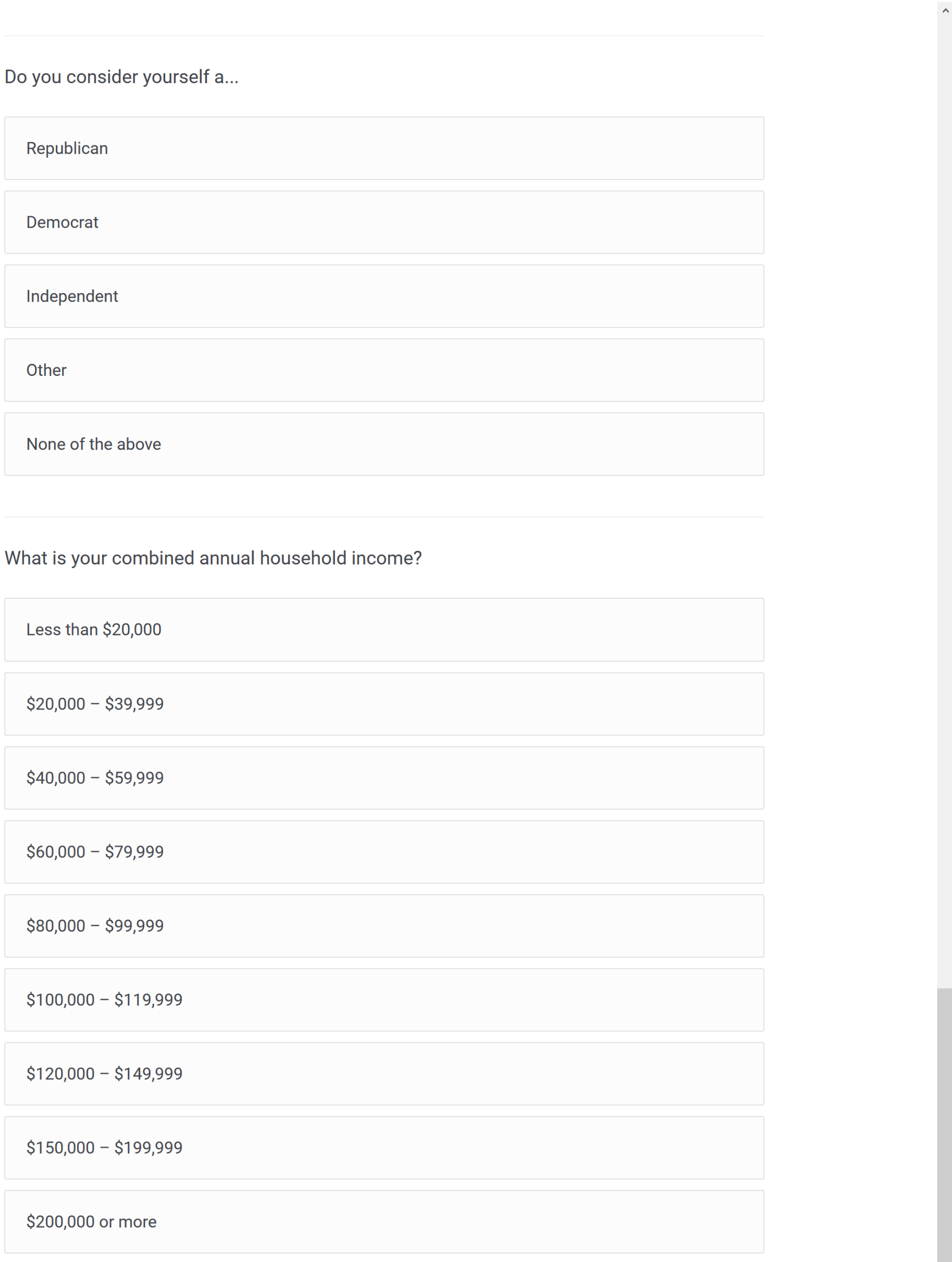
Thinking about *economic issues*, which of the following best describes your attitudes?

- Very liberal
- Liberal
- Slightly liberal
- Moderate
- Slightly conservative
- Conservative
- Very conservative

Thinking about *social issues*, which of the following best describes your attitudes?

- Very liberal
- Liberal
- Slightly liberal
- Moderate
- Slightly conservative
- Conservative
- Very conservative

Figure A.13: Demographic questions (5)



Do you consider yourself a...

- Republican
- Democrat
- Independent
- Other
- None of the above

What is your combined annual household income?

- Less than $20,000
- $20,000 – $39,999
- $40,000 – $59,999
- $60,000 – $79,999
- $80,000 – $99,999
- $100,000 – $119,999
- $120,000 – $149,999
- $150,000 – $199,999
- $200,000 or more

Figure A.14: Demographic questions (6)

How would you describe your medical insurance coverage?

- Bad or non-existent
- Fair
- Good

Have you been infected with the coronavirus?

- Yes
- No
- Prefer not to answer this

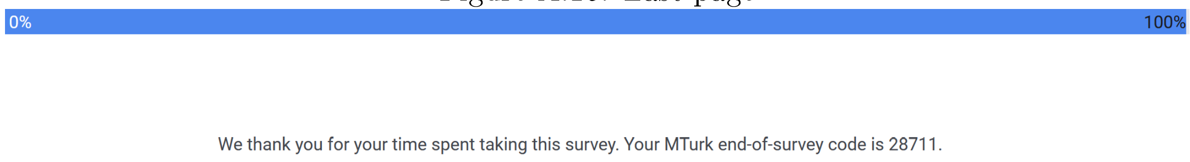
Have someone from your immediate family been infected with the coronavirus?

- Yes
- No
- Prefer not to answer this

Do you have any other comments about the survey, or thoughts you want to share about the coronavirus situation? Thank you again for participating!

Next →

Figure A.15: Last page



A.2 Complementary Versions of the Survey

We define the baseline version of the survey as the version with the module order A–(B↔C)–D–E–F (or without E and F prior to their addition). 9,586 observations in the full sample belong to the baseline version, and 9,135 in the main sample.

Four complementary versions of the survey differ from the baseline version's module order and MTurk filtering criteria, to test the effects of these factors on our results. The following subsections describe the details. The sample from first complementary version, R1, is treated separately and is not a part of our main full and main samples, while the samples from the other three complementary versions, R2, F-first and F-middle, are pooled with the baseline version's sample to generate our full sample ($N = 13,880$) and main sample ($N = 13,156$).

Version R1: testing robustness to experience of MTurk workers. While our main concern in the data collection—which we addressed using several measures—is low-quality MTurk responses, an opposite concern is over sophistication. The exclusion of inexperienced Workers or those with low approval rates generates a sample of highly experienced respondents, which may exhibit diminished effects relative to laypeople in some domains (e.g., Robinson et al. 2019). Specifically, many studies about the public response to COVID-19 were launched during the pandemic, raising the concern that experienced MTurk Workers may have have become COVID experts, with knowledge and attitudes that do not represent laypeople.

To understand our findings' sensitivity to MTurk Workers' experience, we launched another survey version, labeled R1, in parallel to the baseline version for three days between April 27, 2020 and April 29, 2020. The R1 version is identical to the baseline version in any detail except for the MTurk screening criteria, which was set, according to the recommendation in Robinson et al. (2019), to allow only Workers which are located in the US, who have successfully completed *at most 100 HITs* and whose approval rate is *at least 95%* to participate (for the screening criteria of the baseline version and the other three complementary versions, see Appendix A.3). These criteria imply two disjoint sets of respondents in the baseline version and in the R1 survey in the April 27 – April 29 period. 120 slots were opened for this survey version each day, and 309 responses were collected, from whom we keep 296 after full sample exclusion, and 255 after main sample exclusion.

Appendix C.2 investigates the effect of MTurk experience (using an R1 dummy) on some of the main results.

Version R2: testing robustness to the position of infection-risk perceptions elicitation (module E) in the survey. Module E, which elicits perceived state-average infection risk, is located close to the end of the survey and, for some respondents (depending on the randomized order of modules B and C) far from the closely related case perceptions questions in module B. This position is a compromise rather than an optimal setting, as the module was added a few weeks after the beginning of the study, and was placed in the end to minimize the addition's effect on the validity of previous results from the other modules.

To test the sensitivity of our findings to module E's position and to questions E1 and E2's order, we launched another survey version denoted R2 in parallel to the baseline version, for five days between May 4, 2020 and May 8, 2020. The R2 version is identical to the baseline version in any detail, except that questions E1 and E2 are now on two separate pages, 4 and 5, right after modules A–C (pages 1–3) and before module D (page 6). The order of E1 and E2 is random. 120 slots were opened for this survey version each day, and 426 responses were collected, from whom we kept 423 after full sample exclusion, and 408 after main sample exclusion.

Appendix D.2 investigates the order effects on the main results.

Versions F-first and F-middle: testing robustness to the position of health-protective behavior elicitation (module F) in the survey. A similar concern to the one above regards the position of module F in the survey. Its baseline position as the last module after modules A–E, is also a compromise, due to similar reasons to those mentioned above.

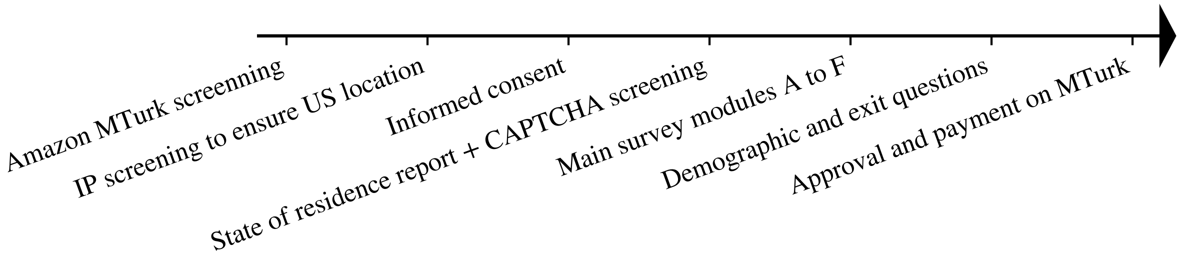
To test the sensitivity of our findings to module F's position, it was randomized for each respondent since June 19, 2020 to be either in the end of the survey (page 6, baseline version), in the middle (page 4, after B, C, and before D; F-middle version) or in the beginning (page 1, before A; F-first version). Versions F-first and F-middle were given the same sampling weight as the baseline version at the cost of reducing its daily sample size, since the validity and robustness of behavior-related results is a central issue. 2,032 observations were collected in the F-first version, and 1,960 in the F-middle version, from whom we kept 1,965 and 1,906 respectively after full sample exclusion, and 1,828 and 1,785 after main sample exclusion.

Table 2 and Appendix D.2 investigate the order effects on the main results.

A.3 MTurk Task and Screening Details

We use the Amazon MTurk crowd-sourcing platform to construct our sample. The survey timeline from the point of view of an MTurk Worker is shown in Fig. A.16.

Figure A.16: Survey timeline from an MTurk Worker's point of view

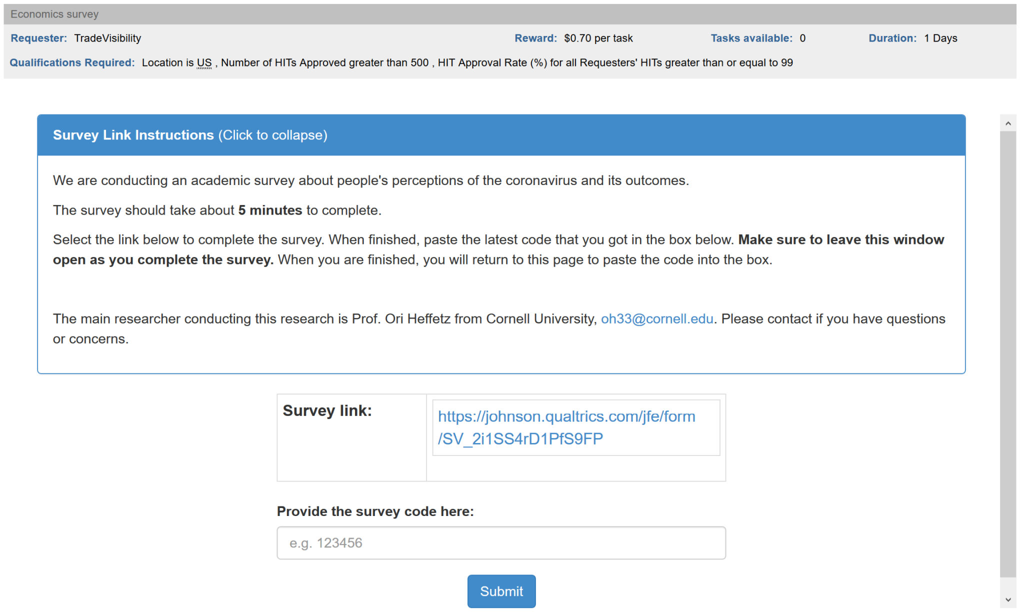


Slots opened. We publish an Amazon MTurk Human-Intelligence-Task (HIT) linking to the survey on a daily basis during the sampling period, usually around noon and no later than evening time ET. We typically opened 120 slots per day to collect all observations, except for versions R1 and R2, which were collected using additional slots on specific days (see Appendix A.2). The exceptions to the 120-slots-per-day rule are the first three days of the survey, March 24, 2020 to March 26, in which 100 slots were opened, the dates May 25 to June 1 and August 19 to August 23, in which 140 slots were opened to deal with a decreasing pool of available Workers that have not yet participated, and the dates of June 2 and June 3, in which 300 and 200 slots were opened respectively in order to observe possible discontinuities in results after resetting the sample. Figure A.19 shows the number of responses collected per day in each of the survey versions.

MTurk screening and task design. Only Workers located in the US, who have successfully completed at least 500 HITs, and whose approval rate is at least 99% are able to see our task recruitment page. This is a recommended practice by studies investigating response quality problems and fraudulent responses in the MTurk platform (e.g., Kennedy et al. 2018). Workers who enter our recruitment page are informed about the payment, which is \$0.7 per survey response (calculated based on a median response time of 5 minutes and a minimum wage of about \$6 per hour) and about the estimated time for completion, 5 minutes, and are provided with the researchers' contact details. If they choose to complete our survey, they are redirected to the survey page using a link. In the last page of survey they are provided with a code that they are requested to submit within the MTurk platform in order to be eligible for payment. Payment is guaranteed no later than two days after submission. Figure A.17 shows the MTurk recruitement page.

Screening non-US respondents. The survey is programmed and hosted on Qualtrics. As recommended by Kennedy et al. (2018) and using the methods they propose, the first page of the survey (see Figure A.1) provides an additional screening of non-US responses,

Figure A.17: MTurk recruitment page



which the MTurk platform may fail to detect, using the API of the website iphub.com. The second page provides an informed consent and a CAPTCHA screening of non-human respondents. Respondents who fail one of these screenings are informed immediately, are not allowed to advance and are not paid.

A.4 Sample Technical Details

Figures A.18, A.19 show the daily sample size by survey version during the sampling period, and the survey duration.

Figure A.18: Sample size by day and by survey version

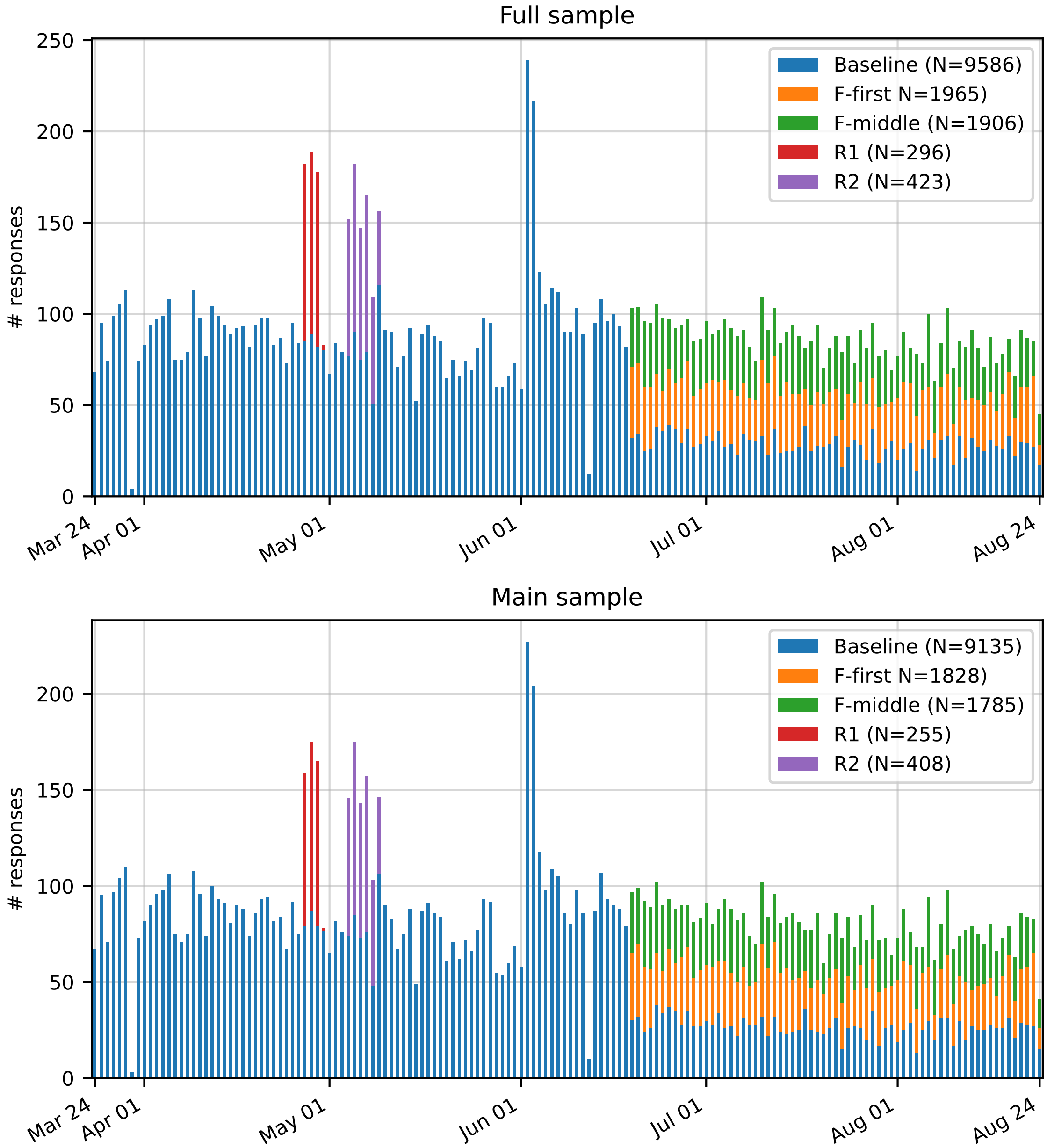


Figure A.19: Distribution of survey completion time and daily median completion times (main sample)

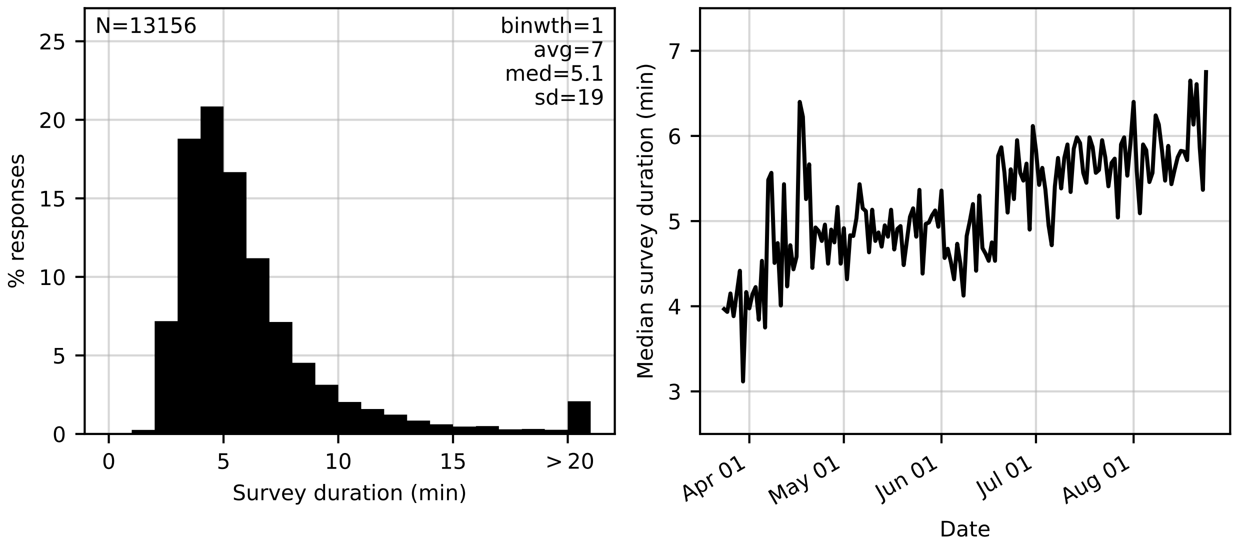
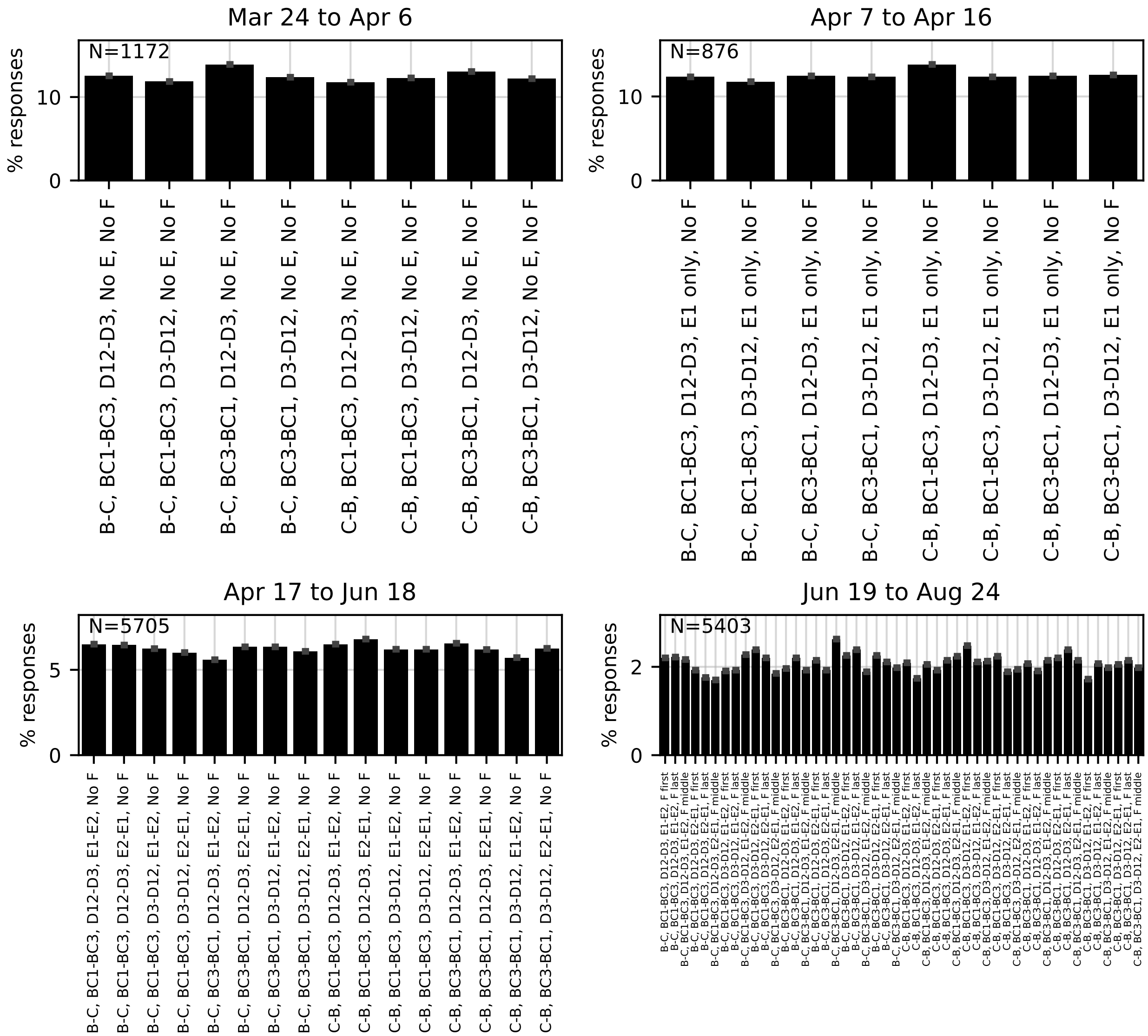


Figure A.20 shows the number of responses randomly allocated to each possible order of the survey and Table A.1 shows that demographics had no statistically-significant relation with any of the randomized orders.

