

On the Accuracy, Media Representation, and Public Perception of Psychological Scientists' Judgments of Societal Change

Cendri A. Hutcherson^{1, 2}, Konstantyn Sharpinskyi³, Michael E. W. Varnum⁴,
Amanda Rotella³, Alexandra S. Wormley⁴, Louis Tay⁵, and Igor Grossmann³

¹ Department of Psychology, University of Toronto Scarborough

² Department of Marketing, Rotman School of Management, University of Toronto

³ Department of Psychology, University of Waterloo

⁴ Department of Psychology, Arizona State University

⁵ Department of Psychological Sciences, Purdue University



At the onset of the COVID-19 pandemic, psychological scientists frequently made on-the-record predictions in public media about how individuals and society would change. Such predictions were often made outside these scientists' areas of expertise, with justifications based on intuition, heuristics, and analogical reasoning (Study 1; $N = 719$ statements). How accurate are these kinds of judgments regarding societal change? In Study 2, we obtained predictions from scientists ($N = 717$) and lay Americans ($N = 394$) in Spring 2020 regarding the direction of change for a range of social and psychological phenomena. We compared them to objective data obtained at 6 months and 1 year. To further probe how experience impacts such judgments, 6 months later (Study 3), we obtained retrospective judgments of societal change for the same domains ($N_{\text{scientists}} = 270$; $N_{\text{laypeople}} = 411$). Bayesian analysis suggested greater credibility of the null hypothesis that scientists' judgments were at chance on average for both prospective and retrospective judgments. Moreover, neither domain-general expertise (i.e., judgmental accuracy of scientists compared to laypeople) nor self-identified domain-specific expertise improved accuracy. In a follow-up study on meta-accuracy (Study 4), we show that the public nevertheless expects psychological scientists to make more accurate predictions about individual and societal change compared to most other scientific disciplines, politicians, and nonscientists, and they prefer to follow their recommendations. These findings raise questions about the role psychological scientists could and should play in helping the public and policymakers plan for future events.

This article was published Online First April 20, 2023.


Cendri A. Hutcherson  <https://orcid.org/0000-0002-4441-4809>


This work was supported by the Social Sciences and Humanities Research Council of Canada Insight Grant 435-2016-1274 (to Cendri A. Hutcherson), the National Science Foundation Grant 2026854 (to Michael E. W. Varnum), and the Social Sciences and Humanities Research Council of Canada Insight Grant 435-2014-0685 and Early Researcher Award ER16-12-169 from the Ontario Ministry of Research and Innovation (both to Igor Grossmann). The authors gratefully acknowledge Alden Lai and the Gallup-Wellness for Planet Earth Global Wellbeing Initiative for providing estimates for well-being, Giving Tuesday for providing estimates for charity domains, and Lynn Vavreck and the Nationscape initiative for providing data on political polarization. The authors have no known conflicts of interest to disclose.

Cendri A. Hutcherson played lead role in writing of original draft, supporting role in data curation, formal analysis, funding acquisition, software and validation and equal role in conceptualization, methodology and visualization. Konstantyn Sharpinskyi played lead role in formal analysis, supporting role in conceptualization, data curation, methodology, project administration, writing of original draft and writing of review and editing and

equal role in visualization. Michael E. W. Varnum played supporting role in conceptualization, funding acquisition, methodology, writing of original draft and writing of review and editing. Amanda Rotella played supporting role in conceptualization, methodology, project administration and writing of review and editing. Alexandra S. Wormley played lead role in validation and supporting role in data curation, formal analysis, visualization and writing of review and editing. Louis Tay played supporting role in methodology and writing of review and editing. Igor Grossmann played lead role in data curation, funding acquisition, supervision and writing of review and editing, supporting role in formal analysis, validation, visualization and writing of original draft and equal role in conceptualization and methodology.

Data and study materials, along with code for reproducible analyses are available at <https://osf.io/9btsy/>. This study's design, data exclusions, and portions of the analysis were preregistered; see <https://osf.io/zxavd/>.

 The data are available at <https://osf.io/9btsy/>.

 The experimental materials are available at <https://osf.io/9btsy/>.

Correspondence concerning this article should be addressed to Igor Grossmann, Department of Psychology, University of Waterloo, PAS 3047, Waterloo, ON N2L 3G1, Canada. Email: igrossma@uwaterloo.ca

Public Significance Statement

At the onset of the COVID-19 crisis, psychological scientists contributed to the public discourse on COVID-related societal change in the news media through intuition-based reasoning and often made predictions outside their area of expertise. We assessed the likely accuracy of such judgments by surveying psychological scientists and laypeople at the onset of the pandemic regarding future societal change in different domains and comparing predictions to actual markers of change at 6 months and 1 year after. We found that psychological scientists and laypeople made similar and largely inaccurate predictions. Neither direct experience, training, nor domain-specific expertise was associated with greater accuracy.

Keywords: scientific intuitions, science communication, COVID-19, forecasting, lay theories of change

Supplemental materials: <https://doi.org/10.1037/amp0001151.supp>

Formal training in the social sciences typically focuses on developing explanatory theories that account for observed, often laboratory-based, phenomena. Although such approaches have resulted in richly detailed causal models of individual human behavior, recent years have witnessed growing calls to understand whether and how to increase their usefulness (Watts, 2017). Can psychological theory scale up to predict larger societal processes (Yarkoni & Westfall, 2017) in ways that enable effective intervention in times of crisis (IJzerman et al., 2020)? At the onset of the COVID-19 pandemic, there were a handful of notable efforts by psychologists and other behavioral and social scientists to provide formal guidance in academic journals about what areas of individual and societal behavior might be affected (Brooks et al., 2020; van Bavel et al., 2020; West et al., 2020). Scientists also attempted to contribute to public understanding of the potential consequences of the pandemic through discussion in public media, such as newspapers and magazines. How often and on what basis did psychological scientists make those judgments? For example, are such public judgments grounded in a more intuitive and heuristic reasoning style, or are they based on some combination of expert knowledge and formal modeling of potential outcomes? Are such judgments of societal change accurate? Here, we ask how psychological scientists made on-the-record judgments about societal change in public media, and formally assess whether the nature of their expertise gives them an advantage in the accuracy¹ of their judgments about future outcomes, compared to an average nonexpert.

Understanding how psychological scientists make public judgments is critical for determining their accuracy and potential usefulness. On the one hand, psychology training and expertise should improve understanding of probability and statistics (Fong & Nisbett, 1991; Nisbett et al., 1987) and reduce mistaken assumptions about human behavior (Gardner & Dalsing, 1986; Gardner & Hund, 1983; Taylor & Kowalski, 2004)—qualities that tend to increase the accuracy of forecasts for discrete geopolitical events (Mellers et al., 2015).

Furthermore, the existence of empirically grounded, causal theories about human responses to social isolation (Hawley & Cacioppo, 2010), financial uncertainty (Artazcoz et al., 2004), and disease threat (Schaller & Park, 2011) should enable psychologists to estimate, at minimum, the direction of changes in psychology and behavior in response to the pandemic. On the other hand, research on forecasting in domains ranging from political (Tetlock, 2005) and economic (Armstrong, 1985) to career-related outcomes (Ægisdóttir et al., 2006; Dawes et al., 1989) suggests that experts are rarely more accurate than simple statistical models (Tetlock, 2005). Moreover, psychological theories are typically applied only at the individual or local level. Different forces may be at play when generalizing to societal processes writ large (Na et al., 2010; Piantadosi et al., 1988).

To better understand the nature of psychological scientists' contribution to public understanding in times of crisis, we first analyzed the world's largest corpus of COVID-19 news reports, tracing the nature of psychological scientists' engagement with the news media during March–May 2020, finding that these judgments were typically made in an intuitive style, relying on analogical reasoning and only occasionally on reference to research, and that more than a quarter of judgments were done outside of scientists' domain-specific expertise (Study 1). Then, we present a systematic investigation into the accuracy of such judgments (in both absolute terms and relative to laypeople) for predicting and retrospectively evaluating aggregate-level changes in human psychology and behavior during the first 6 months of the pandemic (Studies 2 and 3). We find that for most domains, scientific judgments of the kind found in public discourse were either at chance or

¹ Our main operationalization of accuracy concerns prediction of direction of societal change, because most psychological causal models of human behavior or social processes lend themselves to predictions about direction of change (e.g., “if X occurs, violence will increase”) rather than estimates of specific magnitude (e.g., “if X changes by Y amount, violence will increase by Z amount”). **Online Supplemental Results** show similar conclusions when accuracy is operationalized via magnitude or the rank ordering of change across different domains.

largely inaccurate and not more accurate than judgments of laypeople. Domain-specific expertise did not facilitate prediction accuracy. Finally, we show that understanding psychological scientists' accuracy matters because the public expects them to be more accurate in predicting psychological and societal outcomes and prefers to base policy on their recommendations compared to politicians, laypeople, and other scientific disciplines (Study 4).

