**The pandemic fallacy: Inaccuracy of social scientists’ and lay judgments about COVID-19’s societal consequences in America**

Cendri A. Hutcherson1,2, Konstantyn Sharpinskyi3, Michael Varnum4, Amanda Rotella3, Alexandra Wormley4, Louis Tay5, & Igor Grossmann3

1 Department of Psychology, University of Toronto Scarborough, Toronto, ON M1C 1A4, Canada

2 Department of Marketing, Rotman School of Management, University of Toronto, Toronto, ON M5S 3E6, Canada

3 Department of Psychology, University of Waterloo, Waterloo, ON, N2L 3G1, Canada

4 Department of Psychology, Arizona State University, Tempe, AZ, 85287

5 Department of Psychological Sciences, Purdue University, West Lafayette, IN, 47907

**Author Note**

Correspondence should be addressed to Igor Grossmann, PAS 3047, University of Waterloo, Waterloo, ON N2L 3G1, [igrossma@uwaterloo.ca](mailto:igrossma@uwaterloo.ca). This work was supported by a Social Sciences and Humanities Research Council of Canada Insight Grant 435-2016-1274 (to C.H.), a National Science Foundation grant 2026854 (to M.E.W.V.), 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 I.G.).

**Author contributions**: C. H. and I.G. developed the study concept. C.H., K. S., M.E.W.V., A.R, A.W. and I.G. designed the research, A.R., K.S. and I.G. collected the data. K.S. and I.G. analyzed the data. A.W. cross-validated the code and data analysis. C.H., K.S., M. E.W.V. and I.G. drafted the manuscript. L. T. provided critical feedback on data analyses. All authors read and provided feedback on the manuscript.

**Abstract**

Effective management of global crises relies on expert judgment of their societal effects. How accurate are such judgments? In the spring of 2020, we asked behavioral and social scientists (*N* = 717) and lay Americans (*N* = 394) to make predictions about COVID-19 pandemic-related societal change across social and psychological domains. Six months later we obtained retrospectiveassessments for the same domains (*N*scientists = 270; *N*layP = 411) and compared these judgments to objective data to assess estimation accuracy. Scientists were no more accurate than lay people, neither in prospective nor retrospective judgments. Across studies and samples, estimates of the magnitude of change were off by more than 20% and less than half of participants accurately predicted the direction of changes. Taken together, we find that experts and lay people fared poorly at predicting social and psychological consequences of the pandemic and misperceive what effects it may have already had.

*Keywords*: COVID-19; forecasting; scientific intuitions; cultural change, lay theories of change

**The pandemic fallacy: Inaccuracy of social scientists’ and lay judgments about COVID-19’s societal consequences in America**

Social scientists[[1]](#footnote-2) have offered many suggestions about the impact of the COVID-19 pandemic on human psychology, as well as policy recommendations aimed at countering its damaging consequences (*1*–*4*). These efforts are not ad-hoc, relying on theories and research regarding the effects of social isolation (*5*) or the threat of infectious disease (*6*–*10*). The value of these efforts rests on the premise that social scientists’ expertise grants them an enhanced ability to recognize the societal consequences of the current crisis, relative to non-experts—a thesis which so far has not been tested.

A growing body of scholarship raises doubt about expert judgment of political (*11*), economic (*12*), and career-related outcomes (*13*, *14*): in these domains, experts are rarely more accurate than simple statistical models or even when comparing to non-human primates (*11*). Social scientists tend to emphasize causal explanation over out-of-sample prediction (*15*). As a result, they often favor complex models that may be overly sensitive to the particularities of one’s sample or context and not generalizable to new scenarios (*16*). When facing an uncertain global event (*17*) with scant directly applicable research, social scientists may fall back on the same kinds of error-prone and intuition-driven heuristics(*18*–*20*) and biases that shape the reasoning of lay people (*21*).

At the same time, there are reasons to expect greater accuracy among social science experts compared to lay people or to chimps. Although prior research has found that expert judgment often comes up short (*11*), this evidence comes from predictions about the likelihood of singular geopolitical events, which may be inherently unpredictable and subject to chaos-theory type “butterfly effects” (*22*). Additionally, social science training and expertise improves understanding of probability and statistics (*23*, *24*), and reduces mistaken assumptions about human behavior (*25*–*27*)—qualities that tend to increase the accuracy of forecasts in non-psychological domains (*28*). Especially when considering aggregate psychological responses (which reduce the unpredictability of individual variation), the existence of empirically-grounded theories about human responses to things like social isolation (*5*), financial uncertainty (*29*), and disease threat (*6*) suggest that social scientists might be well-positioned to estimate the magnitude, rank-order, or at least the direction of changes in psychology and behavior in response to the pandemic.

While some research has examined lay predictions about slow-moving social, technological and economic trends (*30*), no work to date has provided a systematic evaluation of the ability of social scientists to estimate societal-level psychological change in response to an acute crisis. Nor has prior work compared such estimates to naïve reasoning. Here, we fill this gap, using the initial six months of the COVID-19 pandemic as a naturalistic experiment. We present the first systematic investigation into the accuracy of social scientists compared to lay people in both predicting and retrospectively evaluating aggregate-level changes in human psychology in response to a global crisis. To measure accuracy, we compared estimates to objective indicators of change across ten domains. We also assessed the extent to which individual differences in confidence (*31*, *32*) and construal level (i.e., concrete vs. abstract reasoning) (*33*)might be associated with the accuracy of such estimates.

**Selecting domains of social and psychological change in response to the pandemic**

We considered a subset of specific domains for which a robust body of theoretical and empirical work links these variables to pathogen-related threats. Based on theories that suggest that intergroup processes are affected by evolutionary and ontogenetic pressures related to pathogen stress (*6*–*10*), we examined judgments of prejudice, 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 in response to environmental threat and pathogen-related unpredictability (*34*, *35*), we assessed judgments of delay of gratification and birth rates. Finally, based on theories about how human mental and affective well-being is influenced by stressors (*36*), including social isolation (*5*), we assessed judgments of depression, loneliness, and life satisfaction.