Figure A.20: Distribution of responses randomized into each possible order of the survey (main sample)



Notes: Modules E and F were added to the study in the middle of the sampling period, hence the figure is divided to four periods: (1) prior to adding E and F, (2) after adding E1, (3) after adding E2, (4) after adding F.

Table A.1: Demographic variables relation with survey order

	C–B	BC3–BC1	D3–D12	E2–E1 or E1 only	F first/middle
At least 3 people in household	0.01	0.02*	−0.00	0.01	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
At least 3 people above 18 in household	−0.01	−0.02	0.02	0.01	−0.01
	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
Female	0.01	0.00	−0.02	0.01	−0.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Hispanic	0.01	0.00	−0.01	−0.02	0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Age at least 40	−0.00	0.01	0.01	−0.01	−0.04
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Not white	−0.02	0.01	0.01	−0.01	−0.05
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
Education less than 4-year college	0.01	0.01	0.01	0.00	−0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
Not married	−0.00	0.01	−0.00	−0.01	−0.05
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Not working	0.00	−0.01	0.03*	−0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Non liberal economic attitudes	0.03*	−0.04***	−0.01	0.01	−0.06
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
Non liberal social attitudes	−0.00	0.02	0.00	−0.00	0.02
	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
Republican	−0.03**	0.01	−0.00	−0.00	0.01
	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)
Not Democrat or Republican	−0.02*	0.01	−0.00	−0.00	0.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
Combined annual less than 60K	−0.00	−0.00	−0.01	−0.00	−0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
Fair or bad economic insurance	0.00	0.01	0.00	−0.01	0.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Have been infected	−0.03	−0.04	0.03	−0.06	−0.04
	(0.03)	(0.03)	(0.03)	(0.03)	(0.07)
Family member has been infected	0.00	0.01	−0.01	0.02	0.04
	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
Obs	13156	13156	13156	11108	5403
R^2	0.00	0.00	0.00	0.00	0.00

Notes: OLS regressions. Dependent variables: survey-order dummies. Independent variables: demographic characteristics coded binarily. In parentheses: Driscoll-Kraay standard errors using Bartlett's kernel and a bandwidth of 4 days.
$^{*}p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.005$

A.5 Second Survey

We incorporated versions of our three core questions about case perceptions, risk perceptions and protective behaviors in a survey conducted between February 8, 2021 and March 10, 2021, primarily aimed to study US residents' use of COVID-19 stimulus checks from the government (Feldman and Heffetz 2021). The last three pages of that survey, shown with random order, each include one of our three questions, as shown in the screenshots below. Questions are quoted in footnote 20 in the paper.

The survey was conducted by Qualtrics, who screened their respondents population based on age, region, gender, income, race and Hispanic-origin quotas to generate a sample that matches US adult population on these demographics. 1,530 observations were collected, which do not require further data exclusion.

9 responses do not have answers to all 3 private-behavior questions and also to all 4 questions. These responses are excluded from behavior-relevant analyses.

Screenshots. The following figures show screenshots of the survey.

Figure A.21: First page: intro page of last section

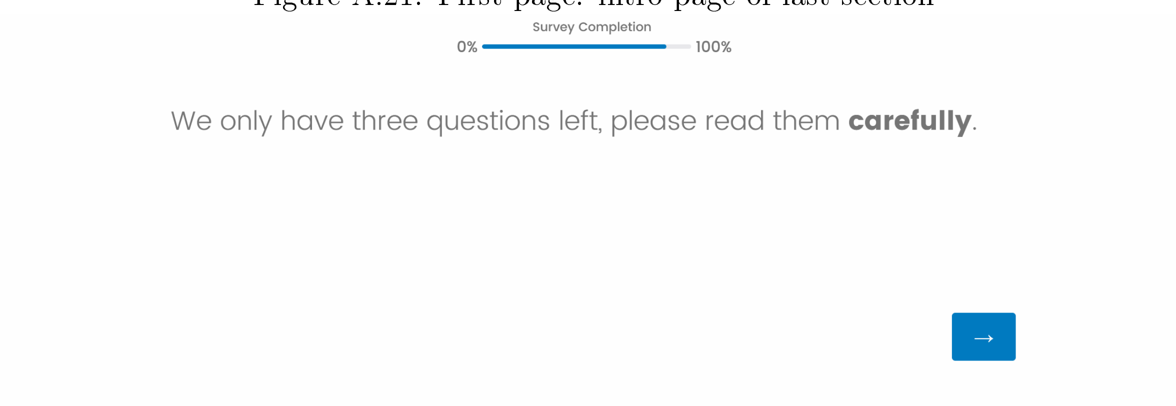


Figure A.22: Modified case perceptions question

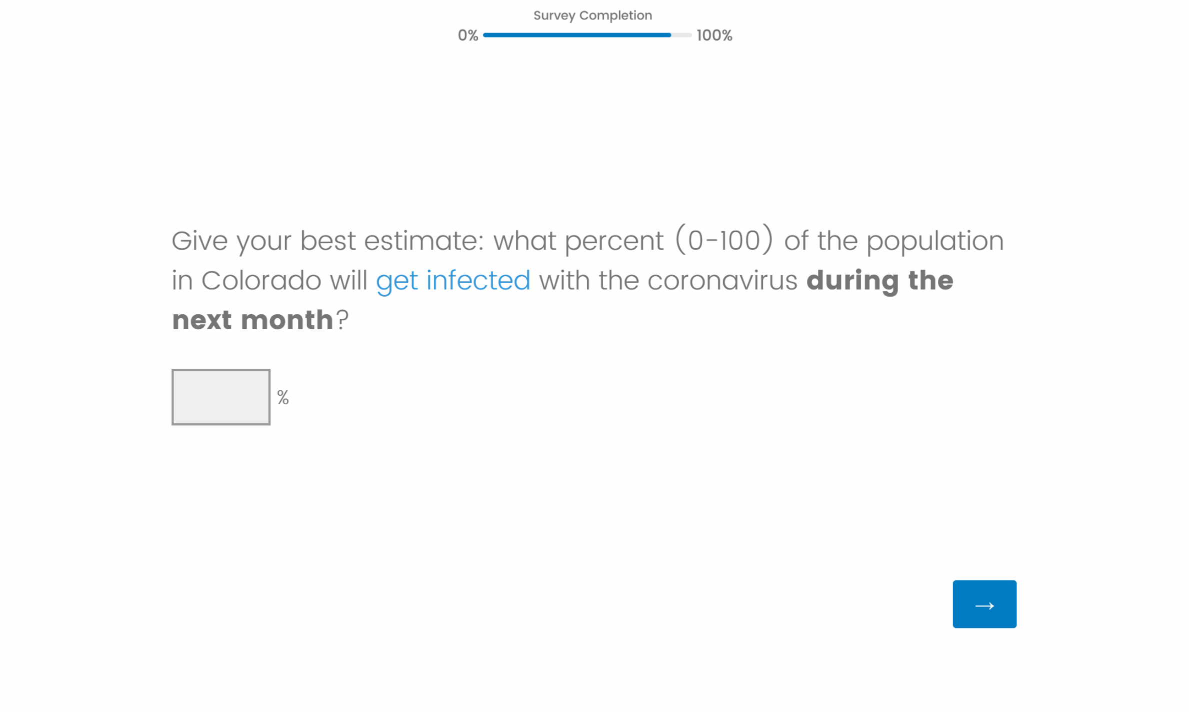


Figure A.23: Behavior questions

Survey Completion

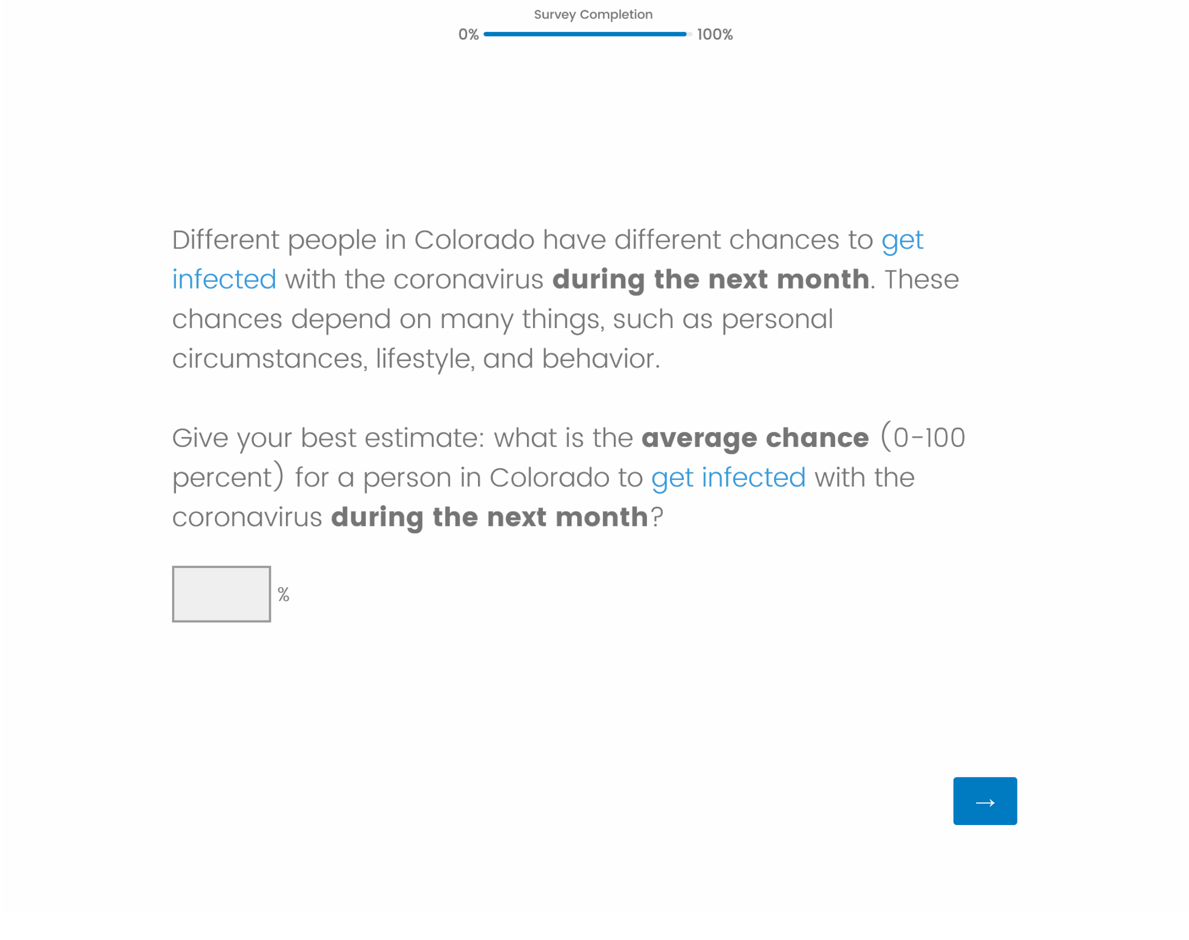
0% 100%

Which of the following have you done **in the last seven days** to keep yourself safe from coronavirus?
Only answer "Yes" for actions that you took or decisions that you made *personally*.

	No	Yes
Washed or sanitized your hands at least 5 **more** times per day **than your pre-pandemic habit**	○	○
Cleaned or sanitized groceries	○	○
Avoided touching your face	○	○
Avoided public spaces, gatherings or crowds	○	○

→

Figure A.24: Last page: slightly modified risk-perception question



B Descriptive Statistics

Figures B.1, B.2 show the distribution of demographic characteristics and US-state of residence in the main sample.

Figure B.1: Demographic characteristics of the main sample

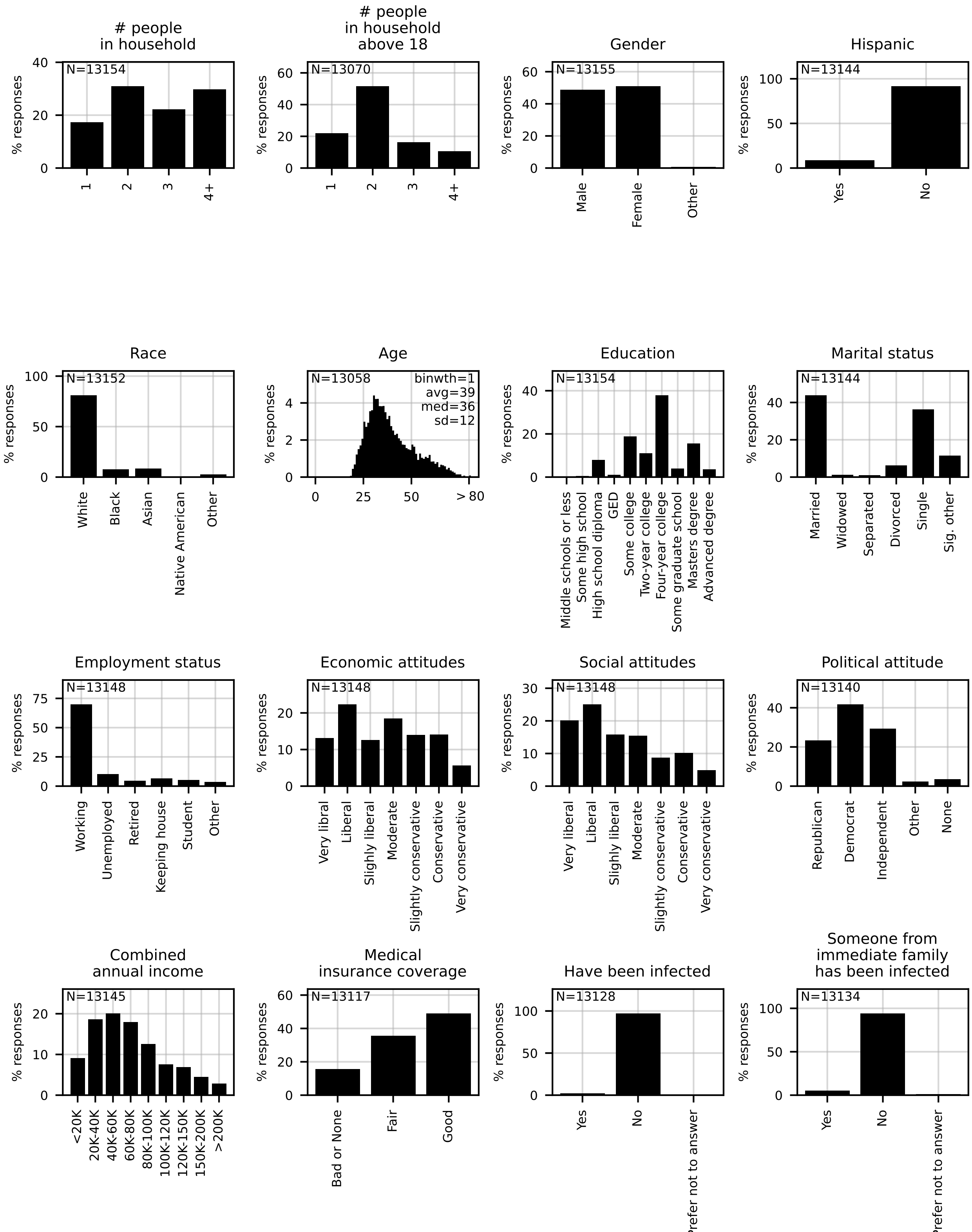
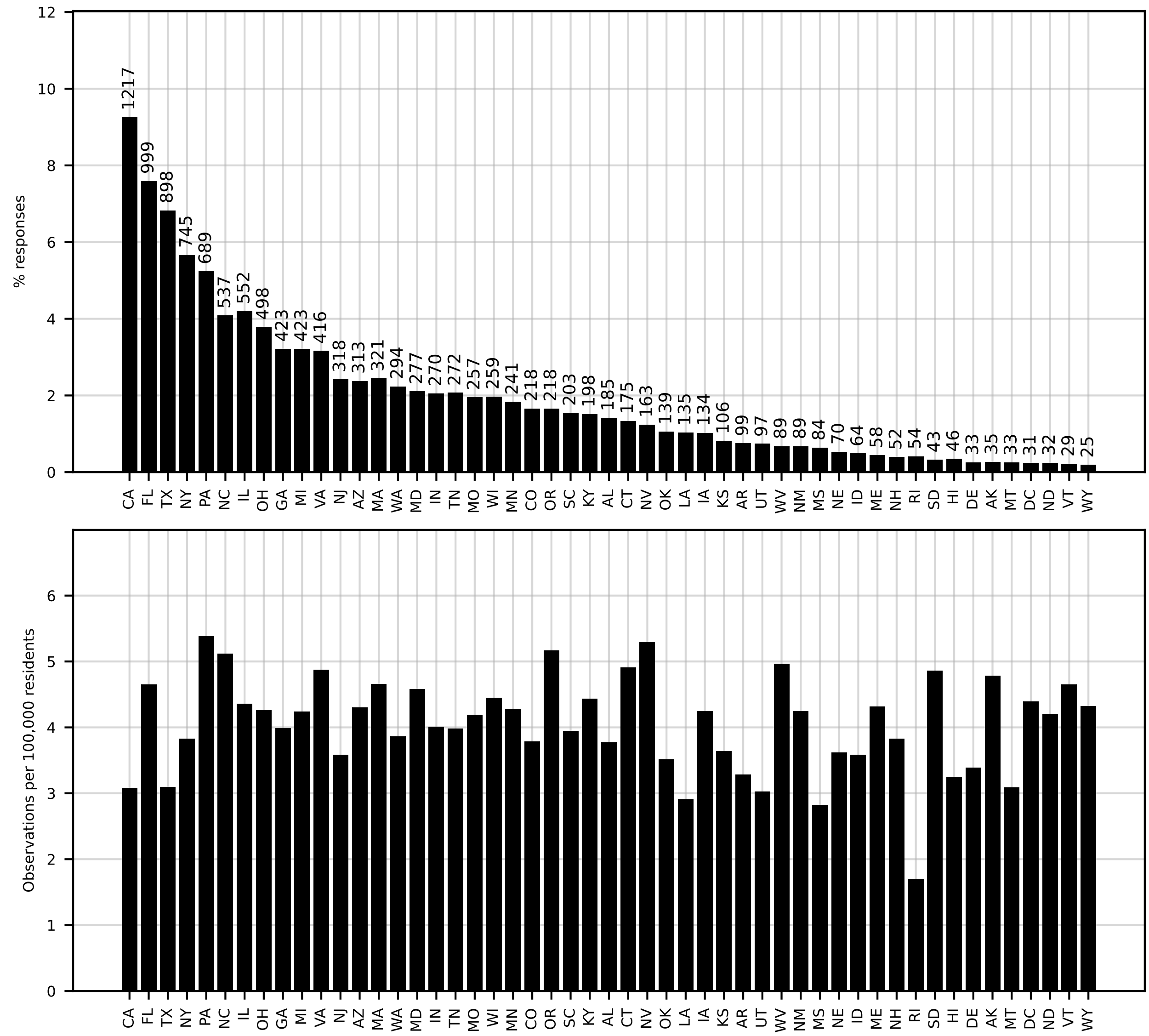


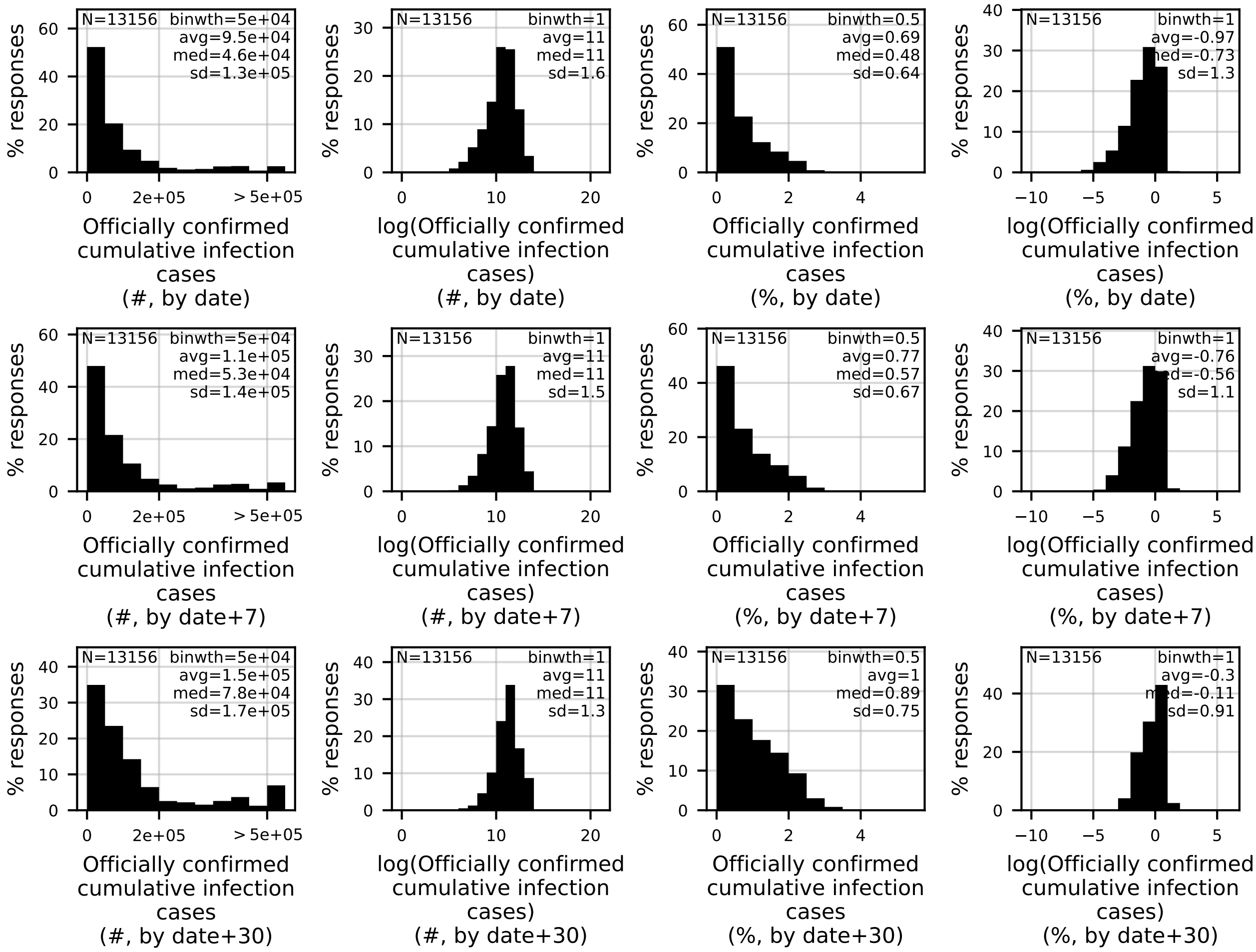
Figure B.2: US-state of residence distribution in the main sample



Notes: Numbers above bars: absolute numbers of responses from each state.