Method

The project was approved by the Office of Research Ethics at the University of Waterloo (Nos. 42123 and 43189). Preregistration, materials, methods, code, and reproducible analyses are available on the Open Science Framework at <https://osf.io/9btsy/> (Hutcherson et al., 2023).

Study 1: News Media Engagement of Psychological Scientists

In Study 1, we examined how psychological scientists talked about the pandemic in the news media. To this end, we used The Coronavirus Corpus (<https://www.english-corpora.org/corona/>)—an extensive record of over 1.8 million texts appearing in online newspapers and magazines in 20 different English-speaking countries—to identify online articles in newspapers and magazines that contained an interview with an academic psychologist regarding some aspect of the pandemic. To identify candidate articles, we first located texts containing key search terms (e.g., psychologist, psychology professor, psychology researcher), limited to a publication date between March 15, 2020, and May 15, 2020. From this, we identified a subset of texts containing interviews with psychological scientists about the pandemic (see Supplemental Figure S1, for a flowchart detailing the definition and identification of psychological scientists). These articles were then reviewed by hand to remove duplicates and apply additional exclusions (e.g., not actually containing an interview with a psychologist despite containing key search terms). This produced a database of 169 unique English-language articles appearing in a wide variety of outlets, including the *New York Times*, *Wall Street Journal*, and other high-quality news sources, presenting judgments by 213 individual scientists. Because these scientists frequently commented on multiple distinct topics (e.g., effects of the pandemic on depression and also children's cognitive development), this yielded 719 unique judgments about the consequences of the pandemic, which were coded by three independent raters for whether the expert's judgment on a given topic fell within their particular domain of expertise, whether it was an observation about the present or a prediction for future outcomes, and what type of justification was given (i.e., none, current events, historical analogy, research, other), as well as the certainty of language used (interrater agreement 79%–85%, see Supplemental Methods, for details).

Study 2a: Psychological Scientists' Predictions About Societal Change

Participants

In the first 2 days of April 2020, we recruited psychology experts by circulating a call for forecasts on listservs and mailing lists for the Society for Personality and Social Psychology, the Cognitive Science Society, Society for Research in Child Development Commons, Association for Behavioral and Cognitive Therapies, and the Society for Judgment and Decision Making. We also posted in relevant Facebook groups, including Psychological Methods, PsychMAP, and COVID-19 groups. Additionally, we contacted colleagues and graduate students at the authors' affiliated departments and institutes.

A total of 470 scientists provided their forecasts in April. Of these, six had incomplete responses, four participants provided nonsensical responses (e.g., age > 900), 57 participants answered all survey questions in less than 5 min (pilot testing with research associates revealed that 5 min is the minimum necessary time to complete the study), and two participants indicated they were undergraduate students. These responses were removed. The final sample ($N = 401$) consisted of participants from 39 countries, with demographics that closely match the membership of psychological societies relevant to these predictions (see Supplemental Table S1).

Procedure

Participants first answered several demographic questions. Participants next predicted cultural change in the United States for 11 domains, presented in a randomized order: implicit and explicit prejudice toward minorities, political polarization, traditionalism, individualism, generalized trust, delay of gratification, expected birth rates, concern for climate change, life satisfaction, and clinical depression (see verbatim questions on Open Science Framework at <https://osf.io/npzcr>; Sharpinsky et al., 2022). Participants provided predictions for 6 months, 1 year, and 2 years in the future on a sliding scale ranging from 50% or greater decrease (−50) to 50% or greater increase (+50). Of these 11 domains, we were able to obtain reliable benchmarks to assess accuracy for seven: polarization, traditionalism, individualism, trust, climate change, life satisfaction, and depression (see Accuracy Analyses section).

Beyond the 11 domains we provided to participants to make forecasts, we were interested in participants' unstructured views about the key societal domains in which one might observe significant changes. After participants predicted cultural change in the above variables, we asked them to identify one key psychological or social issue in the United States not covered in the survey that they thought would change.

Studies 2b and c

Participants

Whereas Study 2a focused on predictions before the initial peak of COVID-19 cases in the United States, Studies 2b and c were conducted after the initial peak. In the last week of April and the first week of May 2020, we recruited another group of psychological scientists using the same methods described in Study 2a. A total of 354 psychological scientists provided their forecasts during this time (98% nonoverlapping with the early April sample). Of these, we removed two who had incomplete responses, 31 who completed the survey too fast according to pilot test estimates (<5 min), and four who indicated they were undergraduate students. The final sample included 316 participants from 26 countries (see [Supplemental Table S1](#), for demographic information).

Concurrently, in the first week of May 2020, we also obtained forecasts from a nationally representative sample of English-speaking U.S. residents via the crowdsourcing U.K.-based company Prolific (<https://www.prolific.co>, Study 2c). To recruit a nationally representative sample, Prolific uses the intended sample size (target $N = 400$) to stratify across age, sex, and ethnicity based on census data from the U.S. Census Bureau. Of the 411 participants who attempted the study, we removed three who had incomplete responses and 14 who completed the survey in less than 5 min. The final sample consisted of 394 participants. Prolific participants received 1.25 USD for completing the survey.

Procedure

Participants in Studies 2b and 2c followed the same general procedure outlined for Study 2a, with the following differences. In addition to the 11 domains of Study 2a, they made predictions in four additional domains: loneliness, religiosity, charitable giving, and prevalence of violent crimes (verbatim questions on <https://osf.io/npzcr>; Sharpinskyi et al., 2022, and in [Supplemental Table S1](#)). For each domain, participants made predictions as in Study 2a but also rated confidence in their predictions on a 5-point scale (1 = *not at all* to 5 = *extremely*). Participants also answered additional demographic questions (see [online Supplemental Material](#), for verbatim items). Of these 15 domains, we were able to obtain reliable benchmarks to assess accuracy for 10: loneliness, charitable giving, violent crimes, polarization, traditionalism, individualism, trust, climate change, life satisfaction, and depression. Thus, in the remainder of the article, we focus on responses from our participants in these domains (see the Accuracy Analyses section). Although this number of domains is not extensive, we think it likely represents a best case scenario for assessing the utility and accuracy of psychological scientists' forecasts, since these are the domains for which (a) psychology has established theories about how change might occur in the face of the pandemic, (b) there was sufficient

interest that high-quality data were being measured during the pandemic, and (c) psychologists were generally more likely to comment in the media. However, we acknowledge that all conclusions we make come with caveats due to the limited number of domains and apply largely to domains in which psychology makes straightforward pandemic-related predictions rather than all possible judgments in general.

Studies 3a and b: Psychological Scientists' Retrospective Judgments About Societal Change

In Studies 3a and b, we aimed to compare prospective predictions from Study 2 to retrospective estimates of changes in these same domains. Study design and data exclusions were preregistered (registration available at <https://osf.io/zxavd/>; Hutcherson, 2022).

Participants

We recruited a new regionally stratified sample of Americans from Prolific. Participants received 1.10 GBP for participation. Exclusion criteria were identical to Study 2, with the exception that we also preregistered to exclude participants who provided estimates for fewer than five domains or indicated at the end of the survey they took part in the April/May prediction studies even though there was no April survey for Prolific and none of the Prolific IDs from the May, 2020, survey matched their Prolific IDs). Of the 445 participants who started the study, we removed 27 who had incomplete responses and seven who indicated they took a forecasting survey in April. The final sample consisted of 411 participants.

We also obtained survey responses from a sample of psychological scientists recruited via mailing lists (e.g., Social and Personality Psychology mailing list, SJDM mailing list) and social media. Similar exclusion criteria were applied to this sample, with the exception that we did not require scientists to be U.S. citizens. A total of 350 psychological scientists provided forecasts during the last week of October/first week of November 2020 (88% nonoverlapping with the forecasting samples in Studies 1–2). Of these, we removed 80 responses because they provided fewer than five domain estimates. The final sample included 270 participants (see [Supplemental Table S1](#)).

Procedure

Participants in Study 3 were asked to provide retrospective assessments of percentage change as well as confidence in their assessments for the same 15 domains as in Study 2b (verbatim questions on <https://osf.io/9btsy/>; Hutcherson et al., 2023, and additional information in [online Supplemental Material](#)). To match instructions in Study 2, participants were instructed to “provide an estimate of how much you think it has changed compared to where the issue stood six months ago (i.e., end of April 2020).” In addition, as an exploratory

analysis, we obtained information about the types of information participants considered when making their judgments, including whether they considered specific news reports or brought to mind vivid personal memories (see [online Supplemental Materials](#), for detail). All other details were as in Study 2b.

Study 4: Lay Perceptions of Scientists

Participants

For Study 4a, we recruited a sample of Americans from Prolific in March 2021. Participants received 1.10 GBP for participation. Of the 220 participants who started the study, we removed 11 who did not provide any responses and six who did not provide a comprehensible answer to an open-ended question at the end of the study. The final sample consisted of 203 participants ($M_{\text{age}} = 33.81$, $SD_{\text{age}} = 13.04$; 57% female/41% male/2% nonbinary; 74% White/7% Latinx/8% Asian American/6% Black/5% other). To supplement these results, we also recruited a sample of academics and policymakers via announcements on social media (Study 4b). Thirty individuals filled out the survey ($M_{\text{age}} = 40.32$, $SD_{\text{age}} = 10.64$; 57% female/39% male/4% nonbinary; 78% White/9% East Indian/9% mixed/4% other).