**Research Overview**

We expected beliefs about societal change to vary by country (*37*, *38*). To standardize the target of change, we asked participants to focus their predictions on a single country – the United States. To assess temporal stability and reliability of judgments, we obtained one sample of social scientists in early April (Study 1; *N* = 401) and another sample of social scientists (*N* = 316), collected in parallel with a nationally representative sample of Americans (*N* = 394), in late April/early May of 2020 (Study 2). Next, to assess whether lived experience alters assessments of change, we conducted a second set of pre-registered surveys (osf.io/9btsy) in new samples of social scientists (*N* = 270) and a nationally representative sample of Americans (*N* = 411) in late October/early November, just before to the US election (Study 3). In these surveys, we asked participants to estimate how much change had occurred in each domain over the last six months. To assess the accuracy of prospective and retrospective judgments, we compared judgments to actual change in a subset of 10 domains for which we could obtain high-quality, national-level data: depression, life satisfaction, generalized trust, loneliness, individualism, traditionalism, political polarization, climate change attitudes, violent crimes, charitable giving (see also estimates for explicit and implicit prejudice in the online supplement).

**Results**

As Figure 1 shows, we found little evidence that predictions of societal change were accurate on average, with the exception of three domains: predictions regarding charitable giving, individualism, climate change and traditionalism showed evidence of accuracy (also see on-line supplement for statistical tests). Retrospective estimates of change were similarly inaccurate, except for climate change beliefs (see Table S20 in the online supplement). However, even for climate change beliefs social scientists were largely inaccurate in estimating the direction of change (see Figure 2). Beyond these few domains, as Figures 1-2 show, estimates were on average strikingly inaccurate compared to objective markers. This inaccuracy was not due to participants confusing the starting point of change in April/May 2020 with the change happening since the beginning of the COVID-19 pandemic in January 2020: Comparisons of objective change for five domains in which we obtained societal estimates for January and October 2020 also revealed significant differences between predictions and objective estimates for each domain except for lay estimates of traditionalism (see Table S21).

Figure 2 quantifies the percentage of accurate responses for each sample, using both strict and more liberal percentage-difference cutoffs as a measure of accuracy. Using a strict criterion (within 1% point of the estimate), in most domains, less than 2 % of each sample were accurate in their forecasts as well as retrospective estimates, with somewhat better forecasts for traditionalism, life satisfaction, generalized trust and depression rates. Using a moderate criterion (within 5% of the estimate), for most domains, less than 10% of each sample was accurate, except for traditionalism, depression, climate change beliefs, generalized trust and charity (for scientists).

Predictions in the time of societal uncertainty may be influenced by a number of conditional factors that only become known in retrospect (e.g., whether or not governments enact fiscal stimulus, mandate or simply encourage social distancing). Do scientists show greater accuracy for retrospective assessment of societal change? Our results suggest that it is not the case. Retrospective estimates showed a similar if not smaller number of accurate estimates. Even when using a liberal criterion (within 20% of the estimate), for most domains, less than 41% of each sample was accurate, and for no domain did we observe a meaningful majority of participants being accurate (60+%). Notably, Figure 2 demonstrates that numbers of accurate estimates were very similar between social scientists and lay people, with most differences within a negligible rate of 5% difference (with the exception of retrospective estimates of depression and climate change beliefs when using moderate and liberal criteria).

So far, we report accuracy separately by domain. It is possible that social scientists estimate domains in relation to each other—after all, social scientists often study how social phenomena or processes are associated with each other and make conditional inferences. Thus, they could be more accurate when judging the rank order of most positive to most negative societal change across domains. To address this question, for each participant we calculated the rank-order correlation ρ between their estimates and objective markers for the same 10 domains. Here, ρ represents the degree of accuracy in estimated compared to objective rank order. To assess significance, we constructed a null distribution of the expected rank-order correlation using 5000 random permutations of the observed outcomes. As Figure 3 shows, social scientists and lay people alike had average rank-order correlations .05 < ρ ≤ .08 that were not significantly different from chance. Permutation tests with random shuffling of domain labels suggest that this degree of correlation is not significantly different from chance. Rank-order accuracy did not vary by sample (social scientists vs. lay people), or judgment type (prospective vs. retrospective), *p*s > .594.

**Accuracy of the direction of societal change**

Our initial analyses suggest that social scientists, and lay people alike, were largely inaccurate in judging the societal effects of the pandemic. One possible rejoinder to this finding is that the magnitude of change may be subject to unpredictable fluctuations and policies implemented over the course of the pandemic. It might be unreasonable to expect precise estimates of change to be accurate. Therefore, we tested whether accuracy improved when examining the direction of predicted change. In comparison to magnitudes, when examining the overall direction of trends, by May 2020 it was reasonable to expect that certain fundamentals will be hard to avoid irrespective of unpredictable policy responses. For instance, social relationships will with high likelihood be more circumscribed or virtual, some people will lose their jobs, and mortality rates will be elevated. Given these plausible consequences on the direction of movement, we sought to explore the accuracy of direction of societal change, estimating it in terms of the percentage of individuals correctly predicting/estimating the direction of changes.

As Figure 2 shows, most scientists correctly predicted the direction of change for charitable giving (60%), political polarization (73%), generalized trust (60%) and depression (89%). However, considering all domains together, most estimates were inaccurate: Results of a generalized linear mixed model with accuracy scores (1 = hit / 0 = miss) as a dependent variable and domains nested in participants revealed accuracy levels of 1% above chance. . Overall accuracy rates of social scientists were not significantly different from those of lay people, χ2 (*df* = 1) = 0.51, *p* = .475. Moreover, estimates did not improve with experience. As supplementary results show, in domains where most predictions were inaccurate, even larger numbers of retrospective assessments were directionally inaccurate, χ2 (*df* = 1) = 150.58, *p* < .001 (see Fig. S16).

**Comparison of expert and lay judgments**

So far, we have focused on comparisons of estimates against accuracy markers for the first six months of societal change. In the next step, we sought to more extensively characterize and compare predictions of expert and lay people. To do this, we examined a wider temporal range of estimates provided by experts and lay people, including prediction for not only six months, but also for a year and two years from April/May 2020. Expert judgments did not significantly vary by degree of expertise or cultural distance from the US , and remained largely identical over time (April vs. May 2020)[[2]](#footnote-3). Moreover, social scientists’ judgments were very similar in trend and slope to judgments made by lay people. For predictions, we observed significant moderation by sample type for only a handful of domains (linear trends for prejudice,generalized trust, and political polarization; see Table S13), Bayes Factor < 2.73. The overall lack of differences between groups remained robust when controlling for key demographic factors (age, gender, household income, and political orientation, see Figure S10). Moreover, lay people’s predictions showed a similar level of agreement to that observed among our scientist sample, with 95 out of 100 predictions falling within 3.9% points of the mean.