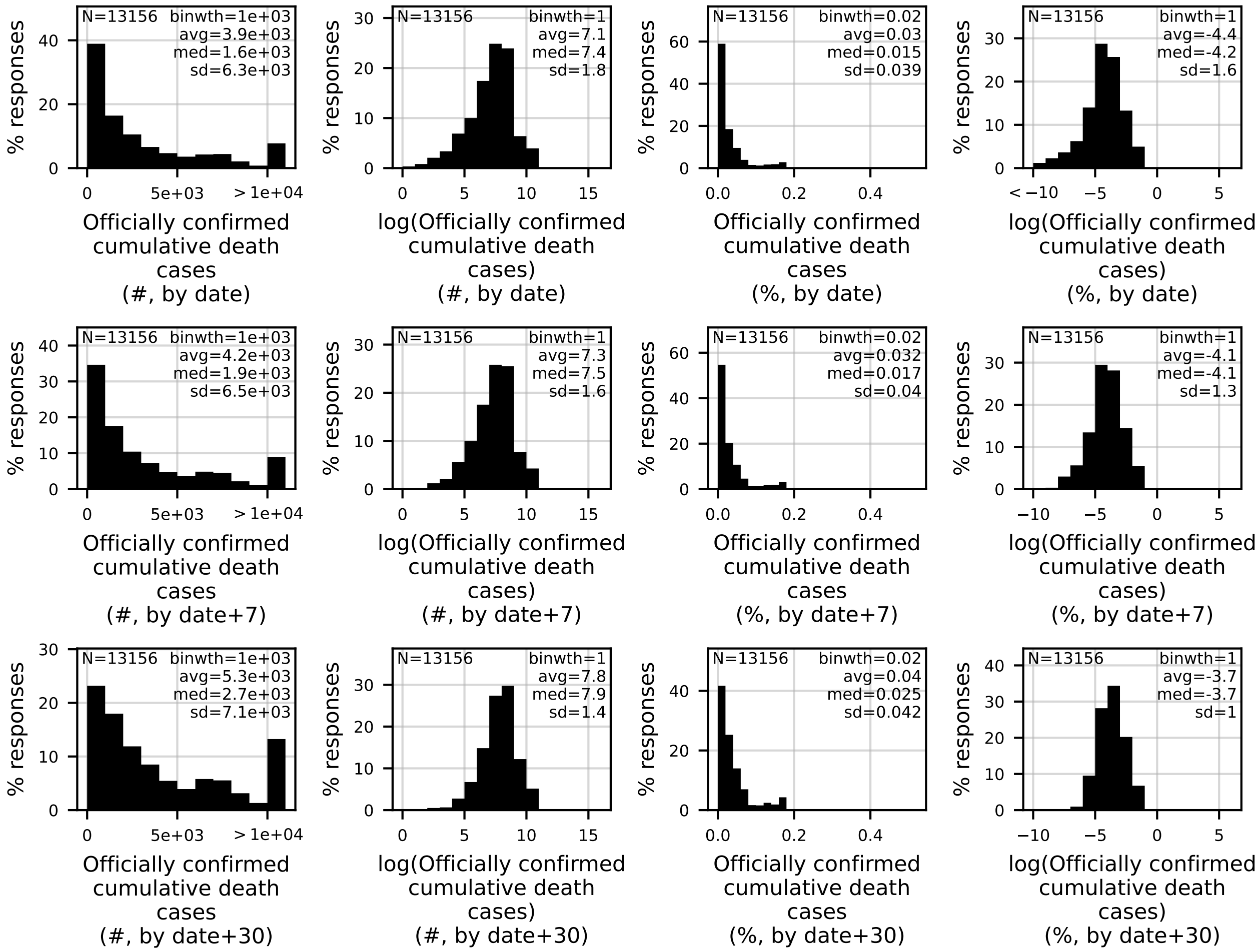
Figures B.3 and B.4 show the distribution of official infection and death counts matched to our survey responses, representing the states and dates sampled in our survey. Figures B.5, B.6, B.7, B.8 show the response distributions to all survey questions. We show these distributions with and without the logarithmic transformation, to emphasize its importance in jointly analyzing values with different orders of magnitude.

Figure B.3: Distribution of officially confirmed cumulative infection cases (main sample)



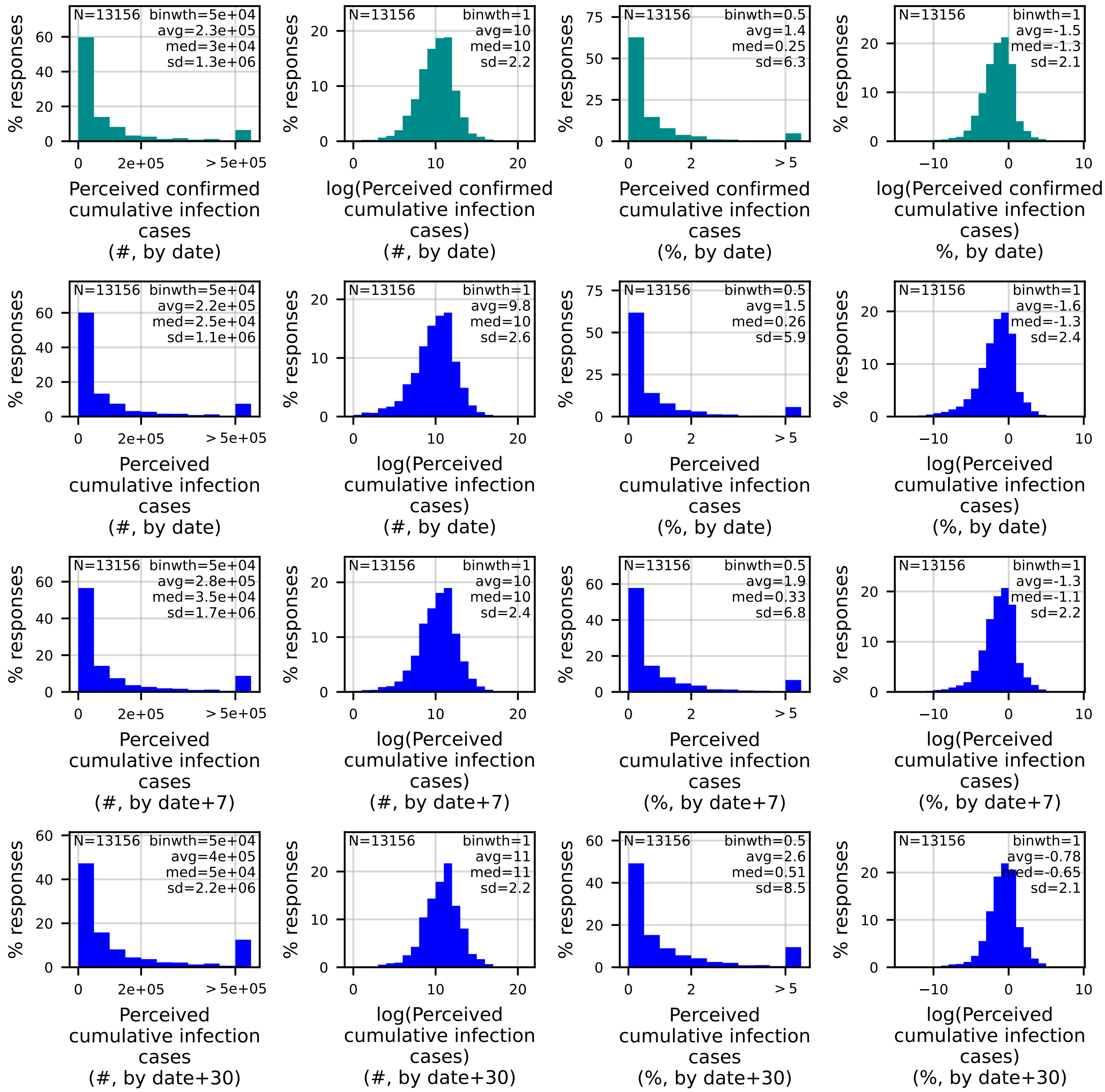
Notes: Both the absolute number of infections (#) and the percent it consists of the state's population (%) are shown. Each distribution is shown without and with a logarithmic transformation, which was applied on data prior to analysis.

Figure B.4: Distribution of officially confirmed cumulative death cases (main sample)



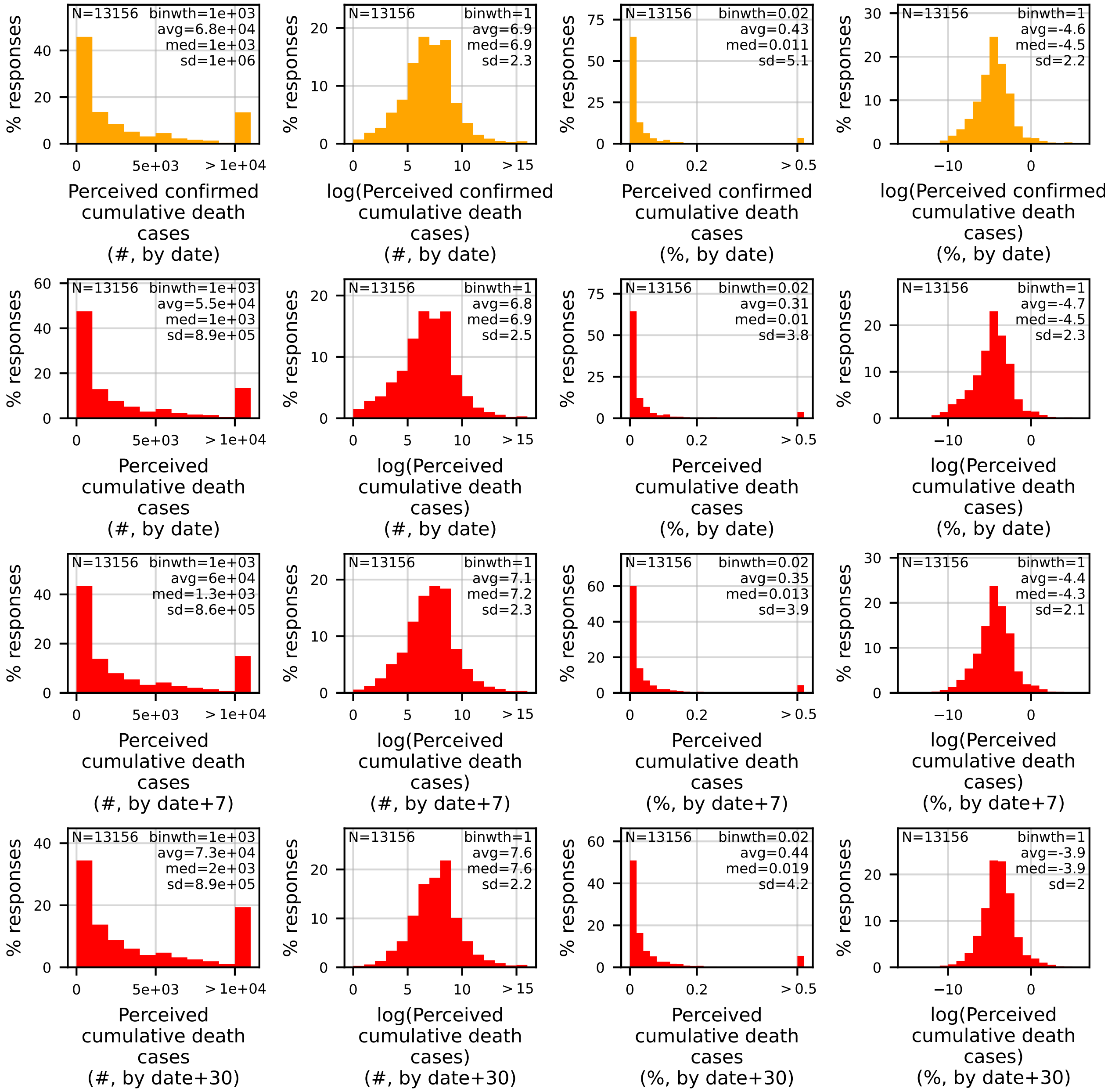
Notes: Both the absolute number of deaths (#) and the percent it consists of the state's population (%) are shown. Each distribution is shown without and with a logarithmic transformation, which was applied on data prior to analysis.

Figure B.5: Distribution of responses to infection case perception questions (modules A, B; main sample)



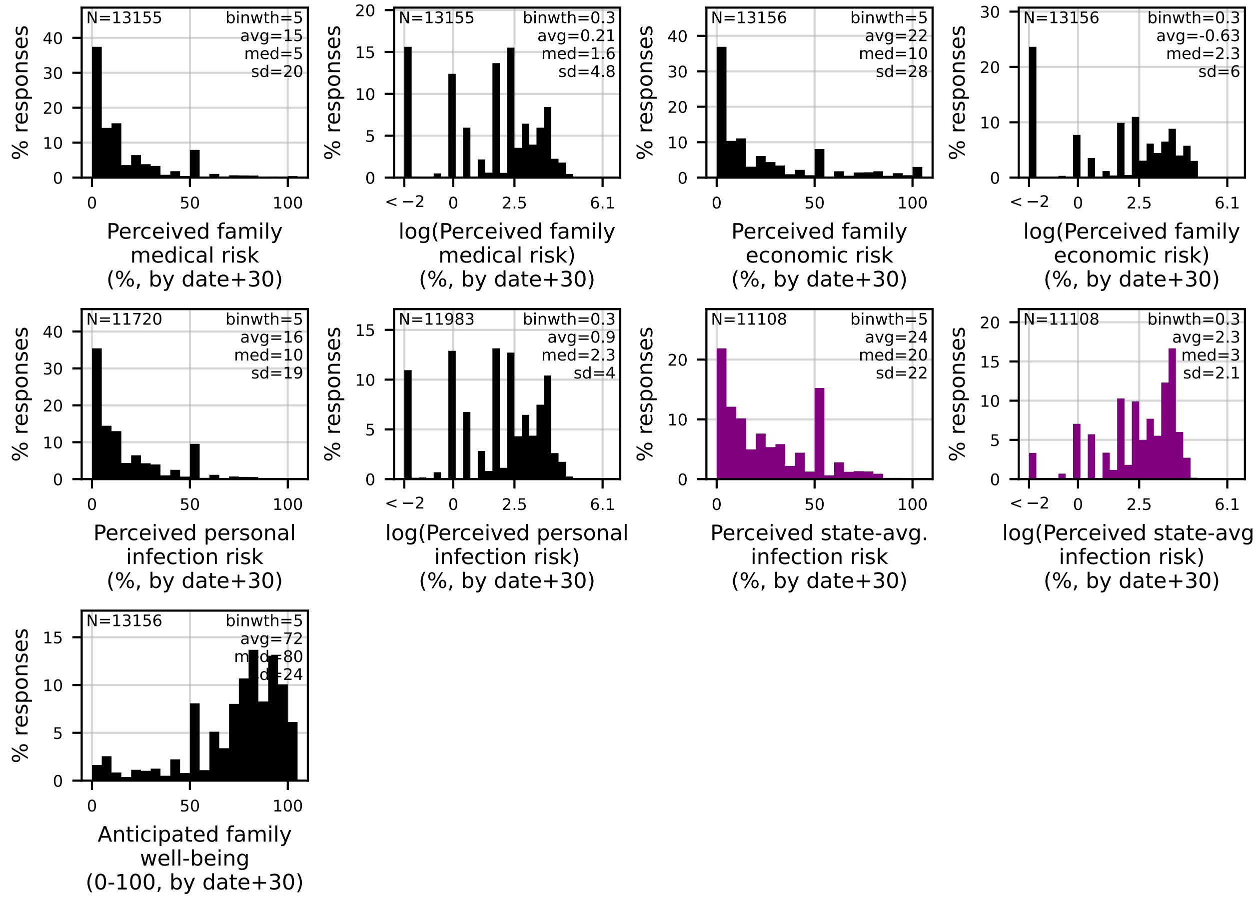
Notes: Each row presents one of four questions, and shows both the absolute number of infections (#) elicited and the percent it consists of the state's population (%). Each distribution is shown without and with a logarithmic transformation, which was applied on data prior to analysis.

Figure B.6: Distribution of responses to death case perception questions (modules A, C; main sample)



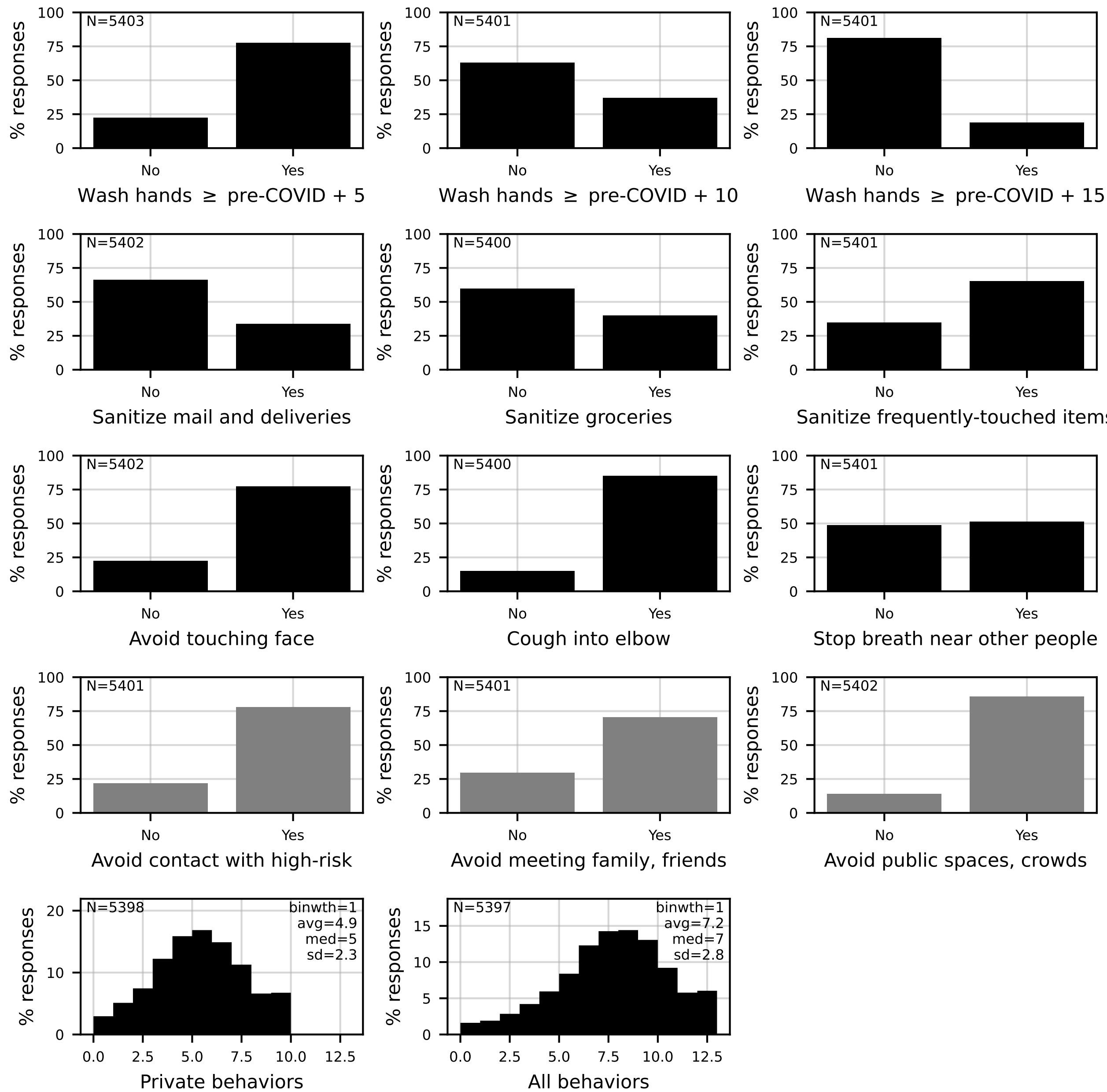
Notes: Each row presents one of four questions, and shows both the absolute number of infections (#) elicited and the percent it consists of the state's population (%). Each distribution is shown without and with a logarithmic transformation, which was applied on data prior to analysis.

Figure B.7: Distribution of responses to risk perception questions and to anticipated family well being (modules D, E; main sample)



Notes: Perceptions are reported as percent chances (%). Each distribution of perceptions is shown without and with a logarithmic transformation, which was applied on data prior to analysis.

Figure B.8: Distribution of responses to health-protective behavior questions (module F; main sample)



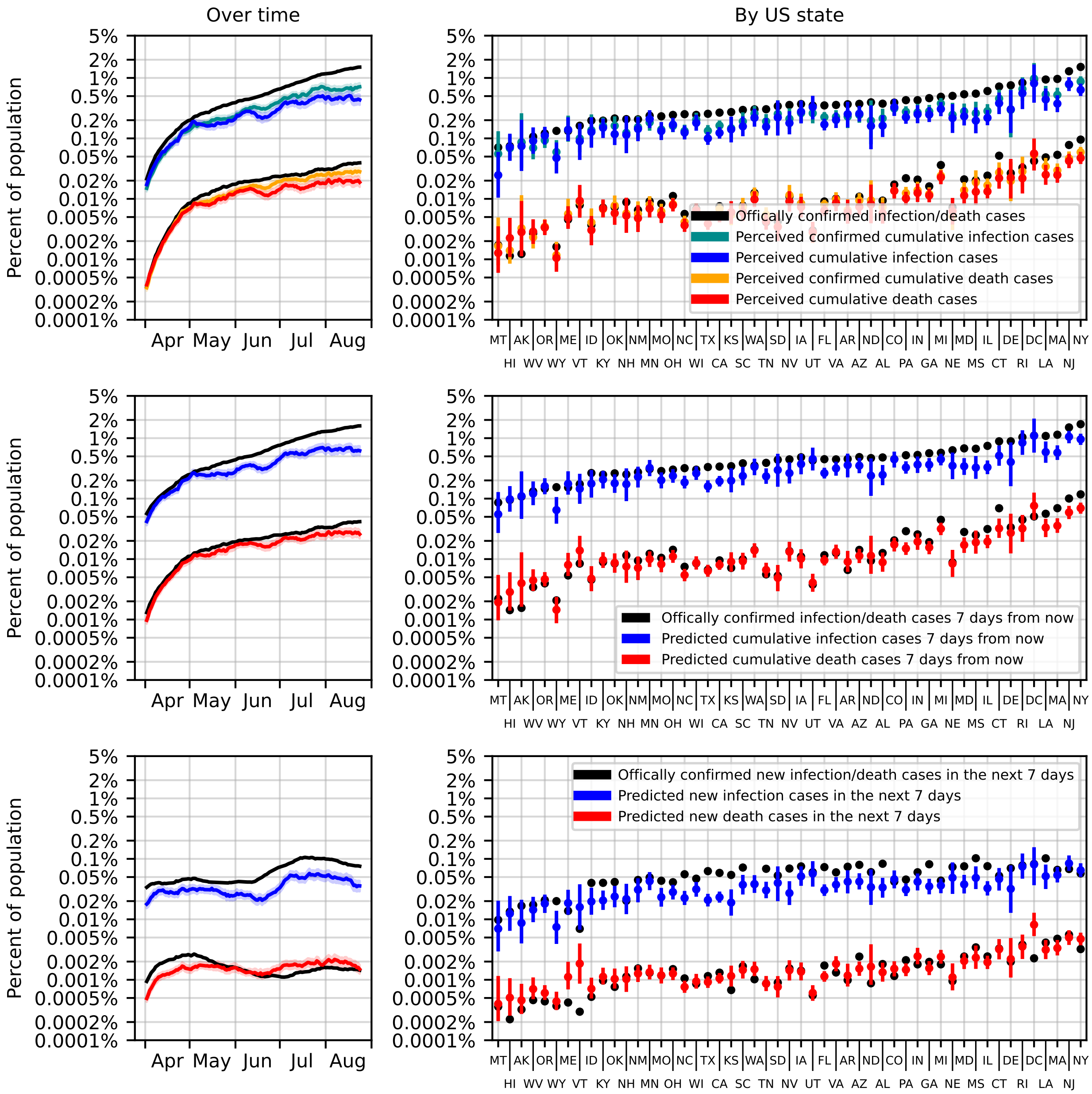
Notes: Upper four rows: 9 private behaviors (black) and 3 public behaviors (gray). Bottom row: distribution of the sum of Yes answers, both when considering only 9 private behaviors and when considering all 12 behaviors.