Procedure

Participants considered different groups of scientists and practitioners, as well as the layperson and rated their possible accuracy when predicting societal change over COVID-19 for depression, life satisfaction, loneliness, violence and related domains, and who would they like to make recommendations for these societal issues. Participants were presented with three sets of questions concerning predictions, preference for recommendations, and ranking of the top three groups they would prefer to ask how the COVID-19 pandemic will affect human behavior and society in the long term. For each set of questions, participants were presented with 10 groups: scientists with expertise in psychology, economics, epidemiology, history, political science, or public health; practitioners with expertise in social work or medicine; as well as politicians and the average American. See the project's Open Science Framework page (<https://osf.io/9btsy/>; Hutcherson et al., 2023), for the precise wording of questions.

We also examined whether participants read prior reports about behavioral science expertise for predicting societal trajectories over COVID-19. Only 7% of the sample indicated vague familiarity with such reports, and excluding these participants did not change the results.

Accuracy Analyses (Studies 2a–c and Studies 3a and b)

We targeted estimates for all domains where we could locate large-scale, nationally representative surveys assessing

the state of that domain in April/early May and in October/early November. Our chief question concerned societal-level change. Thus, we relied on cross-sectional data as long as the estimates were sufficiently large and representative of the U.S. population at large. When possible, we used weighted averages to adjust for representativeness as per the U.S. census. If we could locate multiple sufficiently representative indicators for a given domain, we performed parallel analyses with each. Our sources included the Household Pulse Survey from the National Center for Health Statistics and the U.S. Census Bureau, University of Southern California's Understanding America Survey, Nationscape, Gallup Panels, the National Commission on COVID-19, Criminal Justice, and Giving Tuesday, among others. See [Supplemental Table S2](#), for the exact wording of the questions, and [Supplemental Table S3](#), for more information on each source. When estimates were based on the percentage of the population at the given time point, we calculated the difference score. When the data were based on the sample estimate of a scale-based response, we calculated the percentage change between the initial estimate of the sample in April 2020 and the subsequent estimate half a year later. Ultimately, we quantified societal change in the United States for 10 domains, with most estimates coming from nationally representative surveys and aggregated official reports of crime. We report estimates for two additional benchmarks with lower sampling consistency (prejudice markers from Project Implicit) in the [online Supplemental Material](#). In addition, we obtained objective benchmark data for four of these 10 domains 1 year after the start of the pandemic (five surveys were no longer collecting data, preventing comparable accuracy analyses). Additionally, for 1-year markers, we also obtained U.S. birth rate statistics from the Human Fertility Database (<https://www.humanfertility.org/>), a joint project of the Max Planck Institute for Demographic Research (<http://www.demogr.mpg.de/>) in Rostock, Germany, and the Vienna Institute of Demography (<https://www.oeaw.ac.at/vid/>) in Vienna, Austria.

Our main criterion for accuracy was the direction of change (increase/decrease) as a function of the type of estimate (prospective/retrospective), sample type (lay/expert), and domain type. In addition to frequentist statistics, we ascertained the strength of evidence for or against specific hypotheses about the accuracy of psychological scientists using an estimation of Bayes factors (Rouder et al., 2009) provided by the function *bayesfactor_models* from the *bayestestR* package. In secondary analyses, we compared the magnitude of predicted change to observed change, including both the average estimated change at 6 months as well as the estimated trajectory of change over the full 2-year prediction period. In addition, we performed a number of supplemental analyses quantifying accuracy at 6 months by the percentage of the sample falling within a certain range of observed change as well as rank-order accuracy across domains (i.e., predicting

which domains would show the *most* vs. *least* change). See additional results in the [online Supplemental Material](#), for details.

Results

Study 1: Psychological Scientists' Judgments in the News Media

To understand how psychological scientists' judgments might shape public perceptions, we first sought to understand *how* they typically make such judgments and to document how the topics they discussed aligned with their expertise. This analysis allowed us to answer a crucial question: when communicating to the public, how often do psychological scientists base their judgments on discipline-specific expertise, theory, and models, or instead use an intuitive or heuristic reasoning style that might be shared with nonexperts?

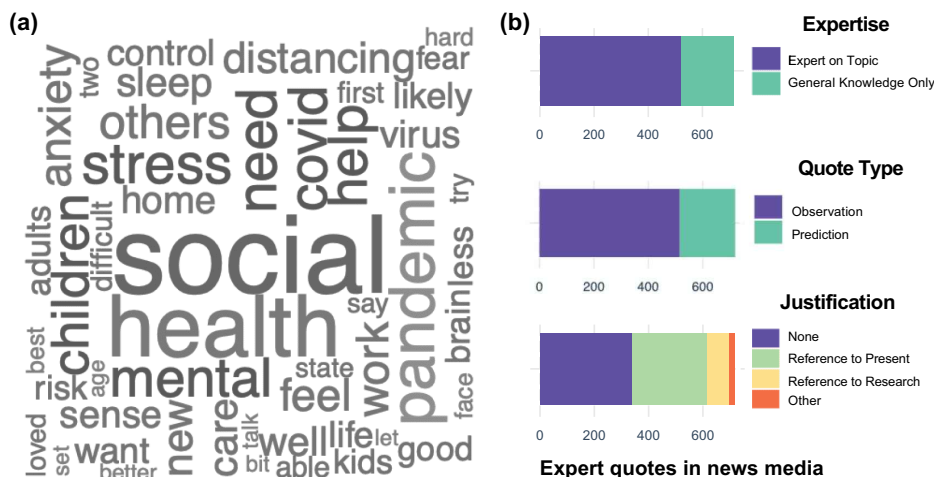
To determine what experts were saying about the societal impact of COVID-19, whether they were making predictions, and how they made them, we analyzed the comments made by psychological scientists to the news media in the first 2 months of the pandemic (see Method section, for details). Analysis of the frequency of content-related words (excluding generic terms like "well," "if," etc., and terms related to "psychological scientist," which were used to identify the articles) indicated that experts spoke on a number of topics focused on mental health, well-being, and various social effects of the pandemic (Figure 1a, b). Though most interviews with experts

concerned observations about the current effects of the pandemic (72% of cases), explicit forecasts about the pandemic's future consequences were also common (28% of cases). When talking to the news media, more than a quarter of judgments were made outside of scientists' area of expertise (27% of cases). We observed no evidence of a difference between scientists speaking within or outside their domain of expertise in the likelihood of making a prediction versus an observation, $\chi^2(1, N = 717) = 1.02, p = .31$.

Finally, we assessed what justification/rationale psychological scientists provided for their judgments. We found that for a sizable fraction of statements (47%), no justification for the judgments was included. When a justification was provided, it rarely referenced research or scientific theory (21% of cases). In most cases (73%), scientists were quoted making intuitive reference to present events (e.g., noting the hoarding of toilet paper when justifying the influence of the pandemic on panic responses).

To determine whether this lack of scientific justification could be attributed simply to omission by journalists, we analyzed separately op-eds in which a psychologist spoke for themselves rather than articles in which they were quoted by a journalist. Although op-eds were more likely to give any kind of justification for a judgment (68% of op-ed judgments vs. 51% of quoted judgments), $\chi^2(1, N = 717) = 6.84, p = .01$, we found no evidence of a significant difference in the likelihood of that justification being based on research (19% of justifications in op-eds, 22% of justifications in other news

Figure 1
Psychological Scientists in the Media



Note. (a) Analysis of the frequency of different words in media interviews with experts shows that they commented on a number of topics, including health, mental well-being, stress, and social relationships. (b) Analysis of these interviews also suggests that psychologists frequently spoke outside their domain-specific topic of expertise, frequently made predictions about future outcomes of the pandemic, and that quoted justifications typically involved intuitive reasoning rather than reference to specific research findings. See the online article for the color version of this figure.

articles), $\chi^2(1, N = 376) = 0.07, p = .80$. Thus, even when psychological scientists were given full control of the narrative via an op-ed format, justifications were either absent or merely reflected references to present-day events.

Across all article types, a significant difference emerged in the type of justification between scientists speaking within or outside their domain of expertise, $\chi^2(3, N = 717) = 9.63, p = .02$. Domain-experts' judgments were significantly more likely to reference research compared to scientists without domain expertise (13% of expert judgments vs. 5% of nonexpert judgments), $z = 3.75, p < .001$, whereas non-experts were equally likely to omit versus provide justifications for their judgments (51% vs. 46%), $z = 1.05, p = .15$. However, when giving a justification, both domain experts and nondomain experts were more frequently quoted referencing current events than research (70% of justifications for domain experts, 82% of justifications for nondomain experts), both binomial tests $p < .001$.