When we compared restrospective estimates of societal change, the picture was very similar to predictions.As Figure 1 indicates, retrospective estimates of societal change were on average similar among social scientists and lay people. Bayesian analyses revealed very strong/strong Bayes Factors > 13.5 in support of the null hypothesis for 7 out of 15 domains, and moderate Bayes Factor > 4.2 in support of the null for another 3 domains. Only delay of gratification (BF = .26) and violent crime rates (BF < .01) showed moderate/strong support of the alternative hypothesis (see Table S15). As with prospective judgments, these differences remained robust when controlling for socio-demographic characteristics (see Figure S15). Similarly, results were not significantly different when comparing social scientists who reported domain-specific expertise to those who did not report such expertise (see Figure S11 and Table S16).

**Prospective vs. retrospective judgments**

In pre-registered analyses, we compared prospective predictions with retrospective estimates. Based on prior literature on forecasting biases (*39*), we predicted that prospective predictions would be more extreme in terms of absolute percentage change from baseline, compared to retrospective estimates, for both expert and lay samples. As Figure 1 shows (also see Figures S13-S14, and Table S20 in the online supplement), social scientists reported significantly less extreme retrospective (vs. prospective) estimates for explicit and implicit prejudice toward minorities, climate change, charity, and religiosity, consistent with this pre-registered hypothesis. However, for all other domains social scientists either showed a statistically significant reversal (i.e., seven domains in which retrospective estimates were greater than prospective predictions: life satisfaction, loneliness, generalized trust, political polarization, delay of gratification, depression rates, and individualism) or no significant differences between prospective and retrospective estimates.

This unexpected result was even more salient among the lay participants, who showed the predicted pattern (prospective > retrospective) for only two domains – birthrates and charity. For eight domains, lay participants reported significantly more extreme retrospective (vs. prospective) estimates, and for other domains we observed no significant differences between forecasts and retrospective assessments (see Figure 1). Thus, on average, retrospective estimates were either similar or more extreme than prospective estimates[[3]](#footnote-4).

**Abstact vs. concrete contrual**

To unpack this unexpected pattern of results, we performed a series of pre-registered analyses involving comparisons of estimates for domains where participants reported recalling *concrete* news and personal experiences, and domains for which participants reported recalling *vivid memories*. We observed significantly more extreme estimates among lay participants recalling concrete (10 out of 15 domains) and vivid experiences (11 out of 15 domains), with no significant differences for the remaining domains. Social scientists showed a similar tendency (see Figure S12), albeit with smaller effect sizes, and not significant for as many domains after false discovery rate correction (concreteness: 6/15; vividness: 7/15; see Table S17). Overall, retrospective estimates for which participants recalled concrete or vivid experiences were more extreme, suggesting that experience accentuated (rather than attenuated) estimates of societal change.

**Confidence in estimates**

Despite the striking similarities between the average predictions of social scientists and lay people, social scientists were consistently significantly *less* confident in their predictions than lay people (see Figure 4), 5.14 < *z*s ≤ 9.86, all *ps* < .001, with differences between the two groups ranging from 0.40 to 0.73 *SD* of the mean. Results were similar in pre-registered analyses for retrospective estimates, 4.77 < *z*s ≤ 9.04, *ps* ≤ .001, with differences between the two groups ranging from 0.40 to 0.70 *SD* of the mean[[4]](#footnote-5). Notably, as Figure 4 also shows, greater confidence was significantly associated with lower accuracy, χ2 (*df* =1) = 91.59, *p* < .001. The latter observation was consistently present for each sample. Further, we observed a significant confidence × survey type (prospective vs. retrospective) interaction, χ2 (*df* =1) = 118.52, *p* < .001, with stronger inverse association between accuracy and retrospective compared to prospective estimates, social scientists: *z* = 4.91, *p* < .001; lay people: *z* = 10.89, *p* < .001. Supplementary analyses also show that the negative association between confidence and accuracy was exacerbated for domains where participants relied on vivid personal experiences, and that it was chiefly due to the positive association between confidence and extremity of one’s estimates. Because societal changes were by far smaller (or even different in direction) than predicted by the majority of participants, greater extremity produced greater inaccuracy.

**Discussion**

The present work suggests good reason to doubt social scientists’ initial judgments about societal consequences of the COVID-19 pandemic. Social scientists were no more accurate than the lay people, for either prospective predictions or estimates of change that had already occurred. Nor were the cues we might naively use to judge whether an expert is likely to be accurate, such as their educational certifications or domain-specific expertise, indicative of greater accuracy. Even worse, confidence in one’s predictions and greater reliance on concrete news, personal experiences, or vivid memories were associated with *lower* accuracy.

Notably, the predictive inaccuracy of experts cannot simply be dismissed as a result of policy makers heeding their cautionary advice and taking actions that mitigated the worst predicted effects of the pandemic. If this had been the case, then we should have observed greater differences between prospective and retrospective judgments (and reduced severity of retrospective judgments) especially in domains like depression and subjective well-being, where policy responses like fiscal stimulus bills might be expected to have the greatest impact. Instead, we find that expert judgments in these domains are if anything *more* inaccurate and *more* extreme in retrospect.

These findings raise questions about how to incorporate the voices of social scientists when formulating public policy responses to global crises. If expert judgments of societal change are largely inaccurate (*12*, *55*), their individual or aggregated recommendations may need to be considered with a grain of salt. On the one hand, they may shed light on general expectations in a society and contextual factors to consider in decision-making, communicated with a certain degree of humility. On the other hand, if such recommendations are taken out of context and presented with more confidence than warranted, they may be counter-productive from both the policy-making and general public perspectives (*56*). For instance, if NGOs read that social scientists expect charity donations to go down over the course of the pandemic, they may reduce fundraising efforts, thereby resulting in a self-fulfilling prophecy (*57*–*59*) of lower donations in return. In such instances, minimal guidelines for assessing confidence in expert judgment may be beneficial (*56*). It is also worth noting in this light that social scientists showed more *meta-accuracy* in their confidence judgments than in their predictions: compared to lay people across all domains, scientists showed greater awareness that they might be wrong. It may be advantageous in future to focus on developing strategies that can improve forecasting accuracy at both the group level (*40*) and the individual level (*32*, *41*), such as training in epistemic humility.