C Detailed Results

C.1 Case Perceptions vs. Official Reports

Figures C.1, C.2, C.3 show that the first main finding of a moderate under-estimation of case perceptions extends to all elicited perceptions about infection and death cases. The under-estimation of death cases is smaller than that of infection cases. The sample used for this analysis excludes 653 further responses from the main sample, in which negative growth of either infection or death cases in the next 7 days or the next 30 days was predicted. It also excludes further 40 responses for which growth in *official* death counts is negative due to ex-post classification of deaths as unrelated to COVID-19 (Section 1.2 explains why we did not apply such corrections backwards in time). This sample has $N = 12,463$.

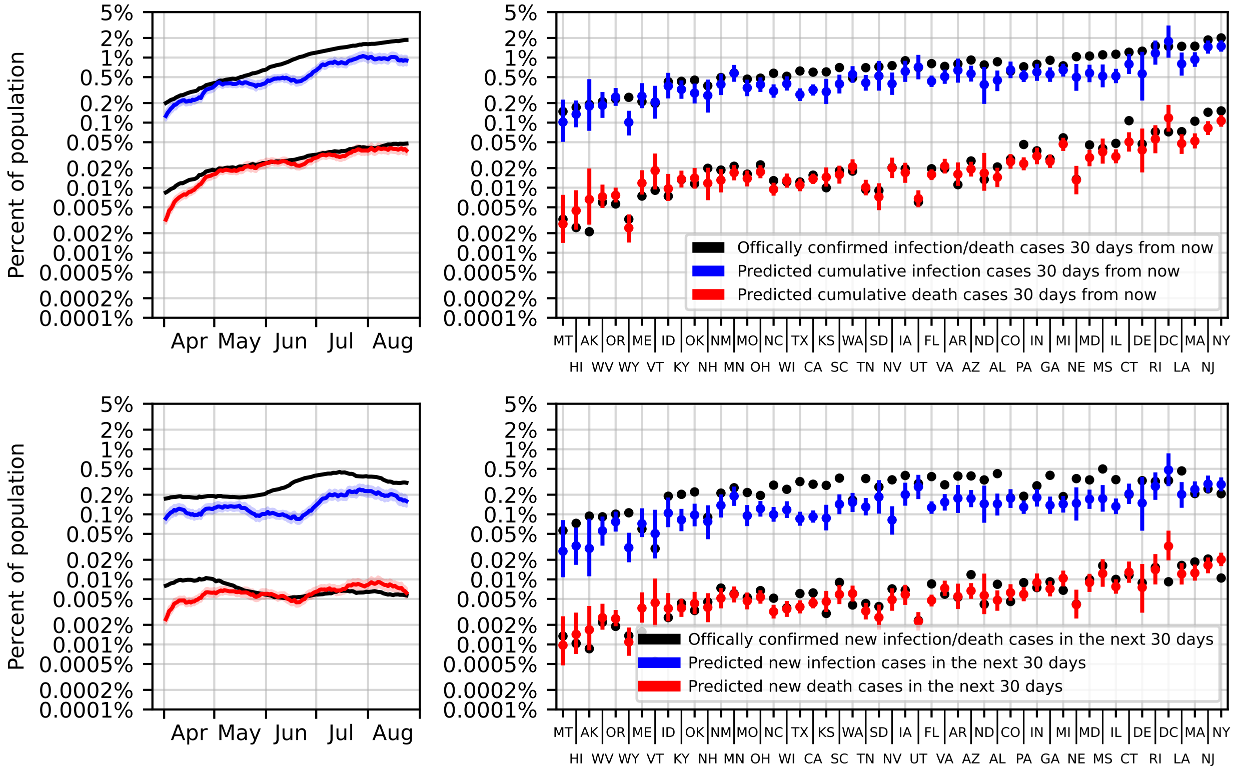
Figure C.1: Perceptions about infection and death cases vs. officially confirmed cases



Notes: Upper panel: current cumulative cases (confirmed and actual). Middle panel: cumulative cases as of 7 days from now. Lower panel: new cases in the next 7 days. The sample used for this analysis excludes observations with a predicted negative growth of either infection or death cases in the next 7 days or the next 30 days, and has $N = 12,463$.

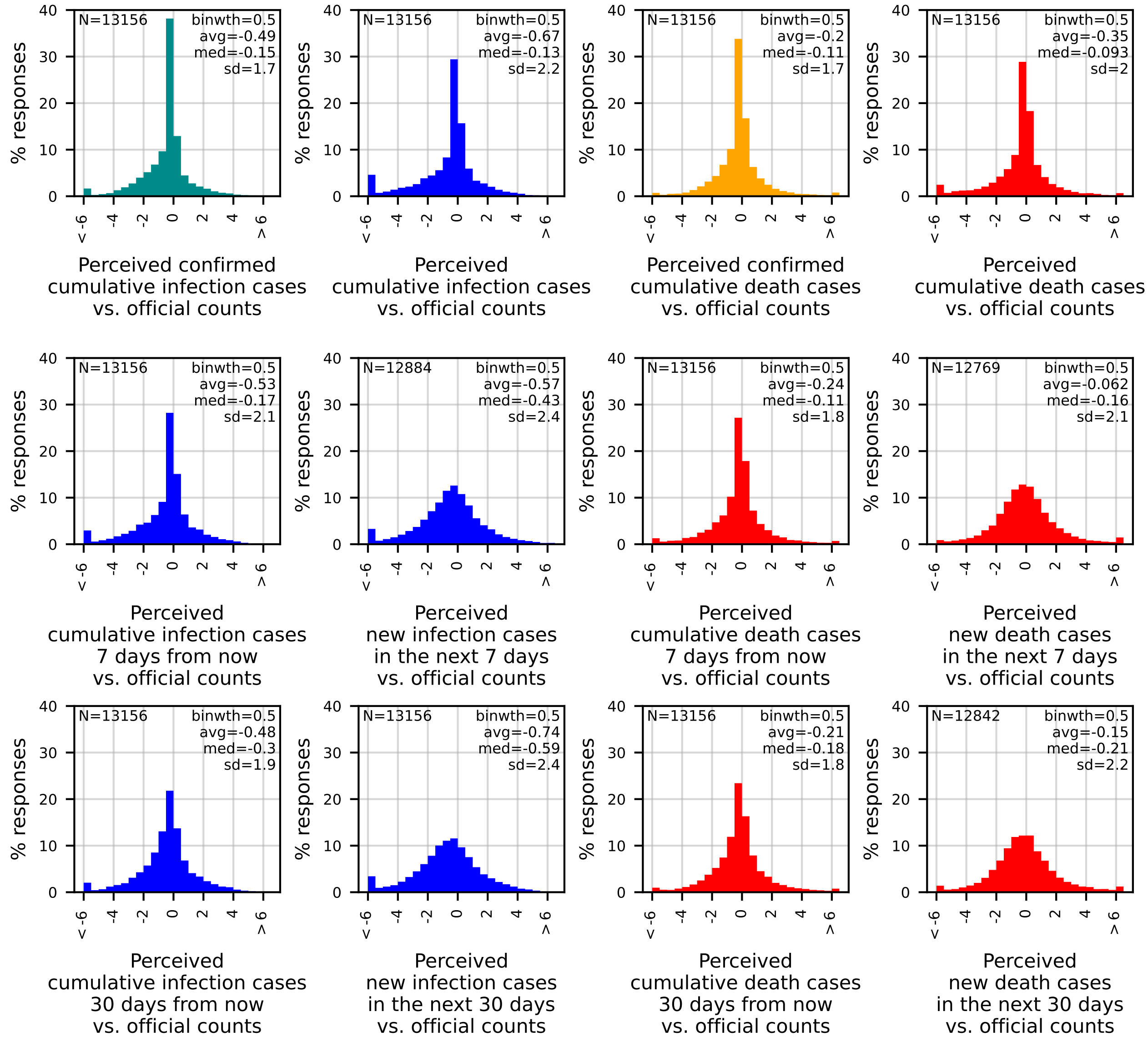
Light-colored areas in the left panels and error bars in the right panels: bootstrapped 95% confidence intervals.

Figure C.2: Perceptions about infection and death cases vs. officially confirmed cases



Notes: Upper panel: cumulative cases as of 30 days from now. Lower panel: new cases in the next 30 days. Same notes as under Figure C.1.

Figure C.3: Distribution of difference between case perceptions and officially confirmed cases

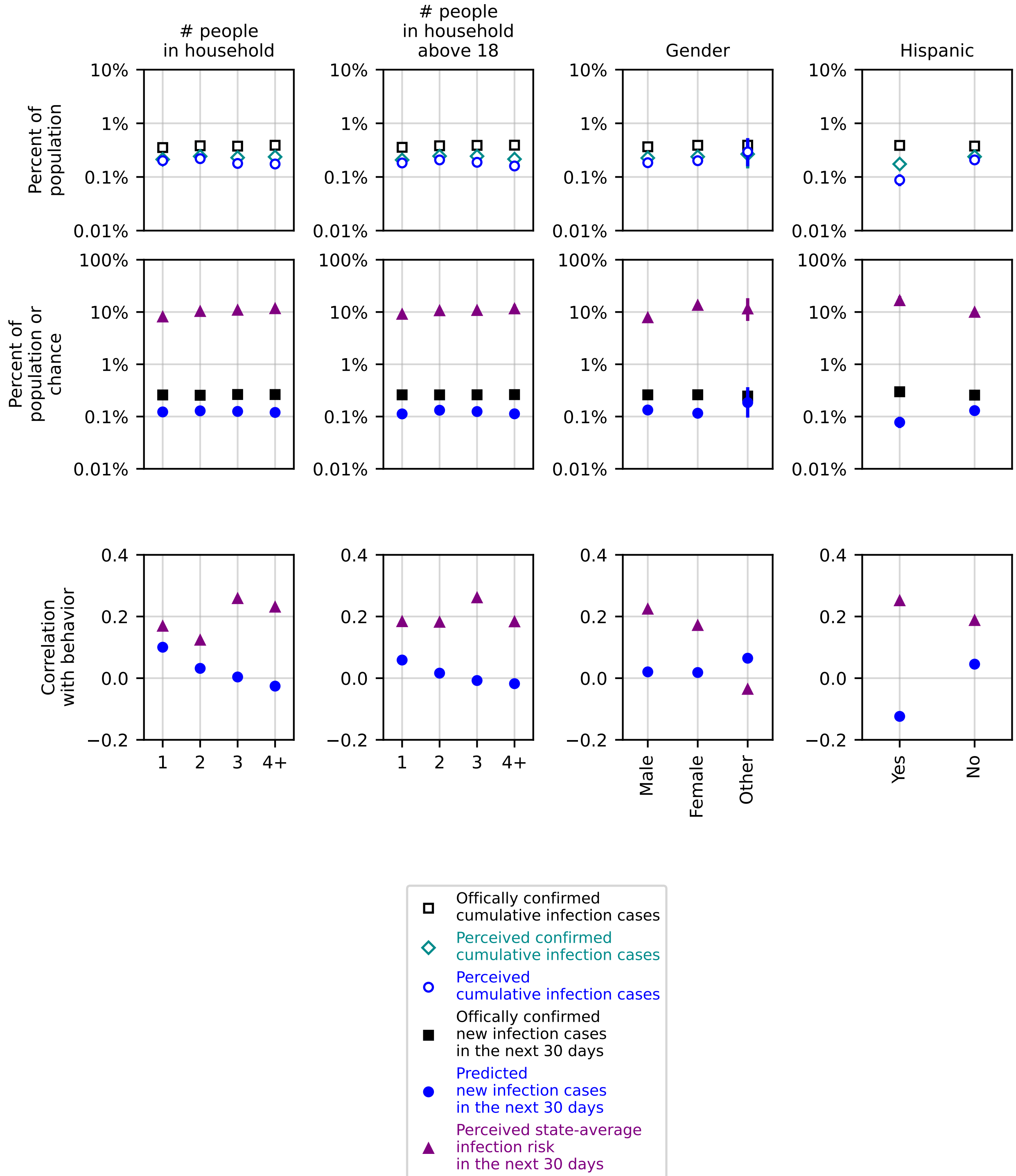


Notes: All quantities are differences between log percentages.

C.2 Main Findings as a Function of Demographics and MTurk Experience

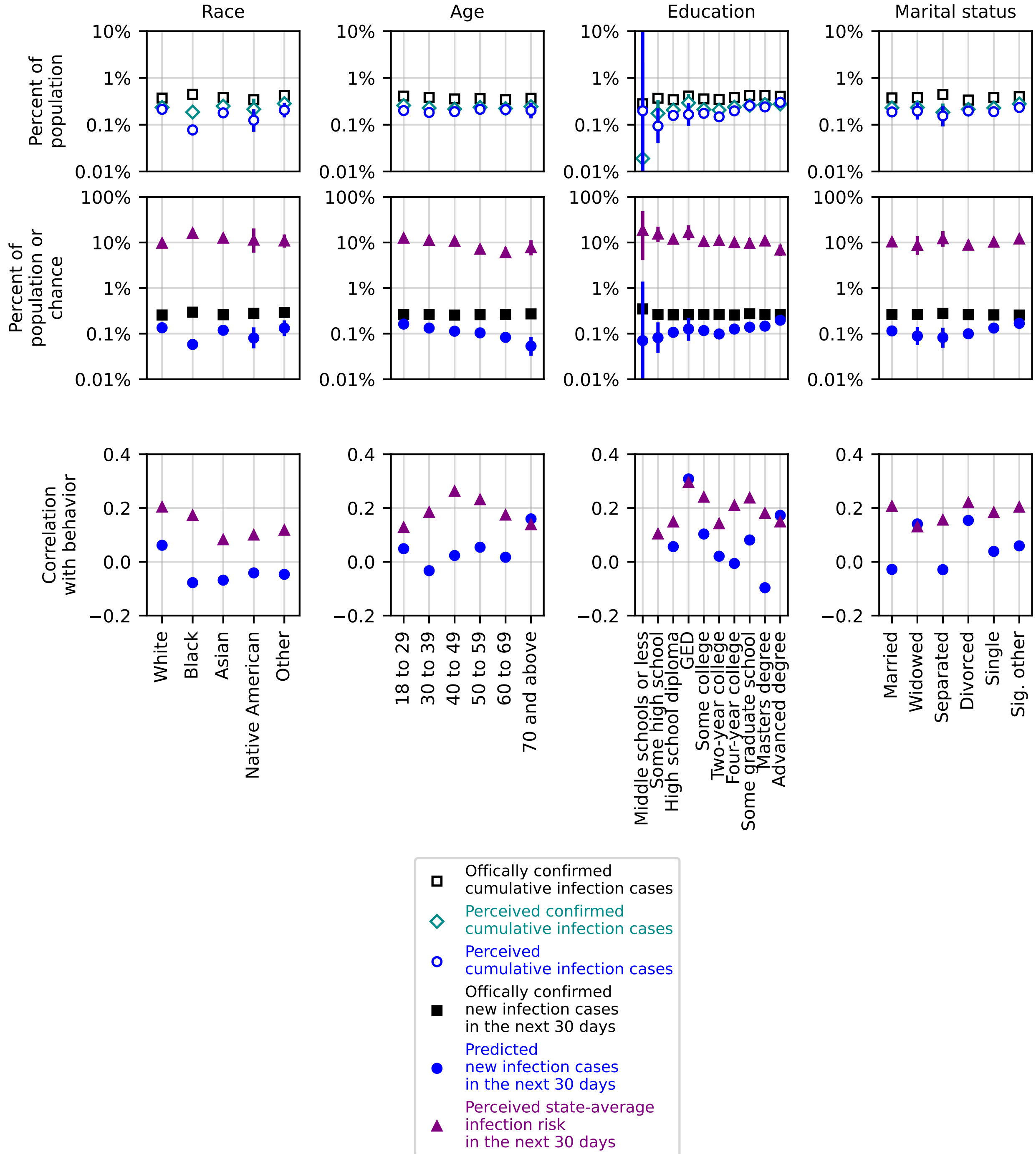
Figures C.4, C.5, C.6, C.7 show Figure 1's results as a function of demographic properties in the sample. Despite some demographic-dependent patterns, our three main findings generally seem to hold across demographic groups.

Figure C.4: Main results from Figure 1 within demographic groups



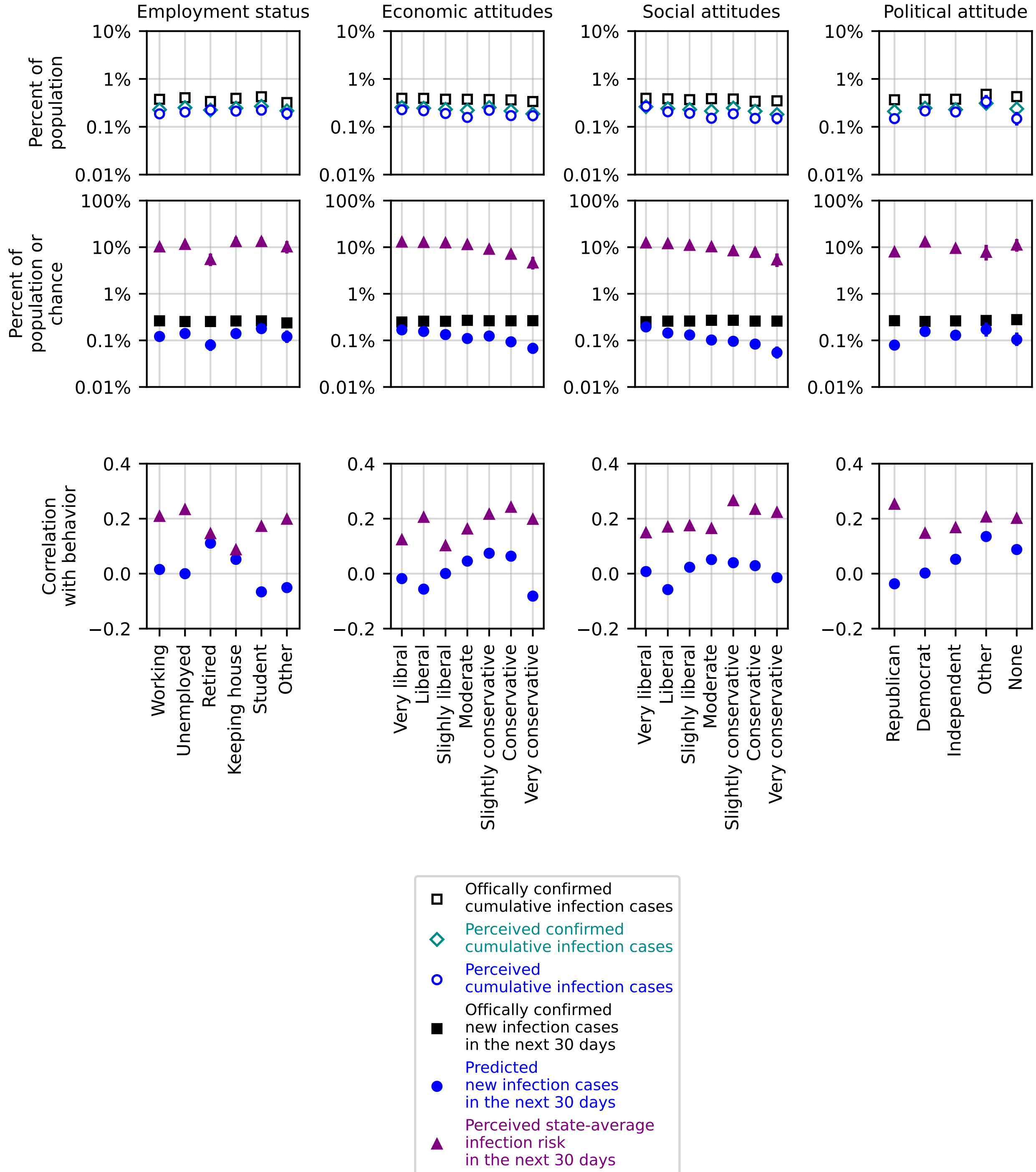
Notes: Top two rows: mean log percent perceptions and official case counts within demographic groups. Error bars (hardly visible): bootstrapped 95% confidence intervals. Lower row: Blue squares: correlation of case perceptions—predicted newly infected population percentage in the next 30 days—with self-reported health-protective behavior. Purple squares: correlation of risk perceptions—perceived state-average infection risk in the next 30 days—with behavior.