The Accuracy of Psychological Scientists' Spontaneous Judgments: Studies 2 and 3

Our analysis of the types of judgments made by psychological scientists in the news media suggested that these judgments might often be made on the spot, without an extensive rationale, or with an intuitive rather than empirical basis for judgment. This observation raises a question about the accuracy of judgments that psychological scientists conveyed to the media in the aftermath of the COVID-19 pandemic.

To address this question, we analyzed predictions about outcomes of the pandemic in the United States from two samples of psychological scientists, one collected in early April 2020 (Study 2a; $N = 401$) and another collected in late April/early May 2020 (Study 2b, $N = 316$). Scientists could make these predictions however they chose, including formal model analysis. Survey completion times and post hoc analysis of self-reported strategies suggest that the majority likely relied on spontaneous, intuitive judgments informed by both training and life experiences, similar to what we observed in news media quotes (see [online Supplemental Results](#), for details), although we acknowledge that such interpretation is speculative. Predictions were obtained about change in different domains (e.g., depression, political polarization) at 6 months, 1 year, and 2 years into the future (see [Supplemental Figures S2 and S3](#), for predicted trajectories). Although we aimed primarily to recruit psychological scientists (composing ~80% of the sample), we also attracted responses from individuals in other behavioral science disciplines, such as economics, political science, and sociology, allowing us to compare psychological and nonpsychological disciplines. However, these analyses indicated little consistent distinction among disciplines on predictions or accuracy (see [Supplemental Tables S12–S13](#), [Figure S12](#), and [Supplemental Results](#)). We thus report statistics for the full

sample here, focusing on other definitions of expertise (e.g., domain-specific training, level of education) as potential moderators of accuracy.

In each of these surveys, we asked our participants to consider specific domains for which a sizable body of theoretical and empirical work links these variables to pathogen-related threats. We focus here on domains for which we could obtain high-quality, national-level data. Based on theories that suggest that intergroup processes are affected by evolutionary and ontogenetic pressures related to pathogen stress (Faulkner et al., 2004; Fincher et al., 2008; Murray et al., 2011; Schaller & Park, 2011; Tybur et al., 2016), we examined judgments of political polarization, cultural values related to traditionalism and individualism, as well as prosocial and antisocial behavior. Based on life history theory, which argues that organisms increase present-focused behavior and reproduction in response to environmental threat and pathogen-related unpredictability (Griskevicius et al., 2011; Horn, 1978), we assessed birth rates. Finally, based on theories about how human mental and affective well-being is influenced by stressors (Kendler et al., 1999), including social isolation (Hawkley & Cacioppo, 2010), we assessed judgments of depression, loneliness, and life satisfaction.

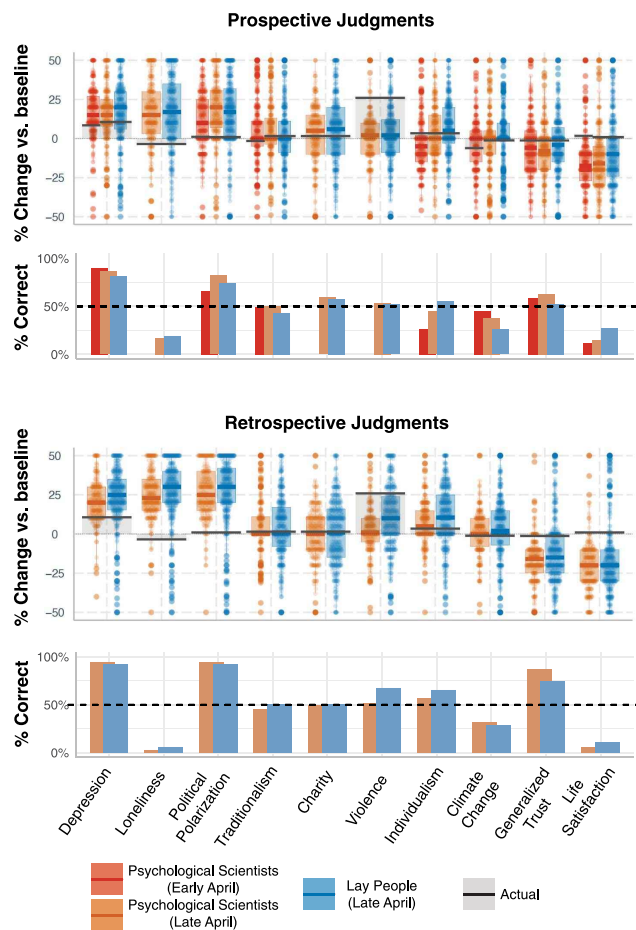
To assess the accuracy of such judgments, we compared predictions for 6 months to ground truth markers of change at 6 months for depression, life satisfaction, generalized trust, loneliness, individualism, traditionalism, political polarization, climate change attitudes, violent crimes, and charitable giving. At 12 months, we were able to obtain high-quality ground truth markers for five domains: depression, loneliness, birth rates, violent crimes, and charitable giving (see also estimates for explicit and implicit prejudice in the [online Supplemental Material](#), for both 6 and 12 months).

Study 2a and b: Were Scientists More Accurate Than Chance in Predicting Societal Change Across Domains?

We answer this question by comparing scientists' predictions for 6 and 12 months into the pandemic against ground truth. At the 6-month mark, we examined the intercept term of a mixed-effects logistic regression (see [online Supplemental Materials](#), for detail) with directional accuracy (1 = correct/0 = incorrect) in each domain as the dependent measure and participant (total $N = 707$) as a random intercept. This yielded an average individual accuracy of 50.5% [49.0, 51.7], a value that was not significantly different from chance, $z = 0.61, p = .54$.

We then investigated how accuracy varied by domain at 6 months. As [Figure 2](#) (top panel) shows, scientists showed above-chance directional accuracy in only four out of 10 domains at 6 months. They correctly predicted increases in depression (89% correct), binomial test against chance accuracy of 50% $p < .001$, corrected for multiple comparisons; political polarization (73% correct), $p < .001$; and charitable

Figure 2
The Accuracy of Prospective (April and May 2020) and Retrospective (October/November 2020) Judgments of Societal Change



Note. Predictions, along with objective markers for 10 available domains, are displayed for prospective (top) and retrospective (bottom) judgments in psychological scientists and laypeople. Boxplots show median and 25/75% confidence intervals. Accuracy (measured as directionally correct predictions) is displayed just below predictions for prospective and retrospective judgments. Prospective data include two separate samples of psychological scientists surveyed in late March/early April and late April/early May). We thus display objective benchmark data separately for the two time periods where it is available. Retrospective data included a single sample of psychologists and laypeople collected in late October/early November. See the online article for the color version of this figure.

giving (59% correct), $p = .03$; and correctly predicted decreases in generalized trust (61% correct), $p < .001$. However, in most of these domains, scientists tended to significantly overestimate the magnitude of changes, all t tests against actual change > 2.77 , $.006 < ps < .001$, corrected for multiple comparisons. The only exception concerned charitable giving, t test against actual change = 1.55, $p = .12$ uncorrected. Moreover, for the remaining six domains, they either failed to predict direction above chance levels or were actually significantly worse than chance. They incorrectly predicted decreases in life satisfaction (13% correct), binomial test

$p < .001$; loneliness (17% correct), $p < .001$; individualism (35% correct), $p < .001$; and concern for climate change (42% correct), $p < .001$, and were no better than chance in predicting changes in traditionalism (50% correct), $p = .97$, and violence (54% correct), $p = .27$. Conclusions did not change when using alternative measures of accuracy, such as absolute deviation or rank ordering of the magnitude of changes across domains (see online Supplemental Results, for details). Conclusions were also similar when making a more granular comparison of early and late April predictions with early and late October markers, respectively (see online Supplemental Results, for relevant details).

Scientists' average individual accuracy for 12-month predictions (Supplemental Figure S9) was even worse than at 6 months and substantially lower than chance, 35.1% [33.3, 37.1], $z = -13.58$, $p < .001$. Moreover, when investigating accuracy by domain, psychological scientists made accurate *directional* predictions for only two out of the five domains for which we had objective markers: birth rate (54% correct) and violence (65% correct), $ps < .02$. However, they nonetheless either over- or underestimated the *magnitude* of change in these domains, $ps < .001$. In addition, directional accuracy for the remaining three domains was significantly worse than chance, all $ps < .001$ (depression—9% correct, loneliness—21% correct, and charity—38% correct).

Study 2c: Expert Predictions Are Not More Accurate Than Lay Predictions

Although psychological scientists' judgments were largely inaccurate overall, it could still be the case that these predictions were more accurate on average than those of laypeople. To test this possibility, we collected predictions from a nationally stratified sample of Americans ($N = 394$) in late April/early May 2020, in parallel with the collection of our second sample of psychological scientists. We then compared the directional accuracy of lay predictions to psychological scientists. Results of a generalized linear mixed model with accuracy of prediction scores for direction of change as a dependent variable and expertise as a predictor revealed no evidence for a difference between psychological scientists and laypeople, $\chi^2(1, N = 1,101) = 2.14$, $p = .14$. Moreover, comparison of a model including group (scientist vs. lay) as a factor to a model without this factor yielded a Bayes factor (Rouder et al., 2009) of 95 in favor of the null. Thus, the evidence increased credibility of the null hypothesis that there was no advantage for scientists over lay people in predictive accuracy, at least in the kinds of domains examined here.