Why are social science experts no more accurate than lay people? We propose two inter-related explanations. First, most social scientists have little training in prediction-oriented (as opposed to explanation-oriented) models (*15*, *16*). The fact that not only the sign but also the magnitude of expert judgments aligned closely with those of the general public supports this interpretation. That there were no major differences between graduate students and tenured faculty further corroborates the absence of training effects in making these predictions. Lack of attention to out-of-sample prediction limits the generalizability of the existing social science theories that experts may draw on to estimate changes in the real world. This latter insight might explain why experts were not only inaccurate in their predictions, but also in retrospective evaluation of change. Second, no formal psychological models of societal change in response to a once-in-a-century event like the pandemic exist (*42*). Without theory and necessary training, social scientists likely based their estimates of societal change on the same naïve theories of human social dynamics as lay people (*18*, *19*).

**Materials and methods summary**

The project was approved by the Office of Research Ethics at the University of Waterloo (#42123). Pre-registration, materials, and methods are available on Openm Science Framework at osf.io/9btsy.

**Study 1**

***Power***

We did not set an *a priori* criterion for sample size recruitment, aiming to recruit the largest number of participants we could in the available time. A power analysis suggested that regression analyses would have 90% power to determine small effects (*f*2 = .1) of time and educational status with a sample size of 130. Our sample sizes more than doubled this number.

***Participants***

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

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 five minutes (pilot testing with research associates revealed that five minutes 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. See Table S1 for demographics information.

***Procedure***

Participants first answered several demographic questions (see Table S1 in the on-line supplement). Participants next predicted cultural change in the USA for 11 domains, presented in a randomized order: implicit and explicit prejudice towards 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 osf.io/npzcr and see Table S2 in the Supplementary Information). 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).

Beyond the eleven 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.

**Study 2**

***Power***

In Study 2, we aimed to compare predictions made by social scientists in April and May 2020, as well as to compare lay and expert predictions. Power analyses using a two-sample *t*-test suggested that to detect medium effects (*d* = .2), a sample size of ~400 participants per group was enough to achieve 80% power. Given the within-subject component of our design (each participating providing 3 ratings for each of 15 domains), the effective power is even larger.

***Participants***

Whereas Study 1 focused on predictions before the initial peak of COVID-19 cases in the US, Study 2 was conducted after the initial peak. In the last week of April and the first week of May 2020, we recruited another group of social scientists using the same methods described in Study 1. A total of 354 social scientists provided their forecasts during the last week of April/first week of May (98% non-overlapping 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 minutes), and four who indicated they were undergraduate students. The final sample included 316 participants from 26 countries (see Table S1 for demographic information).

Concurrently in the first week of May 2020, we also obtained forecasts from lay people via a nationally representative sample of English-speaking US residents via the crowdsourcing UK-based company Prolific (www.prolific.co). 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 US Census Bureau (Prolific, 2020). 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 minutes. The final sample consisted of 394 participants. Prolific participants received 1.10 GBP for completing the survey.

***Procedure***

Participants in Study 2 followed the same general procedure outlined for Study 1. They provided forecasts for the same 11 domains from Study 1, as well as for 4 additional domains: loneliness, religiosity, charitable giving, and prevalence of violent crimes (verbatim questions on osf.io/npzcr and in Table S1 in the Supplementary Information). These domains were included based on preliminary examination of open-ended responses from Study 1 and theoretical considerations about the role of pathogen threat for social change. As in Study 1, participants provided forecasts at 6 months, 1 year, and 2 years on a sliding scale of 50% or greater decrease (-50) to 50% or greater increase (+50). For each domain, participants also rated their confidence in their predictions on a 5-point scale (1 = Not at all to 5 = Extremely). Participants also answered additional demographic questions (see supplement for verbatim items). After completing the first part of the study, participants were asked to identify one key psychological or social issue in the United States not covered in the survey that they thought would change and were invited to make forecasts for another country of their choice.

**Study 3**

In Study 3, we aimed to compare prospective predictions from Study 2 to retrospective estimates of changes in these same domains. To this end, we invited new samples of social scientists and nationally representative samples of Americans to participate in a study that took place half a year after Study 2. Based on the affective forecasting literature, we preregistered the hypothesis that prospective predictions would be more extreme in terms of absolute % change from baseline, compared to retrospective estimates, for both social scientist and lay samples. Following the Study 2 results, we further predicted that lay samples would report greater subjective confidence in their estimates than experts.

***Sample size and power***

To match the Study 2 sample, we targeted a nationally stratified sample of 400 lay individuals. We also aimed to recruit as many social scientists as we could within the 2-week availability period of the survey, targeting between 250-400 social scientists. While the lay sample recruitment gave us full control over the sample size, the social scientist sample recruitment did not. We pre-registered a stopping rule, continuing to recruit social scientists for 2-weeks after advertising the survey via the same venues as Study 2, and terminating data collection after this period. This procedure ensured a roughly homogeneous time period for obtaining retrospective reports. As in Study 2, power analyses for a small effect size (*d* = .2, α = .05/ β = .20) of a two-sample t-test suggested that the sample sizes obtained were adequate, especially when considering the within-subject component of our design (each participating providing 15 ratings, one for each domain)

***Participants***

As in Study 2, we recruited a new nationally representative 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 May, 2020 survey match their Prolific IDs). Of the 445 participants who started the study, we removed 27 who had incomplete responses and 7 who indicated they took a forecasting survey in April. The final sample consisted of 411 participants.

We also aimed to obtain survey responses from a sample of social scientists, recruited via mailing lists (e.g., Social and Personality Psychology mailing list, JDM mailing list) and social media. Similar exclusion criteria were applied to this sample, with the exception that we did not require social scientists to be US citizens. A total of 350 social scientists provided their forecasts during the last week of October/first week of November 2020 (88 % non-overlapping 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 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 2 (verbatim questions on osf.io/9btsy and in Supplementary Information). 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).” We provided an example and a clarification to ensure participants understood the instructions (see Supplementary Information). As in prior studies, participants were presented with domains in a randomized order. For each domain, participants first used a slider to indicate how the given domain (e.g., life satisfaction) “in the United States has changed (in %) from there it stood near the end of April 2020.” As in Studies 1-2, the slider ranged from -50% to +50% and the initial point was centered at zero (no change). As in Study 2, participants subsequently indicated confidence in their estimates on a 5-point scale (“not at all,” “slightly,” “moderately,” “highly,” “extremely”).