Figure C.5: Main results from Figure 1 within demographic groups



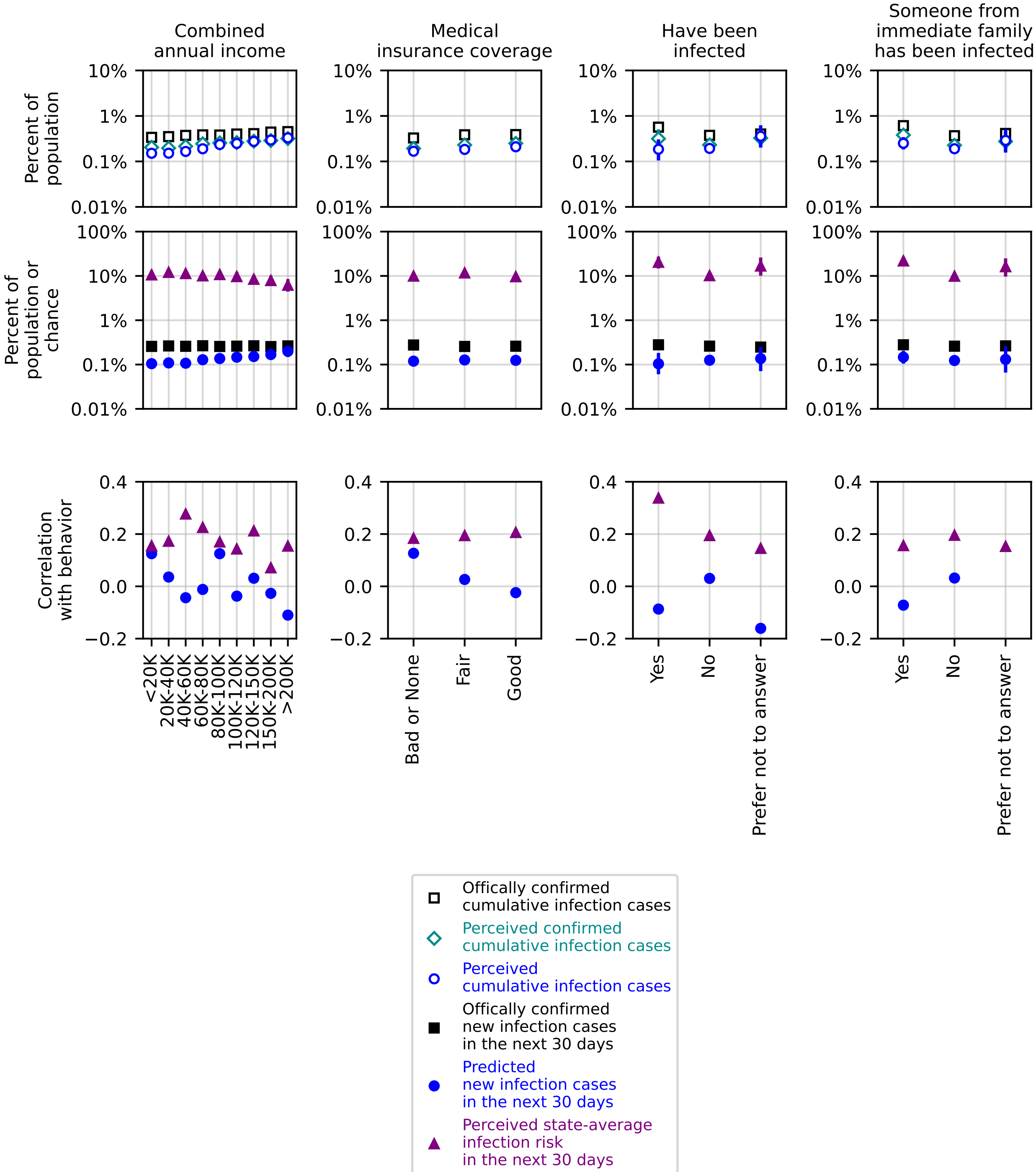
Notes: See under Figure C.4.

Figure C.6: Main results from Figure 1 within demographic groups



Notes: See under Figure C.4.

Figure C.7: Main results from Figure 1 within demographic groups



Notes: See under Figure C.4.

Table C.1 reports regressions of the Present, Future and Risk-Cases gaps' on demographic characteristics, state and day fixed effects. The regressions include a dummy for survey R1, which stands for inexperienced MTurk Workers. The inexperienced sample has a larger Risk-Cases gap than the main sample.

Table C.1: Main gaps as a function of demographic variables

	Present Cases gap	Future Cases gap	Risk-Cases gap
At least 3 people in household	−0.15	0.06	0.07
	(0.06)	(0.05)	(0.08)
At least 3 people above 18 in household	−0.02	−0.12	0.06
	(0.07)	(0.06)	(0.10)
Female	0.03	−0.15	0.65
	(0.04)	(0.04)	(0.06)
Hispanic	−0.64	−0.48	0.82
	(0.11)	(0.09)	(0.12)
Age at least 40	0.06	−0.28	0.05
	(0.04)	(0.04)	(0.05)
Not white	−0.47	−0.50	0.66
	(0.07)	(0.07)	(0.08)
Education less than 4-year college	−0.13	−0.11	0.23
	(0.04)	(0.04)	(0.05)
Not married	0.11	0.13	−0.25
	(0.06)	(0.06)	(0.07)
Not working	0.18	0.17	−0.19
	(0.04)	(0.04)	(0.07)
Non liberal economic attitudes	0.08	−0.01	−0.15
	(0.06)	(0.05)	(0.07)
Non liberal social attitudes	−0.22	−0.31	0.19
	(0.06)	(0.06)	(0.08)
Republican	−0.24	−0.40	0.10
	(0.08)	(0.09)	(0.10)
Not Democrat or Republican	0.01	−0.09	−0.07
	(0.05)	(0.05)	(0.06)
Combined annual less than 60K	−0.33	−0.26	0.42
	(0.06)	(0.05)	(0.06)
Fair or bad economic insurance	−0.04	0.02	0.05
	(0.04)	(0.04)	(0.05)
Have been infected	−0.07	−0.24	0.44
	(0.23)	(0.20)	(0.24)
Family member has been infected	0.07	0.25	0.32
	(0.10)	(0.13)	(0.15)
Inexperienced Worker (Survey R1)	−0.41	−0.15	0.74
	(0.03)	(0.07)	(0.17)
State FE	Yes	Yes	Yes
Day FE	Yes	Yes	Yes
Obs	13411	13411	11363
R^2	0.08	0.09	0.08

Notes: OLS regressions. Dependent variables: Present Cases gap; Future Cases gap; Risk-Cases gap (all as log percentages). Independent variables: demographic characteristics (binarized, see Appendix A.1); MTurk experience; state and day fixed effects. Sample size is larger than in the main analysis, since inexperienced MTurk Workers participated in survey version R1, whose sample is not included in the main analysis.
In parentheses: Driscoll-Kraay standard errors using Bartlett's kernel and a bandwidth of 4 days.

Table C.2 shows a regression of self-reported health-protective behaviors on Risk perceptions (perceived state-average infection risk in the next 30 days), Case perceptions (predicted newly infected population percentage in the next 30 days) and demographic characteristics, controlling for state and day fixed effects. As shown in Table 2, including demographic variables does not change the relations between perceptions and behavior. Some demographics that emerge as relatively important shifters of protective behavior include having someone from the immediate family infected (associated with increased protective behavior) and right-leaning political attitudes (decreased protective behavior, especially public behaviors).

Table C.2: Behavior as a function of demographic variables

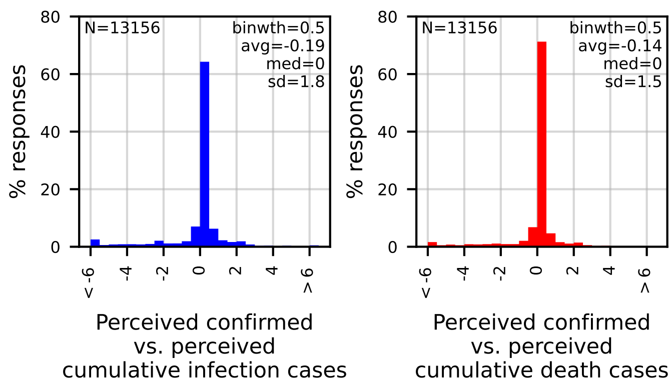
	Private behaviors	All behaviors
Risk perceptions	0.19	0.23
	(0.02)	(0.02)
Case perceptions	0.00	0.03
	(0.01)	(0.02)
At least 3 people in household	0.13	0.10
	(0.08)	(0.09)
At least 3 people above 18 in household	0.11	0.14
	(0.08)	(0.12)
Female	0.13	0.17
	(0.06)	(0.07)
Hispanic	0.35	0.37
	(0.13)	(0.16)
Age at least 40	−0.20	−0.17
	(0.05)	(0.06)
Not white	0.55	0.60
	(0.05)	(0.05)
Education less than 4-year college	−0.16	−0.22
	(0.05)	(0.06)
Not married	−0.23	−0.28
	(0.07)	(0.08)
Not working	−0.22	−0.23
	(0.06)	(0.07)
Non liberal economic attitudes	−0.27	−0.40
	(0.08)	(0.10)
Non liberal social attitudes	−0.02	−0.20
	(0.09)	(0.11)
Republican	−0.13	−0.36
	(0.08)	(0.12)
Not Democrat or Republican	−0.27	−0.33
	(0.10)	(0.12)
Combined annual less than 60K	0.12	0.12
	(0.06)	(0.07)
Fair or bad economic insurance	−0.20	−0.25
	(0.07)	(0.09)
Have been infected	−0.32	−0.53
	(0.23)	(0.26)
Family member has been infected	0.52	0.57
	(0.15)	(0.20)
State FE	Yes	Yes
Day FE	Yes	Yes
Obs	5398	5397
R^2	0.11	0.11

Notes: OLS regressions. Dependent variable: number of self-reported health-protective behaviors, out of nine private behaviors or all twelve private and public behaviors. Independent variables: Risk perceptions: log perceived state-average infection risk in the next 30 days; Case perceptions: log predicted newly infected population percentage in the next 30 days; demographic characteristics (binarized, see Appendix A.1); state and day fixed effects.
In parentheses: Driscoll-Kraay standard errors using Bartlett's kernel and a bandwidth of 4 days.

C.3 Perceived Cases vs. Perceived Confirmed Cases

Figure C.8 shows that the distribution of the difference between perceived (*actual*) cumulative infection cases (deaths) and perceived *confirmed* cumulative infection cases (deaths) is concentrated around zero. While the average difference in the left panel is -17 percent, which may suggest an overall perception of over detection/reporting of COVID-19 infection cases, the negative sign is not robust to random survey order. Responses who see the module order B–C and the prediction horizon order today–week–month (0–7–30) have a positive difference of 27 percent, while responses with the orders B–C $\times$ 30–7–0, C–B $\times$ 0–7–30 and C–B $\times$ 30–7–0 have average differences of -31, -6 and -41 percent respectively. See Figure D.3 in Appendix D.2.

Figure C.8: Distribution of difference between perceived confirmed cases and perceived actual cases



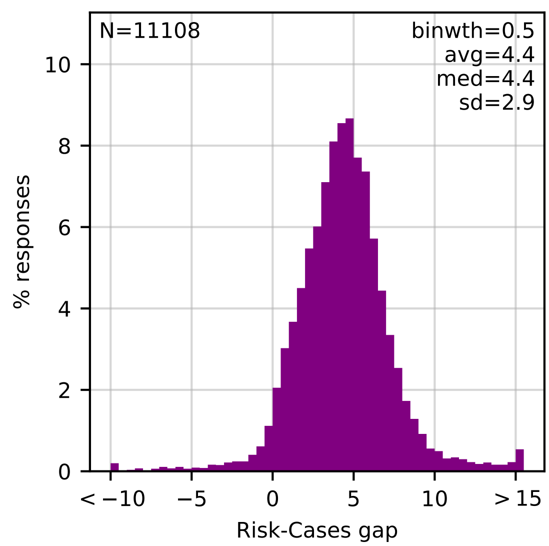
Notes: All quantities are differences between log percentages.

In addition, the average negative difference seems to be driven by outliers, since 5,517 responses perceive (actual) infection cases greater than confirmed cases, 4,901 perceive them as equal, and only 2,738 perceive the former smaller than the latter. Excluding the top and bottom 5 percentiles, the average difference increases to -5 percent, and excluding the top and bottom 10 percentiles, it further increases to 3 percent.

C.4 Distribution of the Risk-Cases Gap

Figure C.9 shows the distribution of the Risk-Cases gap, i.e., the log difference between perceived state-average infection risk in the next 30 days and predicted newly infected population percentage in the next 30 days. Both the mean and median difference indicate a gap of 4.4 log points, and 96 percent of the distribution is above the consistent-beliefs benchmark of zero.

Figure C.9: Distribution of the Risk-Cases gap

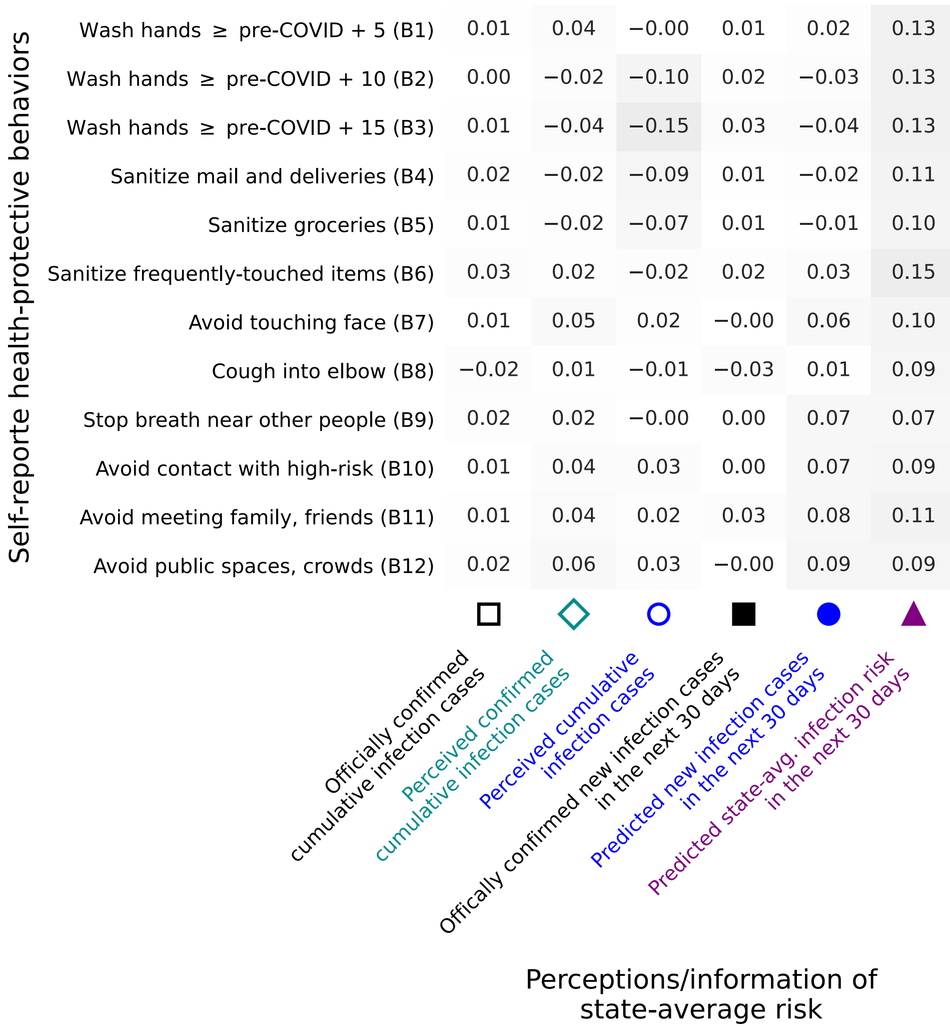


Notes: Risk-Cases gap: the log difference between perceived state-average infection risk in the next 30 days and predicted newly infected population percentage in the next 30 days. All quantities are log percentages.

C.5 Relations Between Single Behaviors and Perceptions

The third main finding relies on an aggregate behavior measure, summing all protective behaviors reported as adopted. Figure C.10 reports the correlations from Figure 1 with each behavior separately. Risk perceptions generally remain the strongest predictor of the first nine private behaviors, while case perceptions have a similar predictive power of the last three public behaviors.

Figure C.10: Correlations of perceptions and single self-reported health-protective behaviors



Notes: Correlations of officially confirmed cases, case perceptions and risk perceptions shown in Figure 1 with all twelve self-reported health-protective behaviors separately, listed on the vertical axis.

Tables C.3 and C.4 show regressions of each self-reported behavior separately on risk and case perceptions while controlling for demographics, state and day fixed effects. This reduces public behaviors' endogeneity with state regulations, which are affected by case counts. Risk perceptions are generally more strongly related to behavior than Case perceptions.

Table C.3: Relations of perceptions and single private self-reported health-protective behaviors

	B1	B2	B3	B4	B5	B6	B7	B8	B9
Risk perceptions	0.024	0.026	0.022	0.016	0.017	0.024	0.016	0.012	0.013
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.002)	(0.004)
Case perceptions	0.001	−0.005	−0.004	−0.001	−0.002	0.004	0.006	−0.002	0.010
	(0.002)	(0.002)	(0.001)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs	5403	5401	5401	5402	5400	5401	5402	5400	5401
R^2	0.074	0.093	0.107	0.110	0.095	0.098	0.069	0.073	0.071

Notes: OLS regressions. Dependent variables: single (not aggregated) self-reported health-protective behaviors (see full list in Figure C.10 above). Independent variables: Risk perceptions: (log) perceived state-average infection risk in the next 30 days; Case perceptions: (log) predicted newly infected population-percentage in the next 30 days; demographics; state and day fixed effects.
In parentheses: Driscoll-Kraay standard errors using Bartlett's kernel and a bandwidth of 4 days.

Table C.4: Relations of perceptions and single public self-reported health-protective behaviors

	B10	B11	B12
Risk perceptions	0.014	0.017	0.011
	(0.003)	(0.003)	(0.003)
Case perceptions	0.006	0.009	0.009
	(0.002)	(0.003)	(0.002)
Demographics	Yes	Yes	Yes
State fixed-effects	Yes	Yes	Yes
Day fixed-effects	Yes	Yes	Yes
Obs	5401	5401	5402
R^2	0.066	0.124	0.100

Notes: same as under Table C.4.

C.6 Within-Respondents Relations of Perceptions and Outcomes

Since we allowed all MTurk Workers to re-participate in the survey once more beginning on June 2, 2020, we have some panel data: the same 2,618 main-sample respondents participated once before the cutoff date and once after it. This large subsample enables testing some of the relations shown in Figure 4 within respondents. Tables C.5, C.6, C.7, C.8 show regressions of all outcome variables except behavior (which was elicited only after June 2, 2020) on risk perceptions and case perceptions, with and without individual fixed effects.

Our third main finding that risk perceptions are more strongly correlated with outcome variables than case perceptions holds at the within-individual level for medical risk outcomes, but only marginally for economic risk outcomes. We cannot support nor reject this finding for anticipated well-being.

Table C.5: Within-individuals relation between perceptions and outcomes
Dependent variable: Perceived personal infection risk

	(1)	(2)
Risk perceptions	0.97	0.72
	(0.02)	(0.08)
Case perceptions	0.16	0.07
	(0.02)	(0.04)
Constant	−0.95	−0.59
	(0.09)	(0.21)
Individual FE	No	Yes
Day FE	No	Yes
Obs	4205	4205
R^2	0.30	0.87

Notes: OLS regressions. Dependent variable: (log) perceived personal infection risk in the next 30 days. Independent variables: Risk perceptions: (log) perceived state-average infection risk in the next 30 days; Case perceptions: (log) predicted newly infected population-percentage in the next 30 days; individual and day fixed effects. Regressions use the subsample of respondents that completed the survey twice.
In parentheses: Driscoll-Kraay standard errors using Bartlett's kernel and a bandwidth of 4 days.

Table C.6: Within-individuals relation between perceptions and outcomes
Dependent variable: Perceived family medical risk

	(1)	(2)
Risk perceptions	0.67	0.34
	(0.03)	(0.08)
Case perceptions	0.25	0.03
	(0.04)	(0.04)
Constant	−0.73	−0.47
	(0.13)	(0.19)
Individual FE	No	Yes
Day FE	No	Yes
Obs	4262	4262
R^2	0.12	0.81

Notes: Same as under Table C.5, with (log) perceived family medical risk in the next 30 days as the dependent variable.

Table C.7: Within-individuals relation between perceptions and outcomes
Dependent variable: Perceived family economic risk

	(1)	(2)
Risk perceptions	0.49	0.16
	(0.05)	(0.07)
Case perceptions	0.21	−0.05
	(0.04)	(0.08)
Constant	−1.63	−1.45
	(0.19)	(0.27)
Individual FE	No	Yes
Day FE	No	Yes
Obs	4262	4262
R^2	0.04	0.82

Notes: Same as under Table C.5, with (log) perceived family economic risk in the next 30 days as the dependent variable.

Table C.8: Within-individuals relation between perceptions and outcomes
Dependent variable: Predicted family (minus) well-being

	(1)	(2)
Risk perceptions	0.97	0.53
	(0.20)	(0.39)
Case perceptions	0.49	0.19
	(0.19)	(0.36)
Constant	26.02	26.36
	(0.85)	(1.15)
Individual FE	No	Yes
Day FE	No	Yes
Obs	4262	4262
R^2	0.01	0.76

Notes: Same as under Table C.5, with minus predicted family well-being in the next 30 days as the dependent variable.

C.7 Explaining Perceptions' Variation

Table C.9 shows a regression of perceptions of cumulative infections at present ("Present case per."), case perceptions ("Case per."), and risk perceptions ("Risk per.") on officially confirmed cumulative present cases and future cases in the next 30 days, demographic variables and state and day fixed effects. While $R^2 = 0.22$ for cumulative infections at present, it is only 0.11 and 0.08 for case perceptions and risk perceptions, respectively. Using the multi-categorical (non-binarized) demographic variables has little effect on these values, increasing risk perceptions' R-squared to just 0.10.

Table C.9: Perceptions as a function of official information and demographic variables

	Present case per.	Case per.	Risk per.
Officially confirmed cumulative infection cases	0.54	0.69	0.22
	(0.07)	(0.07)	(0.08)
Future official cases	0.04	0.17	0.07
	(0.03)	(0.03)	(0.04)
At least 3 people in household	−0.13	0.08	0.16
	(0.06)	(0.05)	(0.05)
At least 3 people above 18 in household	−0.02	−0.14	−0.07
	(0.07)	(0.06)	(0.06)
Female	0.02	−0.16	0.52
	(0.04)	(0.04)	(0.04)
Hispanic	−0.66	−0.43	0.38
	(0.11)	(0.09)	(0.07)
Age at least 40	0.06	−0.29	−0.24
	(0.04)	(0.04)	(0.04)
Not white	−0.49	−0.49	0.22
	(0.06)	(0.07)	(0.05)
Education less than 4-year college	−0.14	−0.11	0.11
	(0.04)	(0.04)	(0.04)
Not married	0.09	0.11	−0.10
	(0.06)	(0.06)	(0.04)
Not working	0.18	0.16	−0.07
	(0.05)	(0.04)	(0.05)
Non liberal economic attitudes	0.10	0.02	−0.14
	(0.05)	(0.05)	(0.06)
Non liberal social attitudes	−0.24	−0.33	−0.10
	(0.05)	(0.06)	(0.06)
Republican	−0.23	−0.39	−0.31
	(0.08)	(0.08)	(0.06)
Not Democrat or Republican	0.00	−0.09	−0.19
	(0.06)	(0.05)	(0.04)
Combined annual less than 60K	−0.31	−0.26	0.15
	(0.07)	(0.05)	(0.05)
Fair or bad economic insurance	−0.02	0.03	0.08
	(0.03)	(0.03)	(0.04)
Have been infected	−0.12	−0.23	0.22
	(0.23)	(0.19)	(0.14)
Family member has been infected	0.11	0.21	0.59
	(0.10)	(0.13)	(0.08)
State FE	Yes	Yes	Yes
Day FE	Yes	Yes	Yes
Obs	13156	13156	11108
R^2	0.22	0.11	0.08

Notes: OLS regressions. Dependent variables: Present case perceptions: (log) perceived cumulative infected population percentage as of today; Case perceptions: (log) predicted newly infected population-percentage in the next 30 days; Risk perceptions: (log) perceived state-average infection risk in the next 30 days. Independent variables: Present official cases: (log) officially confirmed infected population percentage as of today; Future official cases: (log) officially confirmed newly infected population-percentage in the next 30 days; demographic characteristics (binarized, see Appendix A.1); state and day fixed effects.
In parentheses: Driscoll-Kraay standard errors using Bartlett's kernel and a bandwidth of 4 days.

Table C.10 shows the same regressions using the subsample of respondents who completed the survey twice. The R-squared values of future risk perceptions and risk perceptions significantly increase when adding individual fixed effects, suggesting that the bulk of variation in these beliefs can be explained by stable individual characteristics.