Studies 3a and b: Retrospective Judgments Were Not More Accurate Than Prospective Ones

Our results suggest that the prediction of large-scale trends in psychological and societal outcomes might be difficult for

both scientists and laypeople alike, and that scientists were similar in accuracy to the average American. However, this could occur for many reasons related to chaotic or unpredictable dynamics in response to the pandemic. We reasoned that if experts miscalculated the effects of the pandemic solely due to unforeseeable dynamics, but would otherwise have made more accurate judgments about how the pandemic affects psychological outcomes, then expert judgments of change should be more accurate in retrospect, especially compared to laypeople. In other words, experts should be better able to update their judgments in light of experience and/or direct observation of empirical data (although, for most domains, these data did not yet exist or have not yet been published, likely leaving most scientists to rely on the same kinds of intuitive experiences and knowledge as laypeople).

To assess whether this was the case, we conducted a set of preregistered surveys (<https://osf.io/9btsy/>; Hutcherson et al., 2023) in a third sample of psychological scientists (Study 3a, $N = 270$) and a nationally stratified sample of lay Americans (Study 3b, $N = 411$) in late October/early November, just before the U.S. election. We asked participants to estimate how much change *had already* occurred in the previous 6 months rather than make forecasts of future change.

Our results suggested that retrospective judgments improved slightly compared to scientific predictions half a year prior (Figure 2, bottom). On average, across all domains, psychological scientists had an accuracy rate in retrospective reports of 51.9% [50.0, 53.7], which was slightly but significantly more accurate than prospective reports, odds ratio = 1.25, $z = 3.35$, $p = .001$. However, the domains in which psychological scientists' prospective judgments were inaccurate were the same domains in which their retrospective judgments were inaccurate. Moreover, in domains where most predictions were inaccurate, even larger numbers of retrospective assessments were directionally inaccurate, $\chi^2(1, N = 1,782) = 150.51$, $p < .001$, in large part because predictions became more extreme (see online Supplemental Results). Finally, despite the fact that psychological scientists' accuracy improved in retrospective reports compared to prospective ones, this improvement was nonsignificantly *smaller* than that of laypeople, $\chi^2(1, N = 1,782) = 0.84$, $p = .39$, Bayes factor = 83 in favor of the null.

Domain-Specific Expertise Was Not Linked to Greater Accuracy

Although we did not find any difference between the average scientist and the average layperson, it is possible that experts with extensive training in a specific topic might perform better. Since these are the scientists most likely to be consulted both by the media and public policymakers, it is important to know whether they provide more accurate estimates within their specific knowledge area. We thus

examined whether domain-specific expertise was associated with greater accuracy. To do this, we conducted regression analyses asking whether directional accuracy was significantly different when made within or outside a domain in which an expert self-reported having expertise or training (see online Supplemental Methods, for coding of domain-specific expertise in each study). Operationalizing expertise this way, we found no significant effect of expertise for either prospective predictions, $\chi^2(1, N = 659) = 0.06$, $p = .81$, or retrospective estimates, $\chi^2(1, N = 270) = 0.20$, $p = .66$. We also asked whether the degree of experience more generally (i.e., graduate student, postdoc, untenured, tenured faculty) mattered. Although we did observe an effect of experience on prediction accuracy in specific domains, $\chi^2(18, N = 581) = 31.54$, $p = .03$, this was largely driven by an advantage for *graduate students* over faculty in a small set of domains (see online Supplemental Results, for details). In other words, greater expertise did not seem to confer special ability to consistently and correctly predict outcomes.

Scientists Are Less Confident in Their Estimates Than Laypeople

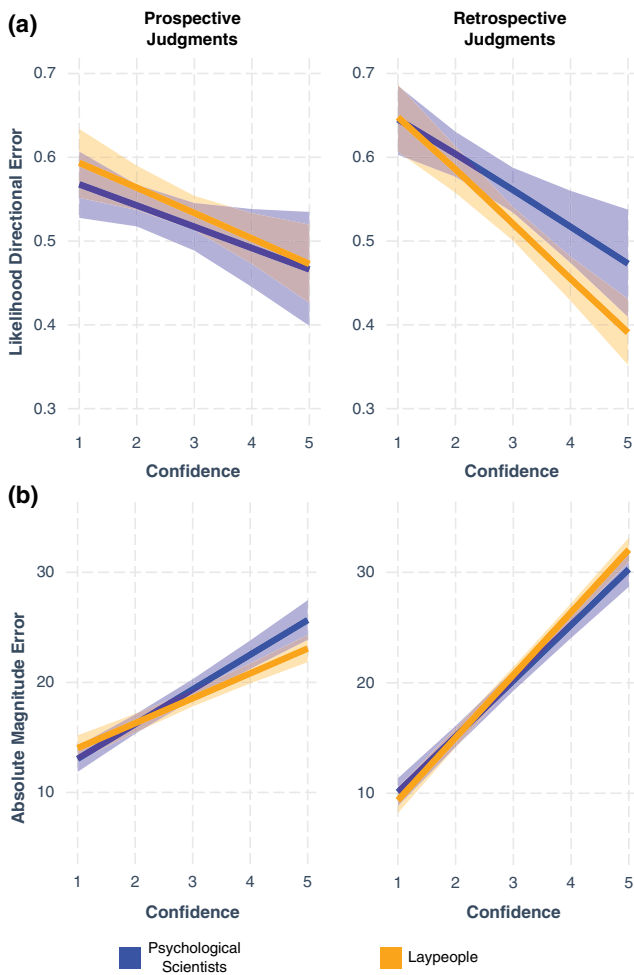
We did observe one notable difference between psychological scientists and lay Americans: scientists were consistently less confident in both their predictions, $5.14 < z_s \leq 9.86$, all $ps < .001$, and their retrospective estimates, $4.77 < z_s \leq 9.04$, $ps \leq .001$.² Greater confidence was simultaneously associated with a greater probability of correctly estimating the direction of societal change, $\chi^2(1, N = 1,387) = 6.02$, $p = .01$, but overestimating its *magnitude*, $\chi^2(1, N = 1,387) = 88.62$, $p < .001$. As Figure 3 shows, these effects were each magnified in retrospective compared to prospective estimates, such that confidence corresponded to fewer errors in directional inaccuracy, $\chi^2(1, N = 1,387) = 5.64$, $p = .02$, but larger errors of magnitude, $\chi^2(1, N = 1,387) = 110.29$, $p < .001$, in retrospective compared to prospective estimates. Thus, for both scientists and lay individuals, predictions made with greater confidence were more likely to get the direction of change correct, yet also to overestimate its magnitude.

Sources of Information When Making Judgments About Societal Change

To understand the kinds of information scientists and laypeople used in constructing their judgments, in Study 3, we asked participants whether they relied on vivid personal experiences and/or news reports when estimating societal change in the last 6 months (see online Supplemental Methods, for details). Both lay individuals and psychological scientists were more likely to report relying on news reports (scientists = 45% of judgments, lay individuals = 41% of

² The effect also held when controlling for political affiliation, ethnicity, age, gender, and income, $3.05 < z_s \leq 7.24$, $.002 < ps \leq .001$.

Figure 3
Confidence and Its Association With Inaccuracy



Note. Panel A: Relationship between confidence and likelihood of directional inaccuracy across domains, as a function of group (lay individuals or psychological scientists) and type of judgment (prospective predictions made in April/May 2020, retrospective estimates made in October/November of 2020). Panel B: Relationship between confidence and average absolute error (i.e., $|\text{estimated change} - \text{actual change}|$). In both panels, lines and error bars display the mixed-effects regression estimated line of best fit. See the online article for the color version of this figure.