***Exploratory measures***

We sought to understand the factors contributing to participants’ reasoning while estimating societal change. As outlined in pre-registered exploratory analyses, we focused on the moderating role of two social-cognitive variables: a) vividness/concreteness of personal experience with a domain; b) extent of news exposure to a domain. We sought to explore whether domains in which participants report concrete visualization of personal experiences or specific news (coded as present/absent) or brought to mind vivid memories that affected them personally (also coded as present/absent) would show more extreme average retrospective estimates. In other words, domains in which retrospective estimates were positive should show more positive estimates if concrete visualizations also accompanied them. On average, domains with negative retrospective estimates should show more negative estimates if concrete visualizations accompanied them.

We also sought to explore how concreteness and news exposure contributes to alignment of retrospective estimates in Study 3 and prospective estimates in Study 2. On the one hand, construal level theory of psychological distance (*33*) would predict that more concrete representation of estimates would lead to greater divergence from abstract prospective estimates. On the other hand, bringing concrete events to one’s mind may result in greater use of heuristics (*43*), biasing one’s retrospective estimates toward the extreme end.

To assess *concreteness* in reasoning about social change, after completing assessments of change and confidence for each domain, participants were presented with a prompt:

As you were reflecting on possible changes in different domains of life in the last half year, for which of the following domains did you consider news reports or specific events that occurred in the last 6 months?

To assess *vividness of memories* in reasoning about social change, participants were presented with a prompt:

For which of these domains did you bring to mind experiences that have occurred in the past 6 months, that affected you personally, and for which you have very vivid memories?

For each question, participants were presented with check-box options, with domains presented in the same randomized order as presented earlier.

**Analytical approach**

***Quantifying the accuracy of estimates***

We targeted national-level estimates for all domains where we could locate large-scale, nationally representative surveys assessing the state 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 US population at large. When possible, we used weighted averages to adjust for representativeness as per the US 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 US Census Bureau, USC’s Understanding America Survey, Nationscape, Gallup Panels, the National Commission on COVID-19, Criminal Justice and Giving Tuesday, among others. See Supplementary Table S2 for each source, and Supplementary Methods for the exact wording of the question. When estimates were based on the percentage of the population at the given time point, we calculated the difference score. When the data was based on the sample estimate of a scale-based response, we calculated percentage change between the initial estimate of the sample in April 2020 and the subsequent estimate half a year later. Ultimately, we were able to quantify societal change in the US 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 supplement.

To examine accuracy, our main criterion was a comparison of prospective and retrospective estimates of change against actual change. To quantify how many participants were accurate in their estimates, we used three benchmarks of decreasing stringency: i) being within 1% point of the actual estimate (bound of half a percent point on each side of the accuracy estimate); ii) being within 5% point of the actual estimate (bound of 2.5% on each side of the accuracy estimate); iii) being within 20% point of the actual estimate (bound of 10% on each side of the accuracy estimate). We compared the percentage of participants within each benchmark by estimate type (prospective / retrospective), sample type (lay / expert) and domain type.

Notably, (in)accuracy can be assessed both by the magnitude of difference in percentage change, but also (in)accuracy in the direction of change (44). Lay and scientific intuitions about change may be accurate in direction (improvement / status quo / decline) even if the magnitude of predicted change is largely inaccurate. Consequently, we also compared accuracy in the direction of change (yes/no) as a function of the type of estimate (prospective / retrospective), sample type (lay / expert) and domain type. This allowed us to measure directional accuracy regardless of whether the estimated percentage change fell close to the true value.

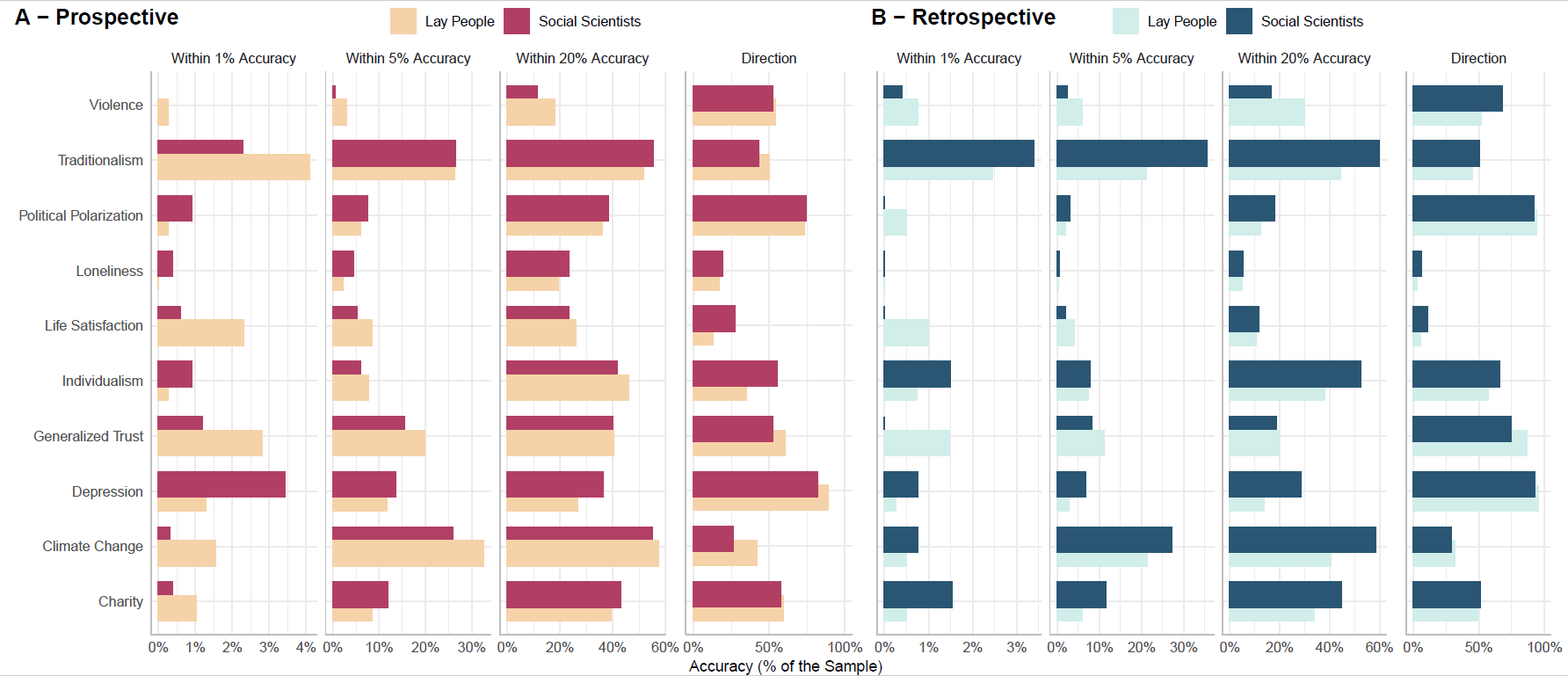
***Statistical modeling and inference criteria***