Table C.10: Perceptions as a function of official information and demographic variables

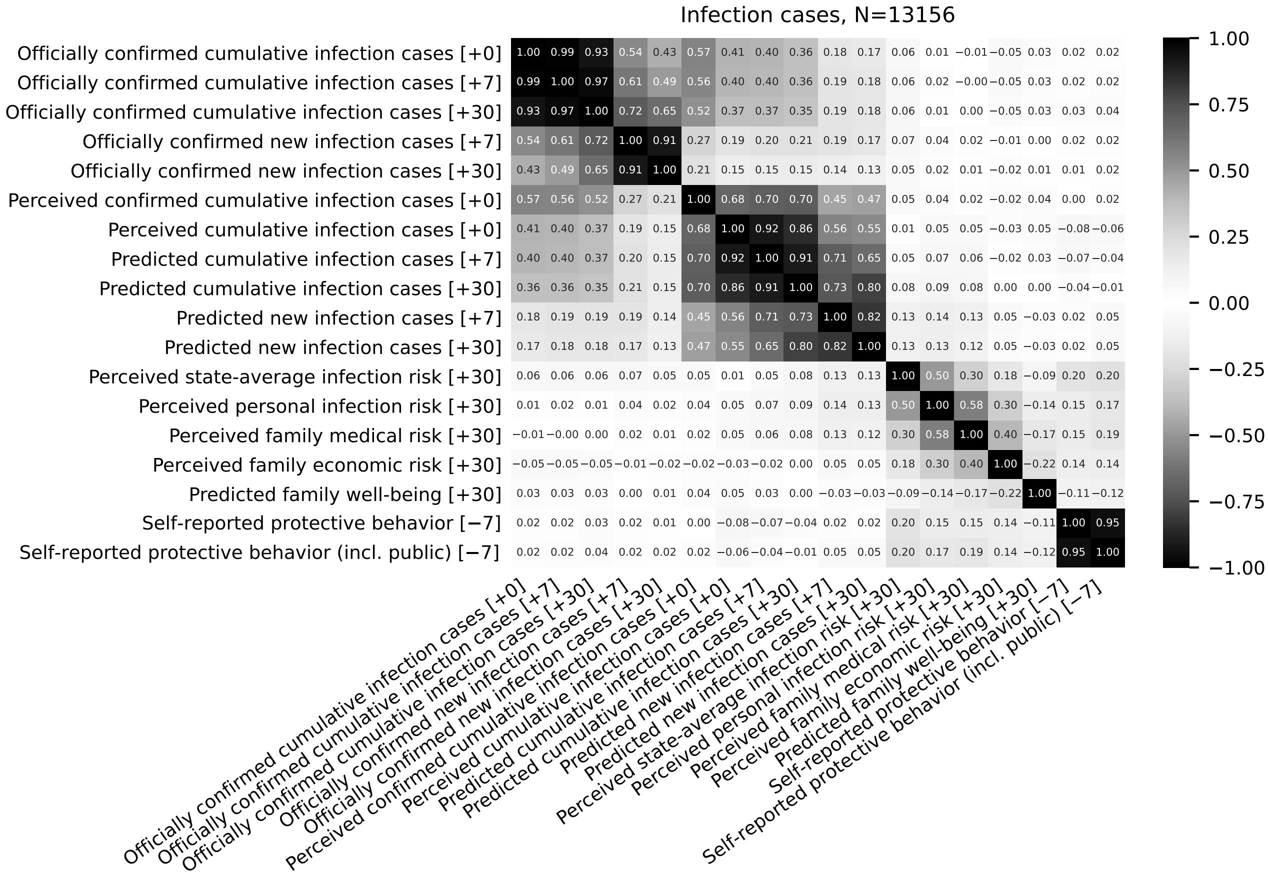
	Present case per.		Case per.		Risk per.	
Present official cases	0.68	0.61	0.74	0.59	0.12	0.13
	(0.09)	(0.06)	(0.10)	(0.09)	(0.13)	(0.16)
Future official cases	0.06	0.09	0.16	0.22	0.07	0.08
	(0.05)	(0.04)	(0.04)	(0.04)	(0.08)	(0.11)
Demographics	Yes	No	Yes	No	Yes	No
State FE	Yes	No	Yes	No	Yes	No
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	No	Yes	No	Yes	No	Yes
Obs	5236	5236	5236	5236	4262	4262
R^2	0.33	0.75	0.17	0.70	0.14	0.81

Notes: OLS regressions using the same variables as in Table C.9, also including individual fixed effects as an independent variable. Regressions use the subsample of respondents that completed the survey twice. In parentheses: Driscoll-Kraay standard errors using Bartlett's kernel and a bandwidth of 4 days.

C.8 Full Correlation Tables

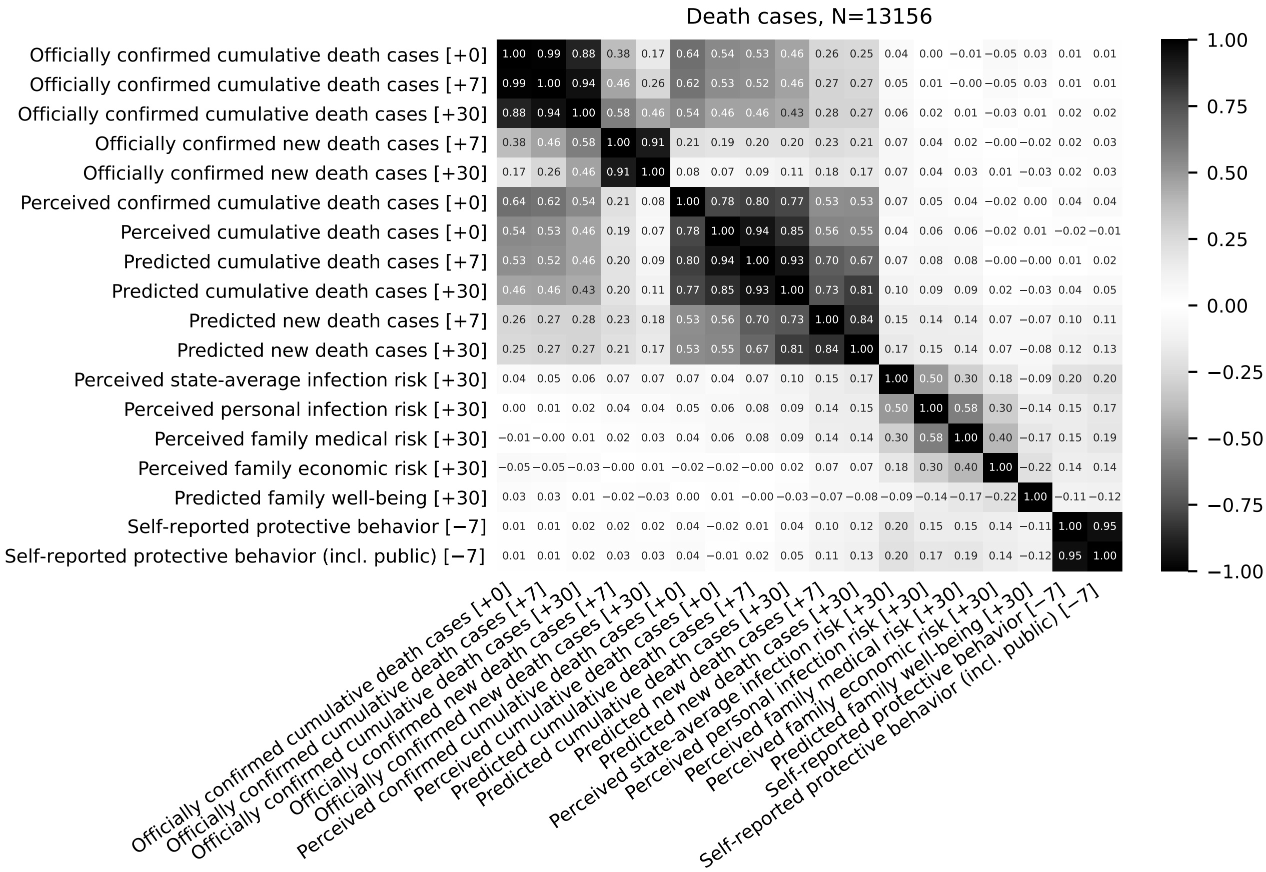
Figures C.11, C.12 explore all correlations between case perceptions, risk perceptions, anticipated well-being and (aggregated) protective behavior. Official reports and case perceptions are correlated with each other and within each; risk perceptions, anticipated well-being and protective behavior are only weakly correlated with them, while being correlated with each other. Perceptions about future deaths are more strongly correlated with behavior than perceptions about future infections.

Figure C.11: Full correlations table (main sample)



Notes: Official case counts, case perceptions and risk perceptions (all as log percentages).

Figure C.12: Full correlations table: case perceptions about deaths rather than infections (main sample)

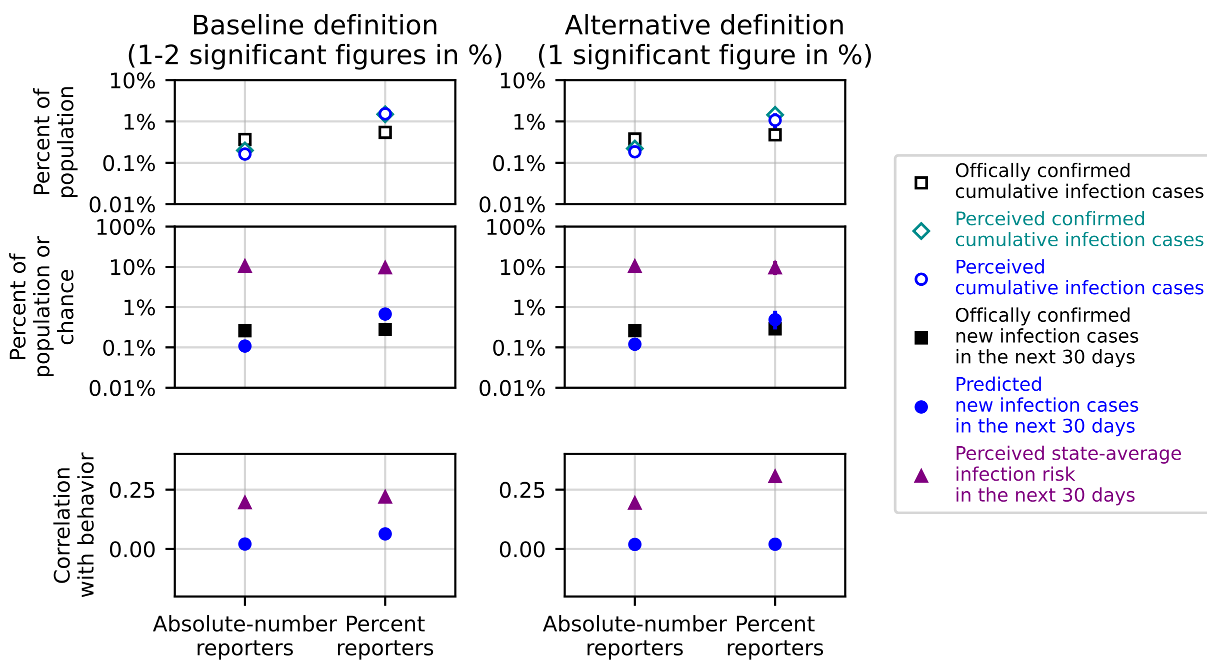


Notes: Official case counts, case perceptions and risk perceptions (all as log percentages).

C.9 The Role of Elicitation Details: More Results

Robustness to percent-reporters classification. We distinguish absolute-number reporters and percent reporters by the rounding pattern of percent responses to the three questions about infection cases (B1–B3). Respondents who report in all three questions percent values with at most two significant figures (e.g., 11%, 0.3%, 0.0052%) are classified as percent reporters, while the rest (e.g., 1.385%; respondents could see no more than three significant figures in the survey interface) are classified as absolute-number reporters. We test an alternative classification using just 1 significant figure. The tradeoff between classification criteria is that there is roughly 10 percent false identification rate when using 1–2 significant figures and only 1 percent rate when using one significant figure, but at the cost of a small subsample. Figure C.13 compares percent reporters and absolute-number reporters using both the baseline definition of 1–2 significant figures (resulting with 8 percent being percent reporters) and an alternative definition of only 1 significant figure (resulting with 3 percent being percent reporters).

Figure C.13: Main results within absolute-number reporters and percent reports: two classification definitions



Notes: Left column: classification of percent reporters as having percent responses with 1 or 2 significant figures (identical to the two left columns in Figure 5). Right column: only 1 significant figure. Bottom panel: correlation coefficients with self-reported protective behavior. Error bars: bootstrapped 95% confidence intervals; mostly smaller than the markers size.

The Risk-Cases gap among percent reporters is a factor of 14 and 22 using the baseline and alternative classification, respectively, and 92, 82 for absolute-number reporters. The Present Cases gap among percent-reporters is a 2.8-fold and a 2.3-fold overestimation of cases using the baseline and alternative classification, respectively. The lower panel shows that the correlations between case perceptions, risk perception and behavior maintain similar relative standings using both classifications.

Controlling for demographics. Self-selection to report percentages or absolute-numbers may be related to demographic characteristics, time and state, which then confound the effect of response format. Tables C.11 and C.12 show the controlled effect of being a percent reporter on the three gaps, which remains qualitatively similar to what shown in Figures 5, C.13: the Present Cases and Future Cases gaps (whose sample-averages are -0.67 and -0.74 log points, respectively) become positive and the Risk-Cases gap (whose sample-average is 4.37 log points) remains positive and large. Table C.13 shows the interacted effects of being percent reporter with risk perceptions vs. with case perceptions on self-reported health-protective behavior, which imply that risk perceptions' relation with behavior remains stronger than case perceptions' both among percent reporters and absolute-number reporters.

Table C.11: Main gaps as a function of being percent reporter (baseline definition)

	Present Cases gap	Future Cases gap	Risk-Cases gap
Percent reporter	2.05	1.72	−1.88
	(0.10)	(0.10)	(0.11)
Demographics	Yes	Yes	Yes
State FE	Yes	Yes	Yes
Day FE	Yes	Yes	Yes
Obs	13411	13411	11363
R^2	0.15	0.14	0.13

Notes: OLS regressions. Dependent variables: Present Cases gap; Future Cases gap; Risk-Cases gap (all as log percentages). Independent variables: dummy for percent reporter (baseline definition of 1-2 significant figures in percent reports); demographics; MTurk experience (survey version R1); state and day fixed effects. In parentheses: Driscoll-Kraay standard errors using Bartlett's kernel and a bandwidth of 4 days.

Table C.12: Main gaps as a function of being percent reporter (alternative definition)

	Present Cases gap	Future Cases gap	Risk-Cases gap
Percent reporter alt. def.	1.81	1.30	−1.41
	(0.21)	(0.23)	(0.23)
Demographics	Yes	Yes	Yes
State FE	Yes	Yes	Yes
Day FE	Yes	Yes	Yes
Obs	13411	13411	11363
R^2	0.11	0.11	0.10

Notes: Same as Table C.11, using percent reporter alternative definition of 1 significant figure to identify percent reports.

Table C.13: Relation of perceptions and behavior as a function of percent reporter

Dependent variable: Self-reported protective behavior

	9 private behaviors		12 behaviors	9 private behaviors		12 behaviors
Risk perceptions	0.22	0.17	0.21	0.22	0.17	0.21
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Case perceptions	0.00	0.02	0.05	−0.00	0.01	0.04
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Percent-reporter	−0.26	−0.34	−0.42			
	(0.10)	(0.09)	(0.12)			
Percent-rep. × Risk per.	0.00	−0.02	0.00			
	(0.06)	(0.05)	(0.06)			
Percent-rep. × Case per.	0.00	−0.00	−0.02			
	(0.03)	(0.03)	(0.04)			
Percent-reporter 2nd def.				−0.00	−0.12	−0.11
				(0.15)	(0.14)	(0.19)
Percent-rep. 2nd def. × Risk per.				0.09	0.04	0.09
				(0.08)	(0.08)	(0.09)
Percent-rep. 2nd def. × Case per.				−0.05	−0.04	−0.06
				(0.04)	(0.04)	(0.05)
Constant	4.36			4.34		
	(0.07)			(0.07)		
Demographics	No	Yes	Yes	No	Yes	Yes
State fixed effects	No	Yes	Yes	No	Yes	Yes
Day fixed effects	No	Yes	Yes	No	Yes	Yes
Obs	5398	5398	5397	5398	5398	5397
R^2	0.04	0.14	0.15	0.04	0.14	0.15

Notes: OLS regressions. Dependent variable: number of self-reported health-protective behaviors, out of nine private behaviors or all twelve private and public behaviors. Independent variables: Risk perceptions: (log) perceived state-average infection risk; Case perceptions: (log) predicted newly infected population-percentage in the next 30 days; Percent-reporter (2nd def.): whether a respondent responded to case perceptions questions using percentages, as defined in the baseline (alternative) classification.
The non-binary interacted variables (Risk perceptions and Case perceptions) are centered around their means to show the mean change of the effect of these variables due to the interaction. In parentheses: Driscoll-Kraay standard errors using Bartlett's kernel and a bandwidth of 4 days.

C.10 Effects of Difficulty-to-Reach and Previous Participation on Perceptions Gaps

Allowing MTurk Workers to re-participate in the survey beginning on June 2, 2020 allows us to identify the extent to which difficulty to reach of respondents, i.e., their (in)availability to take our survey on the platform, affects our main findings.[28] Such effects may confound time trends or cross-group comparisons (Heffetz and Rabin 2013). The MTurk pool of Workers

[28] Re-participation beginning on June 2, 2020 was primarily aimed to efficiently collect more responses. See Section 1.1.

is limited and changes gradually over time, and hence prior to June 2, 2020 easy-to-reach Workers were likely to have already completed our survey during March and April, leaving the more difficult-to-reach Workers to participate only later, on May and June.

Among all our respondents prior to June 2, 2020 (5,986 in the main sample), we proxy easy to reach ones as those who participated again in the survey after this reset date. We compare their perception gaps to those of difficult to reach respondents based on their *first responses only*, to avoid confounding results with specific experience in our survey. Table C.14 shows the results. None of the perception gaps are economically or statistically significantly affected by difficulty to reach.

Table C.14: Effect of difficulty-to-reach on the perception gaps

	PC gap	PC gap	FC gap	FC gap	RC gap	RC gap
Constant	−0.43		−0.53		4.54	
	(0.05)		(0.07)		(0.08)	
Participated twice	0.03	−0.02	−0.10	−0.09	0.03	0.12
	(0.04)	(0.04)	(0.06)	(0.06)	(0.09)	(0.10)
Demographics	No	Yes	No	Yes	No	Yes
State FE	No	Yes	No	Yes	No	Yes
Day FE	No	Yes	No	Yes	No	Yes
Obs	5989	5989	5989	5989	3941	3941
R^2	0.00	0.11	0.00	0.12	0.00	0.14

Notes: OLS regressions. Dependent variables: Present Cases (PC) gap; Future Cases (FC) gap; Risk-Cases (RC) gap (all as log percentages). Independent variables: a dummy for two survey completions by a respondent; demographics, state and day fixed effects. Sample is limited to responses recorded *prior* to June 2, 2020.
In parentheses: Driscoll-Kraay standard errors using Bartlett's kernel and a bandwidth of 4 days.

Similarly, we can compare respondents *after* June 2 by their first/second participation to measure the effect of previous participation on second-participation outcomes. Table C.15 shows that there are no such significant effects.

Table C.15: Effect of previous participation on the perception gaps

	PC gap	PC gap	FC gap	FC gap	RC gap	RC gap
Constant	-1.00		-0.86		4.34	
	(0.08)		(0.07)		(0.07)	
Participated twice	0.32	0.17	-0.05	0.01	-0.19	-0.12
	(0.07)	(0.07)	(0.05)	(0.05)	(0.07)	(0.06)
Demographics	No	Yes	No	Yes	No	Yes
State FE	No	Yes	No	Yes	No	Yes
Day FE	No	Yes	No	Yes	No	Yes
Obs	7167	7167	7167	7167	7167	7167
R^2	0.00	0.10	0.00	0.14	0.00	0.09

Notes: Same as under Table C.14. Sample is limited to responses recorded after June 2, 2020. In parentheses: Driscoll-Kraay standard errors using Bartlett's kernel and a bandwidth of 4 days.

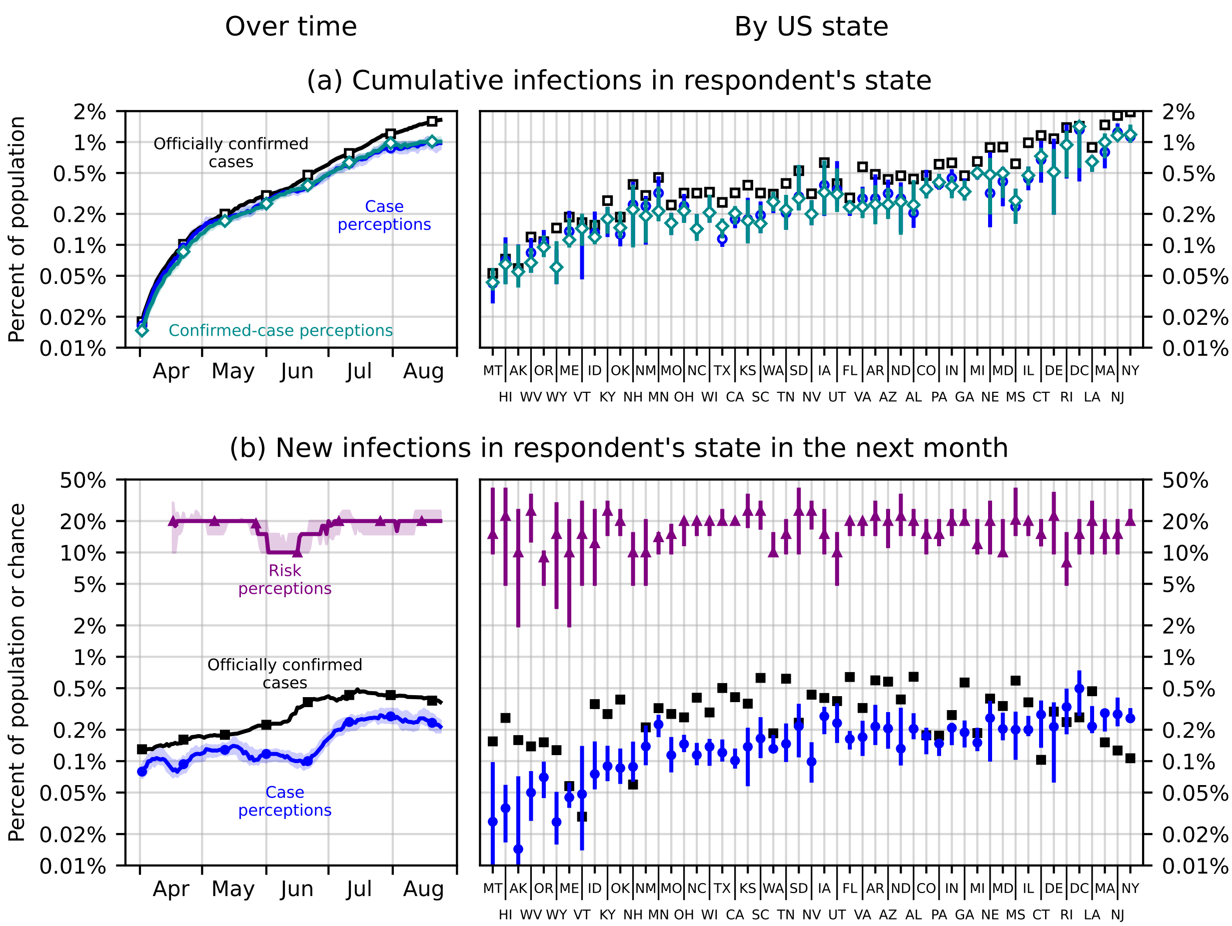
D Robustness Tests

D.1 Robustness to the Logarithmic Transformation and to the Analysis of Population Percentages Rather Than Absolute Numbers of People

All quantities x in the analysis are in the format of a population percentage or a percent chance, and the transformation $\log\left(\frac{1+x\cdot(\text{state pop.})}{1+\text{state pop.}}\cdot 100\right)$ is applied on them prior to analysis. The choice to use population percentages or percent chances is natural since this is a common elicitation format for all perceptions in the survey. The logarithmic transformation is useful since most quantities and percentages of infections and deaths vary be orders of magnitude across states and days, therefore investigating relative differences between quantities rather than absolute differences is preferable (relatedly, the response distributions in Appendix B are closer to symmetric on a logarithmic scale than on a linear scale). The analyzed averages throughout the paper are hence geometric rather than arithmetic.