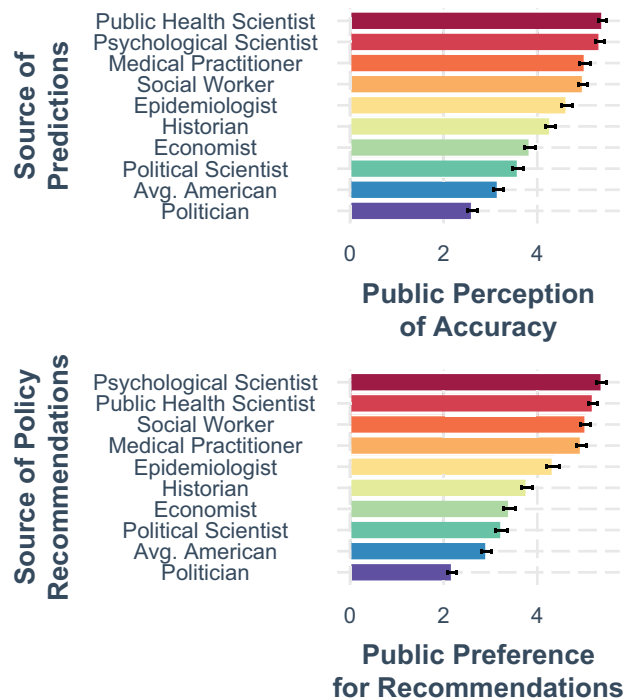
judgments) than personal experiences (scientists = 30% of judgments, lay individuals = 30% of judgments), scientists: $\chi^2(1, N = 270) = 121.32, p < .001$, lay individuals: $\chi^2(1, N = 411) = 78.03, p < .001$. This difference was somewhat larger among scientists, $\chi^2(1, N = 681) = 5.53, p = .02$, due to scientists reporting nonsignificant trends to rely more on news reports, $\chi^2(1, N = 681) = 2.91, p = .09$, and less on personal experience, $\chi^2(1, N = 681) = 0.92, p = .34$. Thus, consistent with the observation that psychological scientists and laypeople did not differ in their estimates or in the accuracy of their estimates, scientists and laypeople showed largely similar use of nonscientific sources of information. Intriguingly, relying on concrete personal experiences was

associated with greater directional accuracy, $\chi^2(1, N = 681) = 26.75, p < .001$, while relying on news articles was associated with lower accuracy, $\chi^2(1, N = 681) = 4.60, p = .03$. Nevertheless, effects on accuracy of considering personal experience and news were similar for scientists and lay individuals: personal experience, $\chi^2(1, N = 681) = 3.09, p = .08$; news, $\chi^2(1, N = 681) = 0.83, p = .36$.

Study 4: The Public’s Preference for Expert Judgments

One might argue that the accuracy of psychological scientists’ estimates of societal change matters chiefly if the public actually values such pronouncements, or would prefer to hear from psychologists as opposed to other sources (e.g., medical practitioners or politicians). We examined the latter hypothesis by surveying a sample of U.S. residents ($N = 203$) about who they expected to be most accurate in predicting societal trends in depression, well-being, violence, and related domains over the first half year of the crisis, and who they would most prefer to consult about how the COVID-19 pandemic would affect human behavior and society (see Method section, for details). As Figure 4 indicates, participants consistently

Figure 4
Comparative Perceptions of Psychological Scientists by the Public



Note. The lay public generally expects psychological scientists to be among the most accurate in predicting consequences of the pandemic for mental and social well-being and prefers to obtain policy recommendations for dealing with these issues from psychological scientists rather than experts in other topics like medicine, economics, or political science. See the online article for the color version of this figure.

ranked psychological scientists at the top of a list of different experts and practitioners, with economists and political scientists in the middle, and politicians ranked below even the average American. U.S. residents assumed psychological scientists would be significantly more accurate than most groups, $2.37 < ts \leq 20.93$, all $ps < .02$, false discovery rate (FDR)-corrected, and significantly more preferred to provide recommendations than most groups, $3.50 < ts \leq 22.47$, $ps < .001$, FDR-corrected, with the exception of public health, which evoked similar levels of preference (see [online Supplemental Materials](#), for further detail).

Discussion

Many psychological scientists were willing to comment publicly on the likely outcomes of the pandemic, justifying their analyses largely based on intuitive reasoning rather than a reference to theoretical or quantitative models (Study 1). Yet, such snap judgments about societal change in the wake of the COVID-19 pandemic were similar to those of laypeople (Studies 2a–c). The small improvements in accuracy that we observed in retrospective judgments were no larger than for laypeople (Studies 3a, b). Nor did we find that scientists with greater scientific training, higher career stage, or domain-specific expertise—that is, the individuals most likely to be consulted by both news outlets and policymakers—were more accurate. We also observed some evidence that the inaccuracy of judgments reported in the news might matter: among both scientists and lay individuals, those who reported relying on news reports when making judgments of change that had occurred over the first 6 months of the pandemic were significantly *less* accurate. These findings stand in contrast to the observation that psychological scientists are *believed* to be more accurate in predicting the pandemic's societal impacts compared to scientists in other disciplines, policymakers, or the lay public (Study 4). Although our conclusions are limited by the small number of specific domains assessed here, they nevertheless suggest that scientists may use their greater knowledge of research and theory to justify, rather than shape, the intuitions they share with the average person.³ This work, along with other forecasting tournaments among social scientists ([The Forecasting et al., 2023](#)), raises important questions about how to improve the accuracy of scientists' predictions regarding the societal effects of events like the COVID-19 pandemic.

Improving scientific accuracy requires some understanding of why psychological scientists were no more accurate than laypeople at predicting the pandemic's societal consequences. We propose two interrelated explanations. First, most psychological scientists have little training in prediction-oriented (as opposed to explanation-oriented) designs and models ([Hofman et al., 2017](#); [Yarkoni & Westfall, 2017](#)). The fact that not only the predicted direction but also the magnitude of societal change judgments aligned closely with those of the

general public supports this interpretation. That there were no major differences in accuracy between graduate students and tenured faculty further corroborates the absence of benefits for experience in psychological science on such judgments. This lack of attention to out-of-sample prediction may limit generalizability of existing psychological theories that experts may draw on to estimate effects in the real world ([Hofman et al., 2017](#); [Yarkoni & Westfall, 2017](#)). Second, although psychologists might be experts at making conditional statements about how and why pandemic-related behaviors might change *if* specific manipulations or policies were adopted ([Ruggeri et al., 2022](#)), formal psychological models of overall societal change in response to a once-in-a-century event like the pandemic are lacking ([Ackerman et al., 2021](#)). Without theory and necessary training to guide them, psychological scientists likely based their estimates on the same naïve theories of human social dynamics as laypeople ([Heider, 1958](#); [Kelly, 1955](#)). Indeed, their judgments were strikingly similar. Thus, we suspect that when scientists make intuitive judgments of the sort that we assessed here, and that appear in news media, they likely rely on exactly the same heuristics and reasoning as lay individuals.

One might also argue that perhaps expert predictions were inaccurate because policymakers heeded their cautionary advice and took actions that mitigated the negative outcomes that psychological scientists predicted. However, if this had been the case, then we likely should have observed greater differences between prospective and retrospective judgments (and reduced extremity of retrospective judgments), especially in domains like depression and subjective well-being, where policy responses might have had the greatest impact. Instead, we find that retrospective judgments in these particular domains are generally *more* inaccurate and extreme.

Our findings also suggest that the level of analysis at which psychologists generally excel (i.e., predicting the behavior of individuals or small groups) may not prepare them to provide judgments at a higher order level of analysis, namely when estimating societal change. This observation raises questions about how to improve both the accuracy and the utility of psychological scientists' expert judgments. At the least, minimal guidelines for assessing confidence in, and interpretation of, expert judgment may be beneficial ([IJzerman et al., 2020](#)). For example, in the present work, both expert and lay participants tended to predict more negative outcomes than actually unfolded, consistent with past research showing a negativity bias in predictions about the collective future ([Shrikanth et al., 2018](#); [Yamashiro & Roediger, 2019](#)). Keeping this tendency in mind might help both experts and policymakers correct for such biases when considering such predictions.

It is also worth noting that psychological scientists reported less confidence that their predictions would come to pass. Thus, even if they are similar to those of laypeople, there may

be benefits to considering psychological scientists' predictions in aggregate if they are weighted in some way to take their level of uncertainty into account, or if such expressions of uncertainty lead to more measured or contingent policy planning.

More broadly, the present findings suggest considerable room for improvement in psychological scientists' ability to predict real-world trends. Indeed, our work, along with prior endeavors, such as the Good Judgment Project, the World after COVID project (Grossmann et al., 2022) and the COVID-19 Tournament of The Forecasting et al., (2023), suggests that forecasting the future is difficult. To the extent that psychologists want to make predictions for such events—which our work shows they seem willing to do and are expected to do so well by the public—then it may be advantageous for psychological scientists to learn strategies that improve forecasting accuracy at both the group (Morgan, 2014) and individual levels (Grossmann et al., 2021; Mellers et al., 2019).

³ It is possible that psychologists may have been more accurate if a larger or different set of domains had been chosen. In the present work, we were limited to a relatively small set of domains by four factors: (a) we only selected domains where there was prior reason to anticipate substantial change as a result of the pandemic, (b) we only selected domains in which psychological scientists were likely to have some knowledge or expertise, (c) we only selected domains with high-quality national data sets, and (d) we could only obtain a limited set of predictions from the psychological scientists participating in our study, who were volunteers with limited time (a pragmatic concern that was especially relevant during the initial pandemic lockdowns). We do not see a clear reason why such judgments would have been more accurate for a different or larger suite of domains, but we remain open to the possibility and hope that others might consider such an ambitious undertaking with more domains in the future.

References

References marked with an asterisk indicate references cited only in the online Supplemental Materials.