To quantify the mean level and trajectory of beliefs about the consequences of the pandemic, we used general linear mixed-effects models, as instantiated by the *lme4* package in *R,* to characterize the effects of time on estimates for each domain. We then used the Benjamini-Hochberg (*45*) false discovery rate correction method for post-hoc comparisons when examining effects across multiple domains.

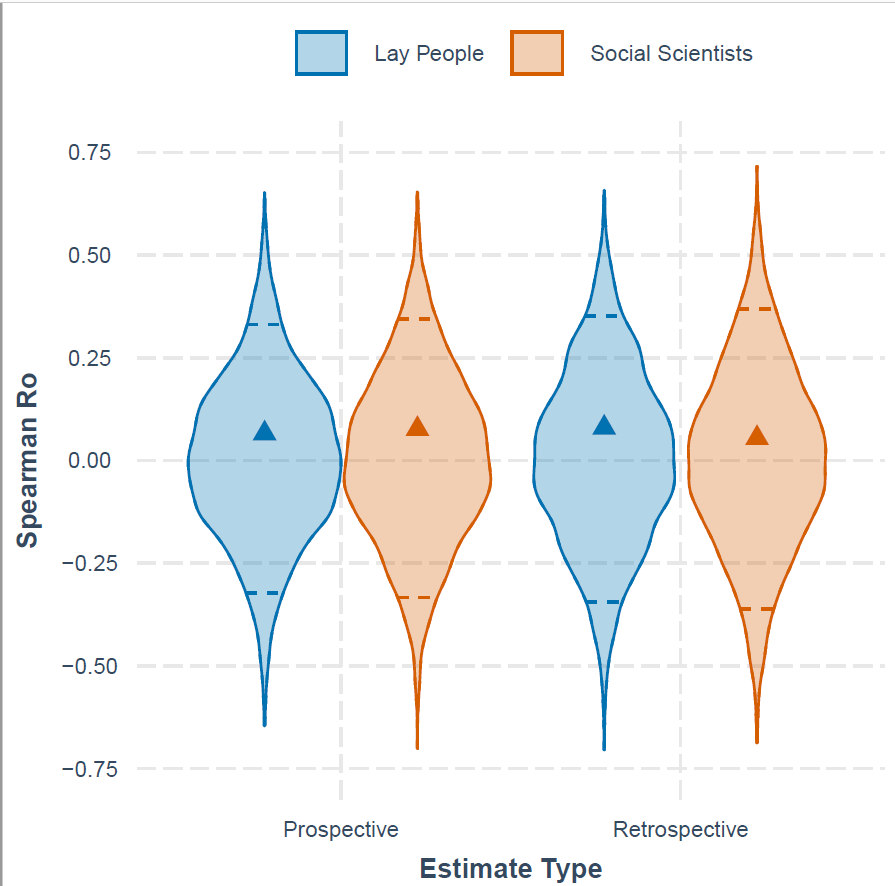
Since null hypothesis testing is inappropriate for assessing equivalence (or hypothesized lack of differences), we used Bayesian analyses [informative priors, *N* ~ (0, 0.5)] to evaluate the strength of evidence for and against group differences across time points and between social scientists and lay people (*46*). In addition to computing Bayesian credible intervals, we also calculated the Bayes factor (BF01) for the null hypothesis, which assumes that the coefficient of interest is equal to zero. Thus, higher BF01values indicate greater evidence in support of the null hypothesis (*47*). We consider a Bayes factor between one and three as indicative of weak, between three and ten of moderate, and finally over ten as strong evidence in support of the null hypothesis (*48*).



*Figure 1*. Prospective (April 2020) and retrospective (October/November 2020) judgment of societal change along with objective markers (dotted line) social scientists and lay people. We included social scientists from both Studies 1 and 2 because their forecasts were largely the same (see Table S10). Error bars indicate 95% confidence interval.



*Figure 2*. Percentage of a given sample that accurately estimated societal change. Panels from left to right represent different accuracy benchmarks: percentage of accurate estimates that (*i*) fall within 1%, (*ii*) 5% and (*iii*) 20% of the accuracy and (*iv*) directional accuracy of the trend.



*Figure 3*. Rank-order accuracy of estimates. Triangles represent the average Spearman’s rank order correlation between individual estimates and observed change across domains. Violin plots represent the expected null distribution of group-averaged correlation coefficients, constructed using 5000 permutations that randomly shuffled domain labels of the observed outcomes. Dashed lines indicate the 95% confidence interval of the null distribution. Average rank-order estimates of each group fell well within this distribution, suggesting that they are not significantly different from estimates expected by chance.