A first concern is that the logarithmic transformation may mechanically generate an artificial negative difference between the average of a noisy variable (such as perceptions) and the average of a less noisy variable (such as official reports), in case the noise is symmetric on a linear scale rather than a log scale. Figure D.1 replicates Figure 1, aggregating the variables at the time or state level using medians, which are invariant to the log transformation, rather than averages. The negative differences between case perceptions and official reports indeed shrink, but maintain the sign. Our first and second main findings—reflected in the Present Cases, Future Cases, and Risk-Cases gaps—remain qualitatively the same when using medians rather than averages of logs.

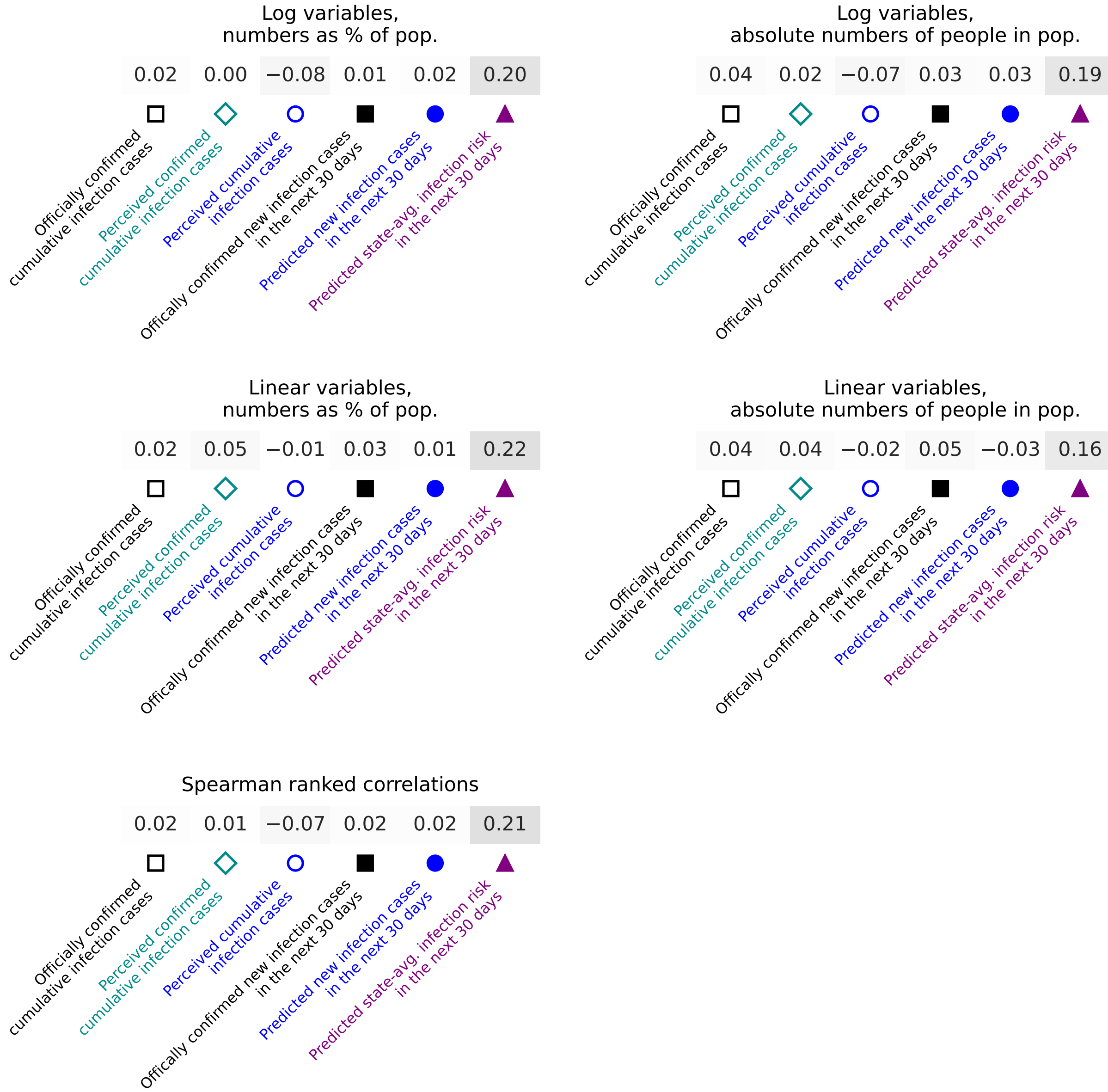
Figure D.1: Main results (as in Figure 1) using medians rather than averages



Notes: Light-colored areas in the left panels and error bars in the right panels indicate bootstrapped 95% confidence intervals.

A second concern is that the logarithmic transformation and/or the choice to analyze case and risk perceptions as population percentages rather than as absolute numbers drives correlations with behavior. Figure D.2 explores the correlations shown in Figure 1's lower bar, where variables are expressed as either state population percentages or absolute numbers of people, and are either log or linearly transformed (i.e., not transformed). A Spearman ranked correlation, which is invariant to any monotonic transformation, is also shown. Our third main finding that risk perceptions are more strongly correlated with behavior than case perceptions is robust to these analysis choices.

Figure D.2: Robustness of correlations with behavior to units and transformations of variables

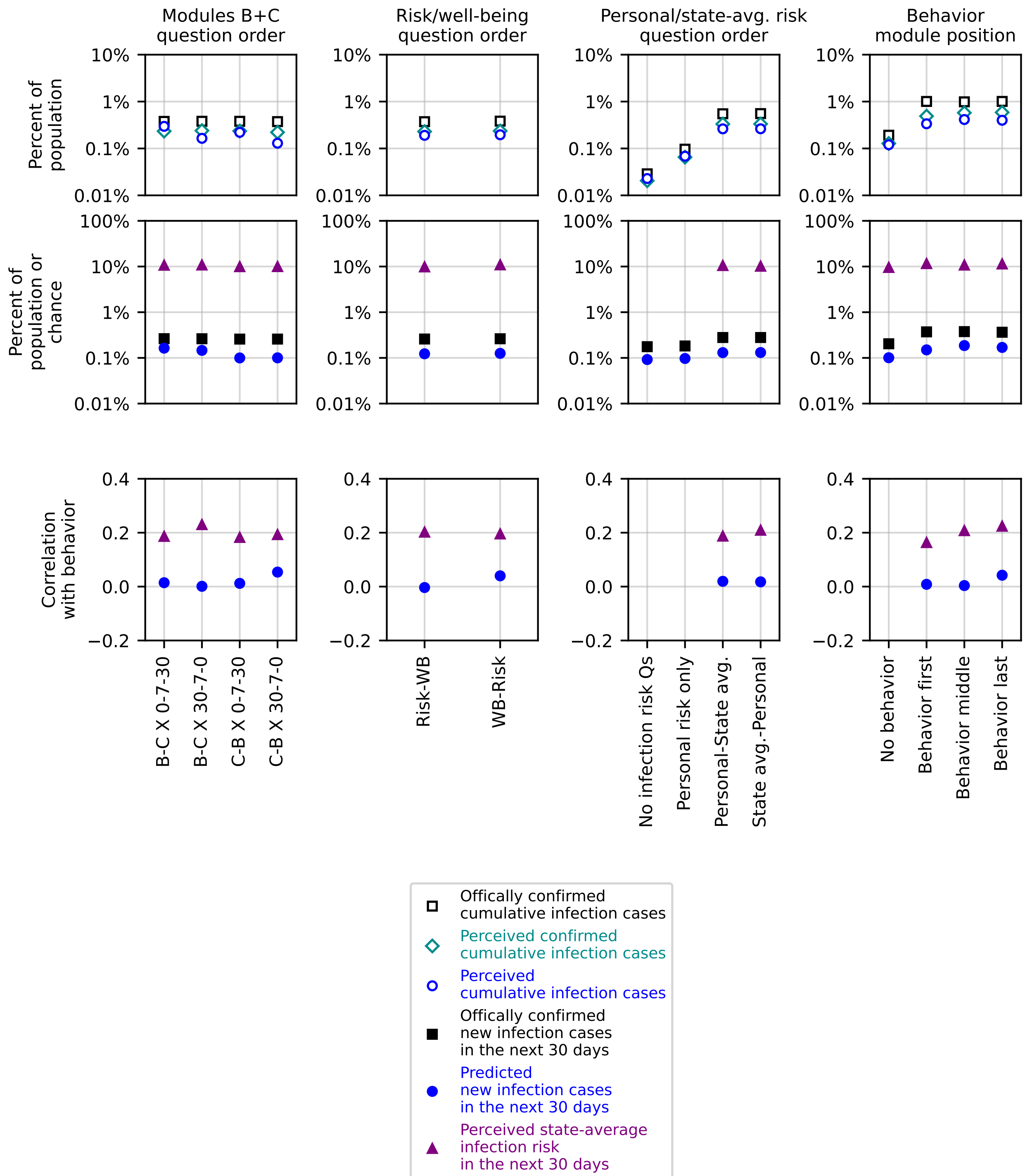


Notes: Correlations with the number of self-reported health-protective behaviors, out of nine private behaviors.

D.2 Robustness to Survey Order

Order effects on main results. Figure D.3 shows the main quantities from Figure 1 across the different survey orders. No single order setting qualitatively changes the main findings regarding Present Cases, Future Cases and Risk-Cases gaps, and the correlations with self-reported behavior. However, as discussed in Section 2.1 and Appendix C.3, the direction of the difference between perceived actual and confirmed cases *is* sensitive to survey order.

Figure D.3: Main results from Figure 1 within survey-order groups



Notes: Top two rows: mean log percent perceptions and official case counts within question-order groups. Error bars (hardly visible): bootstrapped 95% confidence intervals. Lower row: Blue squares: correlation of case perceptions—predicted new infections in the next 30 days—with self-reported health-protective behavior. Purple squares: correlation of risk perceptions—perceived state-average infection risk in the next 30 days—with behavior.

Table D.1 further investigates survey order effects on the main findings of Present Cases, Future Cases and Risk-Cases gap in a controlled regression. Order still has little effect on the main findings. No change of a single order renders the The Present Cases and Future Cases gaps orders-of-magnitudes large, although they may become closer to zero or even slightly positive in very specific combination of orders: for example, the Present Cases gap is slightly positive early in the sample before module E was added, within the survey order B–C $\times$ 0–7–30 (as indicated by the constant in the left column). The Risk-Cases gap remains orders-of-magnitudes large in all survey orders.

Table D.1: Survey order effects on the main gaps

	Present Cases gap	Future Cases gap	Risk-Cases gap
C–B	−0.26	−0.42	0.36
	(0.04)	(0.04)	(0.05)
30–7–0	−0.55	−0.05	0.05
	(0.05)	(0.03)	(0.05)
D3–D12	0.00	0.01	0.08
	(0.04)	(0.04)	(0.06)
E1 added	−0.12	0.02	
	(0.07)	(0.10)	
E2 added	−0.21	−0.17	
	(0.09)	(0.14)	
E2–E1	0.01	0.01	−0.04
	(0.04)	(0.05)	(0.06)
Dates of R2	0.02	0.45	−0.12
	(0.09)	(0.11)	(0.10)
Survey R2	0.09	−0.09	−0.13
	(0.13)	(0.06)	(0.06)
R2 + E2 after B3	0.03	0.15	−0.41
	(0.18)	(0.24)	(0.26)
F added	−0.37	0.02	−0.35
	(0.10)	(0.12)	(0.10)
F first	−0.18	−0.14	0.14
	(0.08)	(0.07)	(0.09)
F middle	0.05	0.07	−0.15
	(0.08)	(0.08)	(0.09)
Constant	0.18	−0.42	4.33
	(0.06)	(0.09)	(0.06)
Obs	13156	13156	11108
R^2	0.03	0.01	0.01

Notes: OLS regressions. Dependent variables: Present Cases gap; Future Cases; Risk-Cases gap (all as log percentages). Independent variables: possible orders of survey modules, including the special orders of version R2 (C–B: module C after B; 30–7–0: elicitation order of case perceptions: 30 days from today, then 7 days from today, then today, rather than the opposite order; D3-D12: well-being elicited before family risk perceptions; E1 added: dummy for all dates since question E1 about personal risk perceptions was added; E2 added: same for question E2 about state-average risk perceptions; E2–E1: state-average risk perceptions elicited before personal ones; Dates of R2: dummy for all dates in which survey R2 was conducted; Survey R2: being shown survey version R2 rather than the baseline version; R2 + E2 after B3: being shown version R2 and within this version have risk perceptions elicited right after case perceptions about 30 days from today; F added: dummy for all dates in which behavior was elicited; F first: behavior is the first module; F middle: behavior is in the middle, between modules B\C and D). In parentheses: Driscoll-Kraay standard errors using Bartlett's kernel and a bandwidth of 4 days.

Notable effects include an increased Risk-Cases gap (in magnitude) in the order C–B, even though case perceptions and and risk perceptions are asked closer to one another; a statistically-suggestive effect of asking E2 right after B3 (the two adjacent) in survey version R2, in the intuitive direction of decreasing the Risk-Cases gap, and a statistically-suggestive effect of having the beshavior module F first in the survey, which increases the absolute values of all gaps due to the lower case perceptions reported.

Effect of behavior module addition on all elicited quantities. Tables D.2, D.3, D.4 show the effect on elicited quantities of incorporating versions F-first and F-middle in the general analysis alongside the baseline version with the behavior module F last.

Table D.2: Effect of the position module of F on elicited quantities in modules A and B

	A1	A2	B1	B2	B3	B2−B1	B3−B1
F first	−0.19	−0.13	−0.18	−0.20	−0.17	−0.16	−0.13
	(0.06)	(0.07)	(0.08)	(0.07)	(0.06)	(0.07)	(0.07)
F middle	−0.02	−0.01	0.04	0.03	0.05	0.06	0.09
	(0.07)	(0.08)	(0.07)	(0.07)	(0.06)	(0.06)	(0.08)
Obs	5398	5398	5398	5398	5398	5273	5398
R^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Notes: OLS regressions. Dependent variables: all quantities elicited in modules A and B (as log percentages), including the new-cases predictions constructed based on them. Independent variables: survey versions F-first; F-middle. As in the main analysis, quantities are in the population percentage units and are log transformed. In parentheses: Driscoll-Kraay standard errors using Bartlett's kernel and a bandwidth of 4 days.

Table D.3: Effect of the position of module F on elicited quantities in module C

	C1	C2	C3	C2−C1	C3−C1
F first	−0.17	−0.17	−0.09	−0.03	0.00
	(0.09)	(0.08)	(0.07)	(0.06)	(0.07)
F middle	−0.00	−0.00	0.03	0.03	0.10
	(0.09)	(0.09)	(0.07)	(0.08)	(0.08)
Obs	5398	5398	5398	5229	5250
R^2	0.00	0.00	0.00	0.00	0.00

Notes: Same as under D.2, except that dependent variables are all quantities elicited in module C (as log percentages), including the new-cases predictions constructed based on them.

Table D.4: Effect of the position of module F on elicited quantities in modules D–F

	D1	D2	D3	E1	E2	F-private	F-all
F first	−0.00	0.02	−0.09	0.08	0.02	−0.18	−0.30
	(0.17)	(0.21)	(0.63)	(0.11)	(0.05)	(0.06)	(0.08)
F middle	−0.06	−0.16	2.03	−0.26	−0.06	−0.18	−0.24
	(0.22)	(0.24)	(0.53)	(0.12)	(0.06)	(0.06)	(0.08)
Obs	5398	5398	5398	5398	5398	5398	5397
R^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Notes: Same as under D.2, except that dependent variables are all quantities elicited in modules D, E and F (as log percentages).

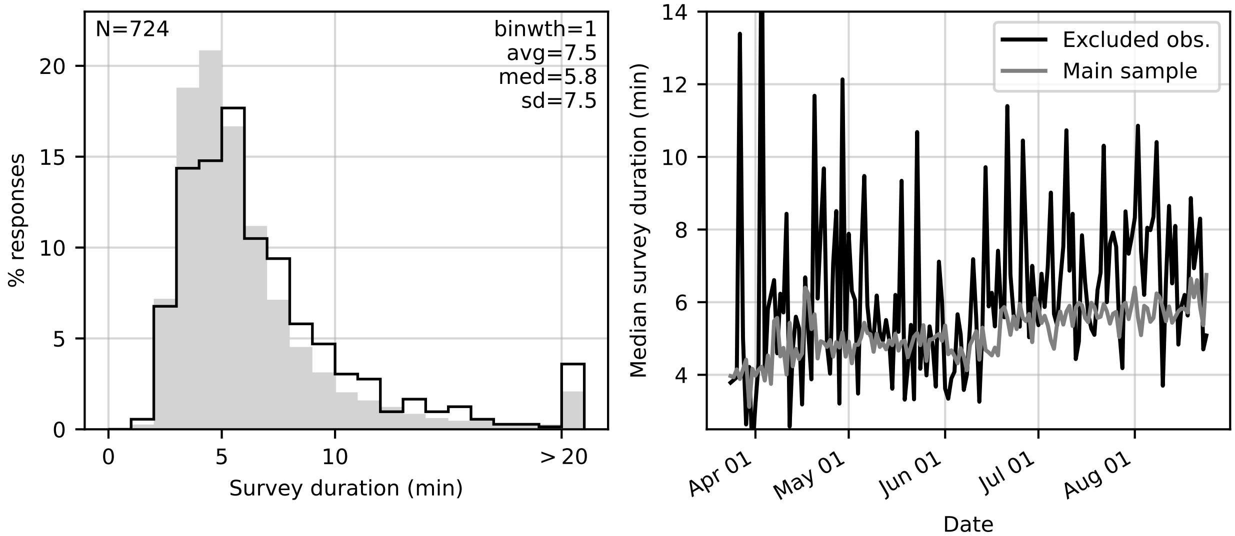
When F is the first module in the survey, respondents report somewhat lower perceptions of infection and death cases than when F is last. Having F in the middle does not have significant effect on elicited quantities in modules A–C. When F is in the middle respondents report somewhat lower economic worries and a higher predicted well being. When F is either first or in the middle respondents report less protective behavior than when F is last. None of these effects is substantial, hence data from these complementary versions is pooled with baseline version in the paper's analysis.

D.3 Robustness to Data Exclusion and Data Quality

724 observations, which are 5.2 percent of the full sample, are excluded from the main sample due to negative growth predictions of cumulative infections in 30 days. We investigate the characteristics of these excluded observations and test the extent to which the main findings hold in the full sample (including excluded observations) using alternative specifications.

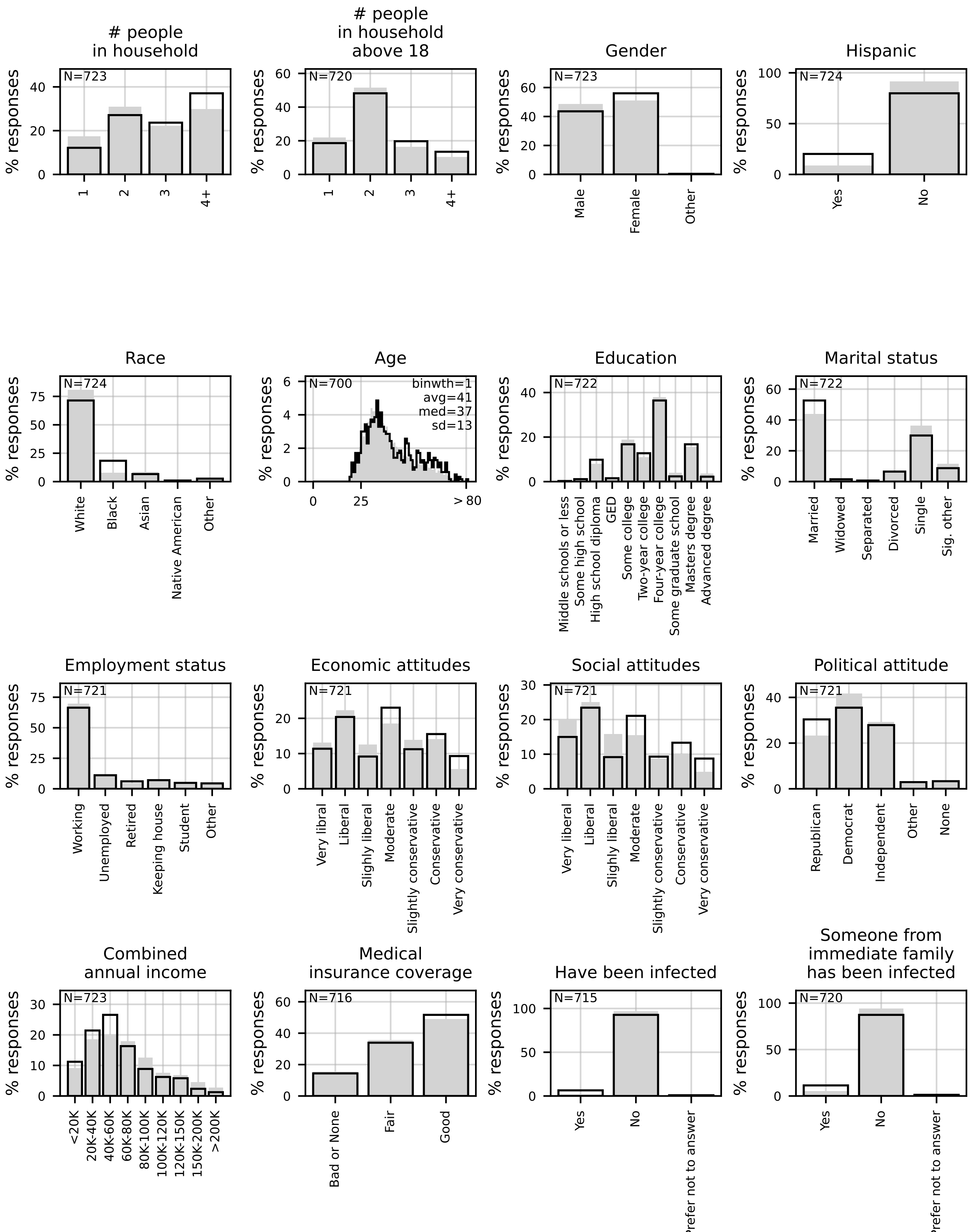
Characteristics of excluded observations. Figures D.4, D.5 and D.6 show the survey duration and demographics of the excluded sample (black lines) vs. those of the main sample (gray). Excluded responses take more time to complete the survey on average, are more conservative-leaning and have less income on average than the main sample's responses, but these differences are not stark.