- Ackerman, J. M., Tybur, J. M., & Blackwell, A. D. (2021). What role does pathogen-avoidance psychology play in pandemics? *Trends in Cognitive Sciences*, 25(3), 177–186. <https://doi.org/10.1016/j.tics.2020.11.008>
- Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. N., Lampropoulos, G. K., Walker, B. S., Cohen, G., & Rush, J. D. (2006). The meta-analysis of Clinical Judgment Project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist*, 34(3), 341–382. <https://doi.org/10.1177/0011000005285875>
- Armstrong, J. S. (1985). *Long-range forecasting: From crystal ball to computer*. Wiley.
- Artazcoz, L., Benach, J., Borrell, C., & Cortès, I. (2004). Unemployment and mental health: Understanding the interactions among gender, family roles, and social class. *American Journal of Public Health*, 94(1), 82–88. <https://doi.org/10.2105/AJPH.94.1.82>
- Boxell, L., Conway, J., Druckman, J. N., & Gentzkow, M. (2020). *Affective polarization did not increase during the coronavirus pandemic*. National Bureau of Economic Research. <https://doi.org/10.3386/w28036>
- Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: Rapid review of the evidence. *Lancet*, 395(10227), 912–920. [https://doi.org/10.1016/S0140-6736\(20\)30460-8](https://doi.org/10.1016/S0140-6736(20)30460-8)
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243(4899), 1668–1674. <https://doi.org/10.1126/science.2648573>
- *Dunwoody, P. T., & Funke, F. (2016). The Aggression-Submission-Conventionalism Scale: Testing a new three factor measure of authoritarianism. *Journal of Social and Political Psychology*, 4(2), 571–600. <https://doi.org/10.5964/jpspp.v4i2.168>
- Faulkner, J., Schaller, M., Park, J. H., & Duncan, L. A. (2004). Evolved disease-avoidance mechanisms and contemporary xenophobic attitudes. *Group Processes & Intergroup Relations*, 7(4), 333–353. <https://doi.org/10.1177/1368430204046142>
- Fincher, C. L., Thornhill, R., Murray, D. R., & Schaller, M. (2008). Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism. *Proceedings of the Royal Society B: Biological Sciences*, 275(1640), 1279–1285. <https://doi.org/10.1098/rspb.2008.0094>
- Fong, G. T., & Nisbett, R. E. (1991). Immediate and delayed transfer of training effects in statistical reasoning. *Journal of Experimental Psychology: General*, 120(1), 34–45. <https://doi.org/10.1037/0096-3445.120.1.34>
- Gardner, R. M., & Dalsing, S. (1986). Misconceptions about psychology among college students. *Teaching of Psychology*, 13(1), 32–34. https://doi.org/10.1207/s15328023top1301_9
- Gardner, R. M., & Hund, R. M. (1983). Misconceptions of psychology among academicians. *Teaching of Psychology*, 10(1), 20–22. https://doi.org/10.1207/s15328023top1001_5
- *Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85(2), 197–216. <https://doi.org/10.1037/0022-3514.85.2.197>
- Griskevicius, V., Tybur, J. M., Delton, A. W., & Robertson, T. E. (2011). The influence of mortality and socioeconomic status on risk and delayed rewards: A life history theory approach. *Journal of Personality and Social Psychology*, 100(6), 1015–1026. <https://doi.org/10.1037/a0022403>
- Grossmann, I., Dorfman, A., Oakes, H., Santos, H. C., Vohs, K. D., & Scholer, A. A. (2021). Training for wisdom: The distanced self-reflection diary method. *Psychological Science*, 32(3), 381–394. <https://doi.org/10.1177/0956797620969170>
- Grossmann, I., Twardus, O., Varnum, M. E. W., Jayawickreme, E., & McLevey, J. (2022). Expert predictions of societal change: Insights from the world after COVID project. *American Psychologist*, 77(2), 276–290. <https://doi.org/10.1037/amp0000903>
- Hawkey, L. C., & Cacioppo, J. T. (2010). Loneliness matters: A theoretical and empirical review of consequences and mechanisms. *Annals of Behavioral Medicine*, 40(2), 218–227. <https://doi.org/10.1007/s12160-010-9210-8>
- *Hehman, E., Calanchini, J., Flake, J. K., & Leitner, J. B. (2019). Establishing construct validity evidence for regional measures of explicit and implicit racial bias. *Journal of Experimental Psychology: General*, 148(6), 1022–1040. <https://doi.org/10.1037/xge0000623>
- Heider, F. (1958). The naive analysis of action. In F. Heider (Ed.), *The psychology of interpersonal relationships* (pp. 79–124). Wiley. <https://doi.org/10.1037/10628-004>
- Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. *Science*, 355(6324), 486–488. <https://doi.org/10.1126/science.aal3856>
- Horn, H. S. (1978). Optimal tactics of reproduction and life-history. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural ecology: An evolutionary approach* (pp. 411–429). Blackwell.
- Hutcherson, C. (2022, September 7). *Retrospective and prospective estimates of the impact of COVID-19*. <https://doi.org/10.17605/OSF.IO/ZXAVD>
- Hutcherson, C., Grossmann, I., Sharpinskyi, C., Wormley, A., & Varnum, M. E. W. (2023, February 5). *Retrospective and prospective estimates of the impact of COVID-19*. <https://osf.io/9btsy>