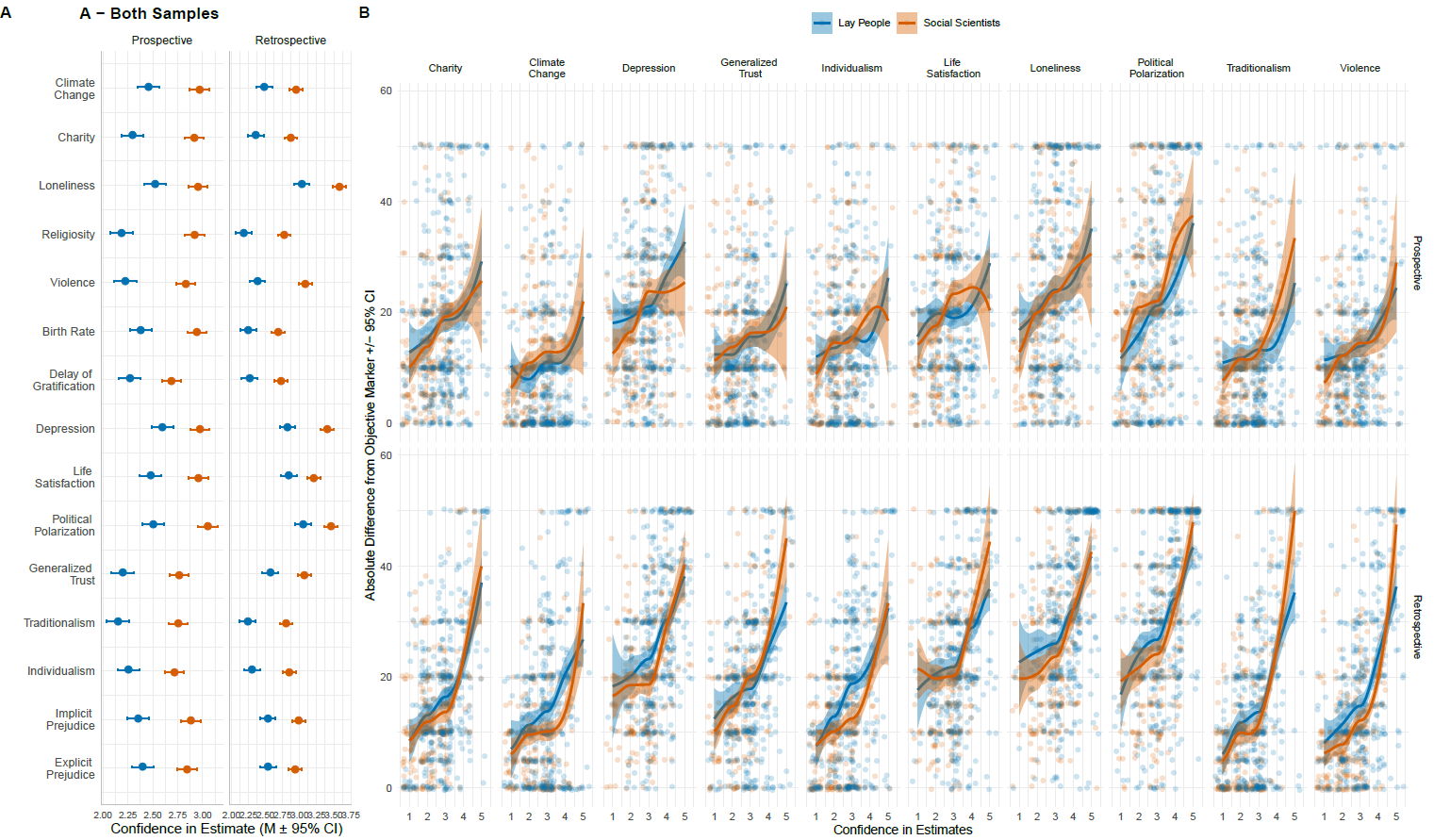


Figure 4. Panel A: Confidence of social scientists and lay people in their forecasted societal change for each of 15 domains for prospective and retrospective estimates, with error bars representing the 95% CI. Panel B: Individual-level inaccuracy for each participant by domain, as a function of confidence and sample type (social scientists or lay people) for prospective and retrospective estimates, with higher scores on the y-axis showing greater inaccuracy. Scatterplots with locally weighted polynomial regression (loess) line of best fit. Shaded region indicates 95% confidence interval.

**References**

1. J. J. Van Bavel, K. Baicker, P. S. Boggio, V. Capraro, A. Cichocka, M. Cikara, M. J. Crockett, A. J. Crum, K. M. Douglas, J. N. Druckman, J. Drury, O. Dube, N. Ellemers, E. J. Finkel, J. H. Fowler, M. Gelfand, S. Han, S. A. Haslam, J. Jetten, S. Kitayama, D. Mobbs, L. E. Napper, D. J. Packer, G. Pennycook, E. Peters, R. E. Petty, D. G. Rand, S. D. Reicher, S. Schnall, A. Shariff, L. J. Skitka, S. S. Smith, C. R. Sunstein, N. Tabri, J. A. Tucker, S. van der Linden, P. van Lange, K. A. Weeden, M. J. A. Wohl, J. Zaki, S. R. Zion, R. Willer, Using social and behavioural science to support COVID-19 pandemic response. *Nat. Hum. Behav.* **4**, 460–471 (2020).

2. J. Zaki, Fighting coronavirus feels like a war. That might bring us together. *Washint. Post* (2020).

3. A. Orlando, How will we remember the coronavirus pandemic? *Discov. Mag.* (2020), (available at https://www.discovermagazine.com/mind/how-will-we-remember-the-coronavirus-pandemic).

4. R. West, S. Michie, G. J. Rubin, R. Amlôt, Applying principles of behaviour change to reduce SARS-CoV-2 transmission. *Nat. Hum. Behav.* **4**, 451–459 (2020).

5. L. C. Hawkley, J. T. Cacioppo, Loneliness Matters: A Theoretical and Empirical Review of Consequences and Mechanisms. *Ann. Behav. Med.* **40**, 218–227 (2010).

6. M. Schaller, J. H. Park, The behavioral immune system (and why it matters). *Curr. Dir. Psychol. Sci.* (2011), doi:10.1177/0963721411402596.

7. C. L. Fincher, R. Thornhill, D. R. Murray, M. Schaller, Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism. *Proc. R. Soc. B Biol. Sci.* (2008), doi:10.1098/rspb.2008.0094.

8. J. Faulkner, M. Schaller, J. H. Park, L. A. Duncan, Evolved disease-avoidance mechanisms and contemporary xenophobic attitudes. *Gr. Process. Intergr. Relations* (2004), doi:10.1177/1368430204046142.

9. D. R. Murray, R. Trudeau, M. Schaller, On the origins of cultural differences in conformity: Four tests of the pathogen prevalence hypothesis. *Personal. Soc. Psychol. Bull.* (2011), doi:10.1177/0146167210394451.