Figure D.4: Distribution of survey completion time and daily median completion time in excluded sample vs. main sample



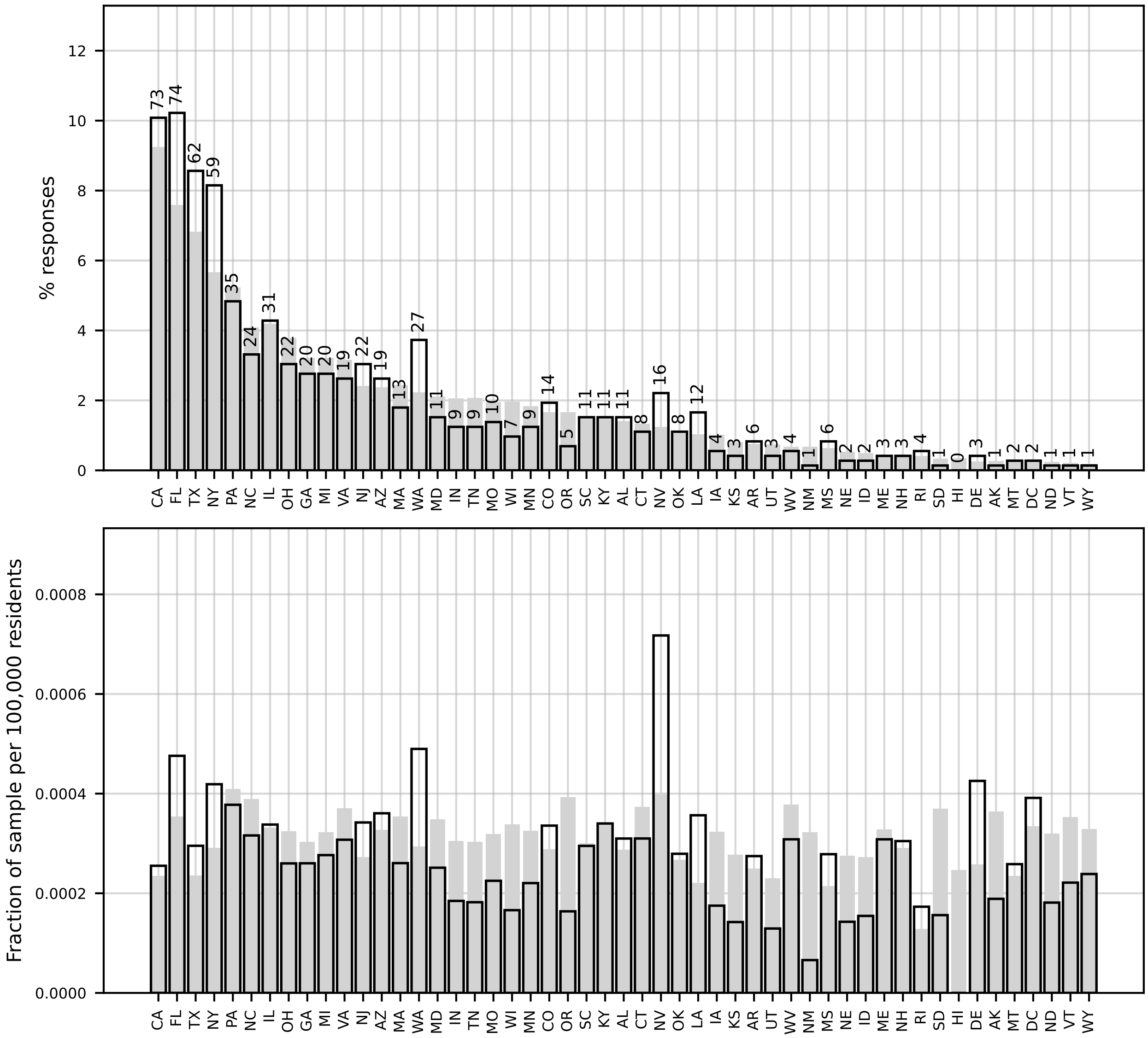
Notes: Black: excluded sample. Gray: main sample. Statistics quoted in the panels refer to the excluded sample.

Figure D.5: Demographic characteristics of excluded sample vs. main sample



Notes: Black: excluded sample. Gray: main sample. Statistics quoted in the panels refer to the excluded sample.

Figure D.6: US-state of residence distribution in excluded sample vs. main sample

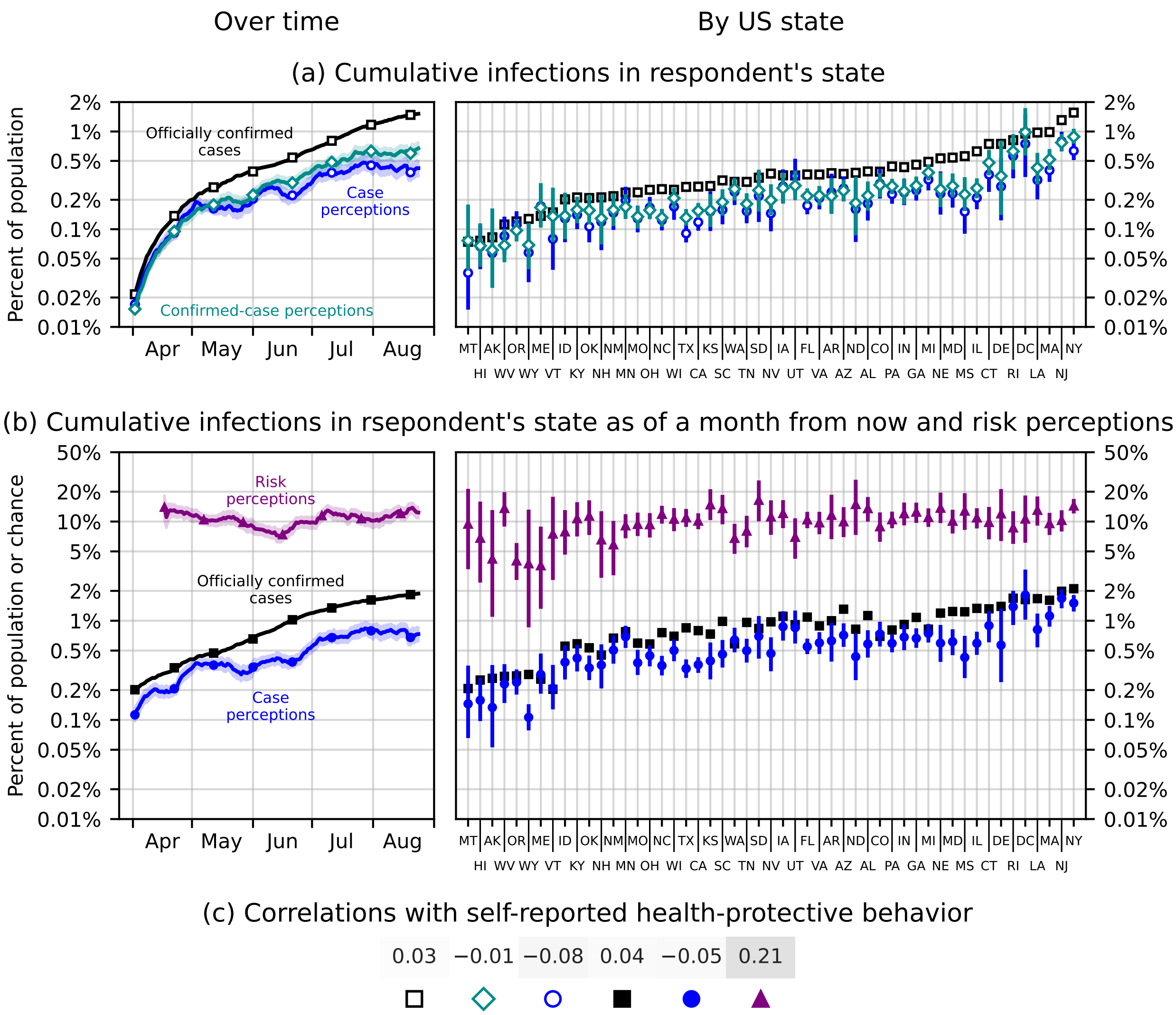


Notes: Black: excluded sample. Gray: main sample. Above bars: absolute numbers of responses from each state in the excluded sample.

Generality of main findings in the full sample. We verify that data exclusion does not drive any of our three main findings by conducting versions of our main analysis on the full sample, rather than the main sample (see Section 1.1). We find that all findings are stable.

First, we repeat the main analysis of the paper using a modified, cumulative, version of case perceptions. Recall that the main analysis constructs case perceptions about *new* infections in the next 30 days, by subtracting perceived cumulative infection cases as of today from predicted cumulative infection cases as of 30 days from now. We therefore have to omit respondents with negative differences between the two. The cumulative analysis is based on case perceptions about *cumulative* cases alone, and can be conducted on the full sample. We define the difference between perceptions of *cumulative* infections in 30 days and the officially confirmed realized numbers as the Cumulative Future Cases gap. The Cumulative Risk-Cases gap is the difference between the these cumulative case perceptions and the risk perceptions used in the main analysis—the state-average infection chance in the next 30 days. Mathematically, since newly infected people in the next 30 days are only a subset of the cumulative number of infected people as of 30 days from now, the Cumulative Risk-Cases gap should be negative. Figure D.7 replicates Figure 1 using cumulative case perceptions in the full, rather than the main sample.

Figure D.7: Cumulative version of main results in the full sample ($N = 13,880$)

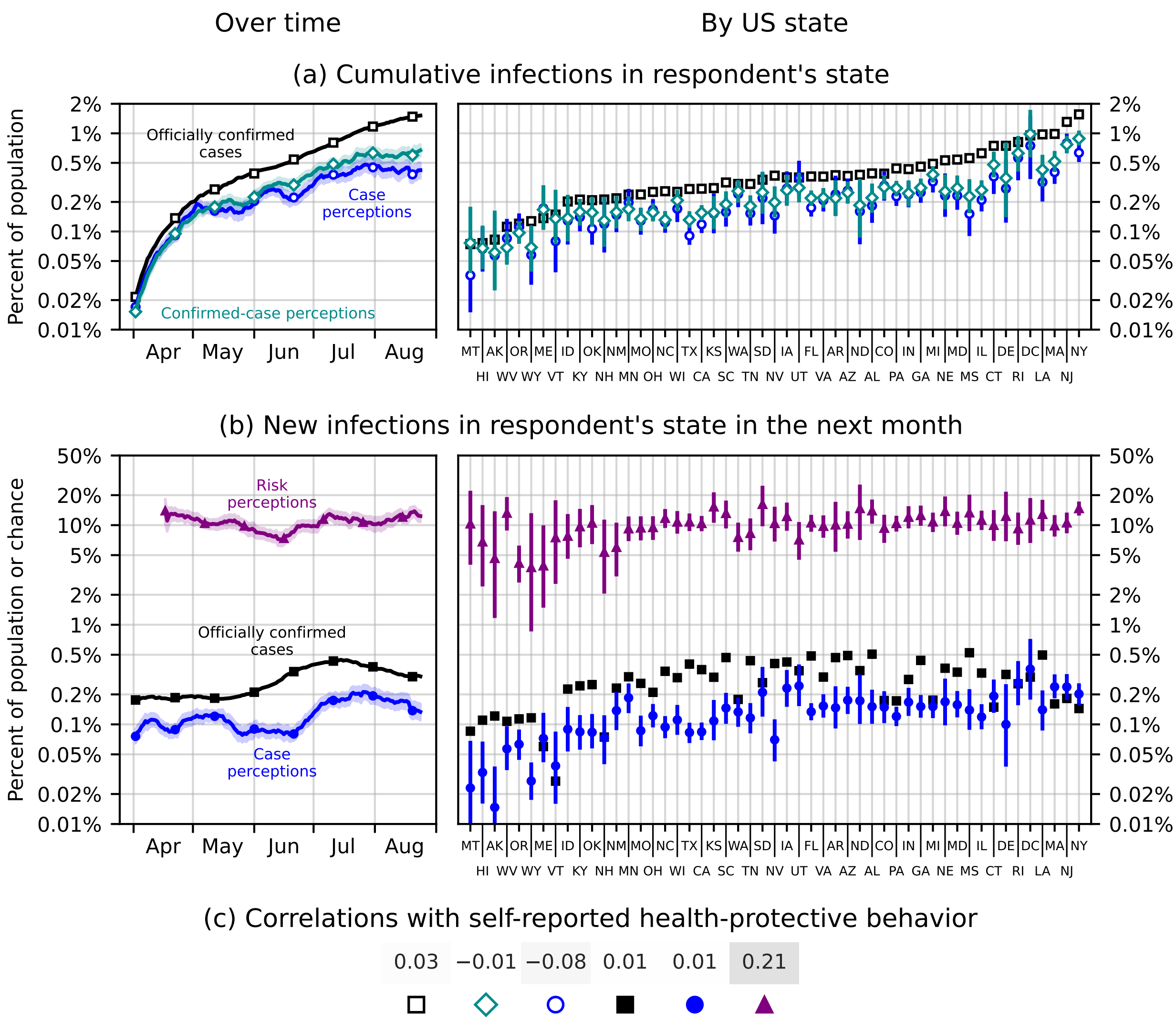


Notes: Light-colored areas in the left panels and error bars in the right panels indicate bootstrapped 95% confidence intervals.

Results are qualitatively similar to the main ones—the Present Cases gap indicates an understatement of current cumulative cases by 49 percent (same as in the main sample), the correlation between perceived and official confirmed infections today is 0.54 (0.57 in the main sample), the Cumulative Future Cases gap indicates an under-prediction of future cumulative cases by 46 percent (38 in the main sample, compared with the non-cumulative Future Cases gap of 52 percent), and the Cumulative Risk-Cases gap indicates that risk perceptions are 21 times larger than cumulative case perceptions on average (19 in the full sample, compared with 79 in the non-cumulative version). This large cumulative Risk-Cases gap indicates that the difference between risk perceptions and case perceptions is substantial, and remains large and hard to rationalize even when including all observations and using a tougher mathematical benchmark. Finally, cumulative case perceptions' correlation with self-reported protective behavior is -0.05 (-0.04 in the main sample, compared with a non-cumulative correlation of 0.02), while risk perceptions' correlation with behavior is 0.21 (0.20 in the main sample).

We conduct a second version of the analysis on the full sample, this time replacing *only excluded observations'* negative predictions of new infection cases in the next 30 days with their predicted cumulative infections cases as of 30 days from today. This accounts for a possibility that respondents with negative differences misunderstood the question as referring to new infections rather than cumulative infections. Again, results are very similar to the main results.

Figure D.8: A version of the main results in the full sample, replacing excluded observations' negative predictions with their cumulative predictions



Notes: Light-colored areas in the left panels and error bars in the right panels indicate bootstrapped 95% confidence intervals.

D.4 Robustness to Matching Between Case Perceptions and Official Reports

Systematic differences between the time of day in which respondents answer the survey and the time of day in which reports are updated at The New York Times may change results regarding case perceptions.

To place bounds on such effects, Table D.5 shows how the Present Cases gap and Future Cases gap change when instead of matching each observation with official reports from the its state-of residence on the same day (the baseline analysis), reports from the previous day or from the next day are being matched. These changes are small relative to the magnitude of the gaps.

Table D.5: Present Cases and Future Cases gap (in log percent differences) as a function of time gap between official reports and survey responses

	Present Cases gap	Future Cases gap
Same-day matching	−0.671	−0.740
	(0.057)	(0.055)
Next-day matching	−0.705	−0.745
	(0.054)	(0.056)
Previous-day matching	−0.635	−0.735
	(0.062)	(0.053)
Obs	13156	13156

Notes: Means are estimated using OLS regressions with a constant only. In parentheses: Driscoll-Kraay standard errors using Bartlett's kernel and a bandwidth of 4 days.

E An Expected-Utility Framework for the Relation Between Beliefs and Behavior

The direction of the relation between case/risk perceptions and risk-mitigating behavior (positive or negative) does not always have a clear benchmark. This appendix develops the relation using a simple EU model and shows that a clear benchmark is more likely to exist when beliefs are about an *exogenous* risk, in principle independent of the one person's choices, and when we have an idea about whether the returns to protective behavior should be increasing or decreasing with baseline risk. The obtained results may differ if beliefs about this exogenous risk are modeled as endogenous, e.g., in models of motivated beliefs or general equilibrium. To keep our underlying model as simple as we can, we abstract from such mechanisms and view the heterogeneity in beliefs about exogenous risk as resulting from heterogeneity in information and in its interpretation, and not from an equilibrium of beliefs and behavior.

Define a person's utility as

$$u(b) = -p(b, \bar{p}) - c(b),$$

where b is the person's protective behavior, $\bar{p}$ is her perception of the exogenous risk (which we consider from now on the relevant exogenous risk itself), e.g., average infection probability in the state (or predicted percent of state infected), $p(b, \bar{p})$ is the endogenous infection risk decreasing in the person's behavior and increasing in the exogenous risk and $c(b)$ is an increasing convex cost function. The FOC is

$$\mathrm{MB}(\bar{p}, b) \equiv -\frac{\partial p(\bar{p}, b)}{\partial b} = c'(b) \equiv \mathrm{MC}(b).$$

At optimum the FOC and an SOC hold:

$$\mathrm{MB}(\bar{p}, b^*) = \mathrm{MC}(b^*),$$

$$\frac{\partial \mathrm{MB}}{\partial b} - \frac{d\mathrm{MC}}{db} < 0,$$

where b^* and p^* are the optimal behavior and endogenous risk respectively.

E.1 Comparative Statics

We are interested in the two comparative statics $\frac{\partial b^*}{\partial \bar{p}}$ and $\frac{\partial b^*}{\partial p^*}$, where $p^* = p^*(\bar{p})$ and $b^* = b^*(\bar{p})$ are the optimal infection probability and behavior. The signs of these derivatives are our benchmarks for the signs of the correlation between risk and behavior, for exogenous and endogenous risk, respectively. Note: these comparative statics are valid *within* a person only, where the relations $p(b, \bar{p})$ and $c(b)$ are indeed fixed.

The relation between behavior and *exogenous* risk, $\frac{\partial b^*}{\partial \bar{p}}$. To obtain the relation between behavior and *exogenous* risk , $\frac{\partial b^*}{\partial \bar{p}}$, take a derivative with respect to $\bar{p}$:

$$\frac{d}{d\bar{p}}\text{MB}\left(\bar{p}, b^*\left(\bar{p}\right)\right) = \frac{d}{d\bar{p}}\text{MC}\left(b^*\left(\bar{p}\right)\right)$$

$$\frac{\partial \text{MB}}{\partial \bar{p}} + \frac{\partial \text{MB}}{\partial b}\frac{db^*}{d\bar{p}} = \frac{d\text{MC}}{db}\frac{db^*}{d\bar{p}}.$$

All derivatives are evaluated at the optimal point. Rearrange to obtain

$$\frac{\partial b^*}{\partial \bar{p}} = \left[\frac{\frac{d\text{MC}}{db} - \frac{\partial \text{MB}}{\partial b}}{\frac{\partial \text{MB}}{\partial \bar{p}}}\right]^{-1}. \tag{1}$$

The relation between behavior and *endogenous* risk, $\frac{\partial b^*}{\partial p^*}$. To obtain the relation between behavior and *endogenous* risk , $\frac{\partial b^*}{\partial p^*}$, take a derivative with respect to b^* and use the relations $\bar{p} = \bar{p}\left(b, p\right)$ (increasing in both b and p; in principle derived from the relation $p\left(b, \bar{p}\right)$) and $p^* = p^*\left(b^*\right)$ (a relation between optimal points):

$$\frac{d}{db}\text{MB}\left(\bar{p}\left(b^*, p^*\left(b^*\right)\right), b^*\right) = \frac{d}{db}\text{MC}\left(b^*\right)$$

$$\frac{\partial \text{MB}}{\partial \bar{p}}\left(\frac{\partial \bar{p}}{\partial b} + \frac{\partial \bar{p}}{\partial p}\frac{\partial p^*}{\partial b^*}\right) + \frac{\partial \text{MB}}{\partial b} = \frac{d\text{MC}}{db}.$$

All derivatives are evaluated at the optimal point. Rearrange to obtain

$$\frac{\partial b^*}{\partial p^*} = \left[\frac{\frac{d\text{MC}}{db} - \frac{\partial \text{MB}}{\partial b}}{\frac{\partial \text{MB}}{\partial \bar{p}}\frac{\partial \bar{p}}{\partial p}} - \frac{\frac{\partial p}{\partial \bar{p}}}{\frac{\partial b}{\partial \bar{p}}}\right]^{-1}. \tag{2}$$

Interpretation. The brackets in both equations 1 and 2 have a first term (in 1 it is the only term) that has a *positive* numerator (due to the SOC), and a denominator whose sign depends on the how the marginal returns to protective behavior change with $\bar{p}$, i.e., on $\frac{\partial \text{MB}}{\partial \bar{p}}$. This makes sense: if protective behavior is perceived as having increasing effectiveness in the exogenous risk, one should invest in more behavior as risk increases. The sign of perceived marginal effectiveness is something that in principle can be measured by panel data or by a designated survey question. However, the brackets in equation 2 have an additional *negative* term whose magnitude depends on the exact relation between b, p and $\bar{p}$, which is harder to measure. We can identify 3 cases:

1. If $\frac{\partial \text{MB}}{\partial \bar{p}} < 0$, then both $\frac{\partial b^*}{\partial \bar{p}} < 0$ and $\frac{\partial b^*}{\partial p^*} < 0$.

2. If $\frac{\partial \text{MB}}{\partial \bar{p}} = 0$, then both $\frac{\partial b^*}{\partial \bar{p}} = \infty$ and $\frac{\partial b^*}{\partial p^*} = \infty$. This is a special case where utility over the dimensions b, p is quasi linear in p, and so optimal solutions have the same value of b^* and varying levels of risk.

3. If $\frac{\partial \text{MB}}{\partial \bar{p}} > 0$, then $\frac{\partial b^*}{\partial \bar{p}} > 0$ but $\frac{\partial b^*}{\partial p^*}$ has an ambiguous sign. In a 2-dimensional framework of b and p subject to the constraint $\bar{p} = \bar{p}\,(b, p)$, $\frac{\partial b^*}{\partial p^*} < 0$ holds when the slopes of iso-$\bar{p}$ lines tangent to the optimal points (b^*, p^*) are increasing in $\bar{p}$.

E.2 Examples

To illustrate the ambiguity of the relation between behavior and endogenous risk $\frac{\partial b^*}{\partial p^*}$ in case 3, we present two examples with $\frac{\partial \text{MB}}{\partial \bar{p}} > 0$.

Example 1: $p = \bar{p}/b$**,** $\textbf{MC} = b$**.** This is an intuitive way to think about protective behavior and the way it mitigates risk. For example, washing hands always removes 90% of germs, thereby reducing risk by a constant ratio. The FOC is $\frac{\bar{p}}{b^2} = b$, so that $b^{*3} = \bar{p}$ and $b^{*2} = p^*$. Hence $\frac{\partial b^*}{\partial \bar{p}} > 0$ and $\frac{\partial b^*}{\partial p^*} > 0$.

Example 2: $p = \bar{p}\,(1 - \alpha b)$ **with** $0 < \alpha < 1$**,** $\textbf{MC} = 1 + \alpha b$**.** The same parameter α in both the risk function and the cost function is assumed to ease algebra. Units are calibrated such that $1 - \alpha b \geq 0$. The FOC is $\alpha \bar{p} = 1 + \alpha b$, so that $\frac{1}{\alpha} + b^* = \bar{p}$ and $\frac{1}{\alpha} - \alpha b^{*2} = p^*$, and then $\frac{\partial b^*}{\partial \bar{p}} > 0$ but $\frac{\partial b^*}{\partial p^*} < 0$.