- Ijzerman, H., Lewis, N. A., Jr., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., Vazire, S., Forscher, P. S., Morey, R. D., Ivory, J. D., & Anvari, F. (2020). Use caution when applying behavioural science to policy. *Nature Human Behaviour*, 4(11), 1092–1094. <https://doi.org/10.1038/s41562-020-00990-w>
- *Jaidka, K., Giorgi, S., Schwartz, H. A., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2020). Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences of the United States of America*, 117(19), 10165–10171. <https://doi.org/10.1073/pnas.1906364117>
- Kelly, G. A. (1955). *The psychology of personal constructs. Volume 1: A theory of personality*. W.W. Norton.
- Kendler, K. S., Karkowski, L. M., & Prescott, C. A. (1999). Causal relationship between stressful life events and the onset of major depression. *The American Journal of Psychiatry*, 156(6), 837–841. <https://doi.org/10.1176/ajp.156.6.837>
- *Kim, H. S., Sherman, D. K., & Updegraff, J. A. (2016). Fear of Ebola: The influence of collectivism on xenophobic threat responses. *Psychological Science*, 27(7), 935–944. <https://doi.org/10.1177/0956797616642596>
- *Leitner, J. B., Hehman, E., Ayduk, O., & Mendoza-Denton, R. (2016). Blacks' death rate due to circulatory diseases is positively related to whites' explicit racial bias: A nationwide investigation using project implicit. *Psychological Science*, 27(10), 1299–1311. <https://doi.org/10.1177/0956797616658450>
- Mellers, B., Stone, E., Murray, T., Minster, A., Rohrbaugh, N., Bishop, M., Chen, E., Baker, J., Hou, Y., Horowitz, M., Ungar, L., & Tetlock, P. (2015). Identifying and cultivating superforecasters as a method of improving probabilistic predictions. *Perspectives on Psychological Science*, 10(3), 267–281. <https://doi.org/10.1177/1745691615577794>
- Mellers, B., Tetlock, P., & Arkes, H. R. (2019). Forecasting tournaments, epistemic humility and attitude depolarization. *Cognition*, 188, 19–26. <https://doi.org/10.1016/j.cognition.2018.10.021>
- Mercer, A. W. (2018). *Selection bias in nonprobability surveys: A causal inference approach*. University of Maryland.
- Morgan, M. G. (2014). Use (and abuse) of expert elicitation in support of decision making for public policy. *Proceedings of the National Academy of Sciences of the United States of America*, 111(20), 7176–7184. <https://doi.org/10.1073/pnas.1319946111>
- Murray, D. R., Trudeau, R., & Schaller, M. (2011). On the origins of cultural differences in conformity: Four tests of the pathogen prevalence hypothesis. *Personality and Social Psychology Bulletin*, 37(3), 318–329. <https://doi.org/10.1177/0146167210394451>
- Na, J., Grossmann, I., Varnum, M. E. W., Kitayama, S., Gonzalez, R., & Nisbett, R. E. (2010). Cultural differences are not always reducible to individual differences. *Proceedings of the National Academy of Sciences of the United States of America*, 107(14), 6192–6197. <https://doi.org/10.1073/pnas.1001911107>
- *Neuberg, S. L., Varnum, M. E. W., Becker, D. v., Ko, A., Pick, C. M. P., & Wormley, A. S. (2023). *Changes in attitudes, self-construal, and prejudice during the COVID-19 pandemic* [Unpublished raw data].
- Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science*, 238(4827), 625–631. <https://doi.org/10.1126/science.3672116>
- *Ofosu, E. K., Chambers, M. K., Chen, J. M., & Hehman, E. (2019). Same-sex marriage legalization associated with reduced implicit and explicit antigay bias. *Proceedings of the National Academy of Sciences of the United States of America*, 116(18), 8846–8851. <https://doi.org/10.1073/pnas.1806000116>
- Piantadosi, S., Byar, D. P., & Green, S. B. (1988). The ecological fallacy. *American Journal of Epidemiology*, 127(5), 893–904. <https://doi.org/10.1093/oxfordjournals.aje.a114892>
- Rosenfeld, R., & Lopez, E. (2020). *Pandemic, social unrest, and crime in US cities*. Council on Criminal Justice.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2), 225–237. <https://doi.org/10.3758/PBR.16.2.225>
- Ruggeri, K., Stock, F., Haslam, S. A., Capraro, V., Boggio, P., Ellemers, N., Cichocka, A., Douglas, K., Rand, D. G., & Cikara, M. (2022). *Evaluating expectations from social and behavioral science about COVID-19 and lessons for the next pandemic*. <https://doi.org/10.31234/osf.io/58udn>
- Schaller, M., & Park, J. H. (2011). The behavioral immune system (and why it matters). *Current Directions in Psychological Science*, 20(2), 99–103. <https://doi.org/10.1177/0963721411402596>
- *Schwartz, H., Eichstaedt, J., Kern, M., Dziurzynski, L., Lucas, R., Agrawal, M., Park, G., Lakshmikanth, S., Jha, S., & Seligman, M. (2021). Characterizing geographic variation in well-being using tweets. *Proceedings of the International AAAI Conference on Web and Social Media*, 7(1), 583–591. <https://doi.org/10.1609/icwsm.v7i1.14442>
- Sharpinskyi, C., Rotella, A. M., & Grossmann, I. (2022, September 5). *Behavioral and social scientists and laypeople make similar predictions for COVID-19's societal consequences*. <https://osf.io/npzcr>
- *Sherman, S. J., Cialdini, R. B., Schwartzman, D. F., & Reynolds, K. D. (1985). Imagining can heighten or lower the perceived likelihood of contracting a disease: The mediating effect of ease of imagery. *Personality and Social Psychology Bulletin*, 11(1), 118–127. <https://doi.org/10.1177/0146167285111011>
- Shrikanth, S., Szpunar, P. M., & Szpunar, K. K. (2018). Staying positive in a dystopian future: A novel dissociation between personal and collective cognition. *Journal of Experimental Psychology: General*, 147(8), 1200–1210. <https://doi.org/10.1037/xge0000421>
- Tausanovitch, C., Vavreck, L., Reny, T., Hayes, A. R., & Rudkin, A. (2019). *Democracy fund + UCLA nationscape methodology and representativeness assessment*. Democracy Fund Voter Study Group.
- Taylor, A. K., & Kowalski, P. (2004). Naïve psychological science: The prevalence, strength, and sources of misconceptions. *The Psychological Record*, 54(1), 15–25. <https://doi.org/10.1007/BF03395459>
- Tetlock, P. E. (2005). *Expert political judgement: How good is it?* Princeton University Press.
- The Forecasting Collaborative, Grossmann, I., Rotella, A., Hutcherson, C. A., Sharpinskyi, K., Varnum, M. E. W., Achter, S., Dhami, M. K., Guo, X. E., Kara-Yakoubian, M., Mandel, D. R., Raes, L., Tay, L., Vie, A., Wagner, L., Adamkovic, M., Arami, A., Arriaga, P., Bandara, K., ... Wilkening, T. (2023). Insights into the accuracy of social scientists' forecasts of societal change. *Nature Human Behaviour*. Advance online publication. <https://doi.org/10.1038/s41562-022-01517-1>
- *Trobe, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, 117(2), 440–463. <https://doi.org/10.1037/a0018963>
- Tybur, J. M., Inbar, Y., Aarøe, L., Barclay, P., Barlow, F. K., de Barra, M., Becker, D. V., Borovoi, L., Choi, I., Choi, J. A., Consedine, N. S., Conway, A., Conway, J. R., Conway, P., Adoric, V. C., Demirci, D. E., Fernández, A. M., Ferreira, D. C., Ishii, K., ... Žeželj, I. (2016). Parasite stress and pathogen avoidance relate to distinct dimensions of political ideology across 30 nations. *Proceedings of the National Academy of Sciences of the United States of America*, 113(44), 12408–12413. <https://doi.org/10.1073/pnas.1607398113>
- van Bavel, J. J. V., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M. J., Crum, A. J., Douglas, K. M., Druckman, J. N., Drury, J., Dube, O., Ellemers, N., Finkel, E. J., Fowler, J. H., Gelfand, M., Han, S., Haslam, S. A., Jetten, J., ... Willer, R. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, 4(5), 460–471. <https://doi.org/10.1038/s41562-020-0884-z>
- Watts, D. J. (2017). Should social science be more solution-oriented? *Nature Human Behaviour*, 1(1), 1–5. <https://doi.org/10.1038/s41562-016-0015>

West, R., Michie, S., Rubin, G. J., & Amlôt, R. (2020). Applying principles of behaviour change to reduce SARS-CoV-2 transmission. *Nature Human Behaviour*, 4(5), 451–459. <https://doi.org/10.1038/s41562-020-0887-9>

*Xu, K., Nosek, B. A., Greenwald, A. G., Ratliff, K. A., Bar-Anan, Y., Umansky, E., Banaji, M. R., Lofaro, N., & Smith, C. (2018, December 21). *Project implicit demo website datasets*. <https://doi.org/10.17605/OSF.IO/T4BNJ>

Yamashiro, J. K., & Roediger, H. L., III. (2019). How we have fallen: Implicit trajectories in collective temporal thought. *Memory*, 27(8), 1158–1166. <https://doi.org/10.1080/09658211.2019.1635161>

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12(6), 1100–1122. <https://doi.org/10.1177/1745691617693393>

Received September 7, 2022
 Revision received January 29, 2023
 Accepted February 2, 2023 ■

UNITED STATES POSTAL SERVICE® Statement of Ownership, Management, and Circulation
 (All Periodicals Publications Except Requester Publications)

American Psychologist Publication Title
 Issue Date: October 2023
 Frequency: 9 times a year
 Issue Number: 9
 Annual Subscription Price: \$12.00
 ISSN: 0003-0651
 USPS PSN: 862-8878
 Publication Title: American Psychologist
 Issue Date: October 2023
 Frequency: 9 times a year
 Issue Number: 9
 Annual Subscription Price: \$12.00
 ISSN: 0003-0651
 USPS PSN: 862-8878

750 First Street, NE, Washington, DC 20002-4242
 Complete Mailing Address of Headquarters or General Business Office of Publisher (Not printer)
 Telephone (include area code): (202) 414-8076
 750 First Street, NE, Washington, DC 20002-4242
 Complete Mailing Address of Publisher, Editor, and Managing Editor (Do not leave blank)
 Publisher Name and complete mailing address:
 American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242
 Managing Editor Name and complete mailing address:
 Harris Cooper, Duke University, Box 90086, Durham, NC 27708-0086
 Rose H. Solot, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242
 Complete Mailing Address:
 American Psychological Association
 750 First Street, NE
 Washington, DC 20002-4242

UNITED STATES POSTAL SERVICE® Statement of Ownership, Management, and Circulation
 (All Periodicals Publications Except Requester Publications)

American Psychologist Publication Title
 Issue Date: September 2023
 Frequency: 9 times a year
 Issue Number: 9
 Annual Subscription Price: \$12.00
 ISSN: 0003-0651
 USPS PSN: 862-8878

750 First Street, NE, Washington, DC 20002-4242
 Complete Mailing Address of Headquarters or General Business Office of Publisher (Not printer)
 Telephone (include area code): (202) 414-8076
 750 First Street, NE, Washington, DC 20002-4242
 Complete Mailing Address of Publisher, Editor, and Managing Editor (Do not leave blank)
 Publisher Name and complete mailing address:
 American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242
 Managing Editor Name and complete mailing address:
 Harris Cooper, Duke University, Box 90086, Durham, NC 27708-0086
 Rose H. Solot, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242
 Complete Mailing Address:
 American Psychological Association
 750 First Street, NE
 Washington, DC 20002-4242

UNITED STATES POSTAL SERVICE® Statement of Ownership, Management, and Circulation
 (All Periodicals Publications Except Requester Publications)

American Psychologist Publication Title
 Issue Date: September 2023
 Frequency: 9 times a year
 Issue Number: 9
 Annual Subscription Price: \$12.00
 ISSN: 0003-0651
 USPS PSN: 862-8878

750 First Street, NE, Washington, DC 20002-4242
 Complete Mailing Address of Headquarters or General Business Office of Publisher (Not printer)
 Telephone (include area code): (202) 414-8076
 750 First Street, NE, Washington, DC 20002-4242
 Complete Mailing Address of Publisher, Editor, and Managing Editor (Do not leave blank)
 Publisher Name and complete mailing address:
 American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242
 Managing Editor Name and complete mailing address:
 Harris Cooper, Duke University, Box 90086, Durham, NC 27708-0086
 Rose H. Solot, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242
 Complete Mailing Address:
 American Psychological Association
 750 First Street, NE
 Washington, DC 20002-4242

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.