10. J. M. Tybur, Y. Inbar, L. Aarøe, P. Barclay, F. K. Barlowe, M. De Barra, D. V. Beckerh, L. Borovoi, I. Choi, J. A. Choik, N. S. Consedine, A. Conway, J. R. Conway, P. Conway, V. C. Adoric, D. E. Demirci, A. M. Fernández, D. C. S. Ferreirat, K. Ishii, I. Jakšic, T. Ji, F. Van Leeuwen, D. M. G. Lewis, N. P. Li, J. C. McIntyre, S. Mukherjee, J. H. Park, B. Pawlowski, M. B. Petersen, D. Pizarro, G. Prodromitis, P. Prokop, M. J. Rantala, L. M. Reynolds, B. Sandin, B. Sevi, D. De Smet, N. Srinivasan, S. Tewari, C. Wilson, J. C. Yong, I. Žezelj, Parasite stress and pathogen avoidance relate to distinct dimensions of political ideology across 30 nations. *Proc. Natl. Acad. Sci. U. S. A.* (2016), doi:10.1073/pnas.1607398113.

11. P. E. Tetlock, *Expert Political Judgement: How Good Is It?* (Princeton University Press, Princeton, New Jersey, 2005).

12. J. S. Armstrong, *Long-range forecasting: From crystal ball to computer.* (John Wiley & Sons, New York, 1985).

13. R. Dawes, D. Faust, P. Meehl, Clinical versus actuarial judgment. *Science (80-. ).* **243**, 1668–1674 (1989).

14. S. Ægisdóttir, M. J. White, P. M. Spengler, A. S. Maugherman, L. A. Anderson, R. S. Cook, C. N. Nichols, G. K. Lampropoulos, B. S. Walker, G. Cohen, J. D. Rush, The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction. *Couns. Psychol.* **34**, 341–382 (2006).

15. T. Yarkoni, J. Westfall, Choosing prediction over explanation in psychology : Lessons from machine learning. *Perspect. Psychol. Sci.* **12**, 1100–1122 (2017).

16. J. M. Hofman, A. Sharma, D. J. Watts, Prediction and explanation in social systems. *Science (80-. ).* **355**, 486–488 (2017).

17. E. G. Clary, A. Tesser, Reactions to Unexpected Events: The Naive Scientist and Interpretive Activity. *Personal. Soc. Psychol. Bull.* (1983), doi:10.1177/0146167283094010.

18. G. A. Kelly, *The psychology of personal constructs. Volume 1: A theory of personality* (WW Norton and Company, 1955).

19. F. Heider, The naive analysis of action. (1958).

20. G. L. Murphy, D. L. Medin, The Role of Theories in Conceptual Coherence. *Psychol. Rev.* (1985), doi:10.1037/0033-295X.92.3.289.

21. A. Tversky, D. Kahneman, Judgment under uncertainty: Heuristics and biases. *Science (80-. ).* (1974), doi:10.1126/science.185.4157.1124.

22. T. Martin, J. M. Hofman, A. Sharma, A. Anderson, D. J. Watts, in *Proceedings of the 25th International Conference on World Wide Web* (International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2016), pp. 683–694.

23. R. E. Nisbett, G. T. Fong, D. R. Lehman, P. W. Cheng, Teaching reasoning. *Science (80-. ).* **238** (1987).

24. G. T. Fong, R. E. Nisbett, Immediate and delayed transfer of training effects in statistical reasoning. *J. Exp. Psychol. Gen.* **120**, 34–45 (1991).

25. R. M. Gardner, S. Dalsing, Misconceptions About Psychology Among College Students. *Teach. Psychol.* (1986), doi:10.1207/s15328023top1301\_9.

26. R. M. Gardner, R. M. Hund, Misconceptions of Psychology Among Academicians. *Teach. Psychol.* (1983), doi:10.1207/s15328023top1001\_5.

27. A. K. Taylor, P. Kowalski, Naïve psychological science: The prevalence, strength, and sources of misconceptions. *Psychol. Rec.* (2004), doi:10.1007/BF03395459.

28. B. Mellers, E. Stone, T. Murray, A. Minster, N. Rohrbaugh, M. Bishop, E. Chen, J. Baker, Y. Hou, M. Horowitz, L. Ungar, P. Tetlock, Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions. *Perspect. Psychol. Sci.* **10**, 267–281 (2015).

29. L. Artazcoz, J. Benach, C. Borrell, I. Cortès, Unemployment and Mental Health: Understanding the Interactions Among Gender, Family Roles, and Social Class. *Am. J. Public Health*. **94**, 82–88 (2004).

30. Y. Kashima, J. Shi, K. Tsuchiya, E. S. Kashima, S. Y. Y. Cheng, M. M. Chao, S. hui Shin, Globalization and folk theory of social change: How globalization relates to societal perceptions about the past and future. *J. Soc. Issues* (2011), doi:10.1111/j.1540-4560.2011.01723.x.

31. G. Tichy, The over-optimism among experts in assessment and foresight. *Technol. Forecast. Soc. Change*. **71**, 341–363 (2004).

32. B. Mellers, P. Tetlock, H. R. Arkes, Forecasting tournaments, epistemic humility and attitude depolarization. *Cognition*. **188**, 19–26 (2019).

33. Y. Trope, N. Liberman, Construal-level theory of psychological distance. *Psychol. Rev.* **117**, 440–463 (2010).

34. H. S. Horn, Optimal tactics of reproduction and life-history. *Behav. Ecol. An Evol. approach*, 411–429 (1978).

35. V. Griskevicius, J. M. Tybur, A. W. Delton, T. E. Robertson, The influence of mortality and socioeconomic status on risk and delayed rewards: A life history theory approach. *J. Pers. Soc. Psychol.* (2011), doi:10.1037/a0022403.

36. K. S. Kendler, L. M. Karkowski, C. A. Prescott, Causal relationship between stressful life events and the onset of major depression. *Am. J. Psychiatry* (1999), doi:10.1176/ajp.156.6.837.

37. I. Grossmann, M. E. W. Varnum, Social Class, Culture, and Cognition. *Soc. Psychol. Personal. Sci.* **2**, 81–89 (2011).

38. L. J. Ji, Y. Su, R. E. Nisbett, Culture, change and prediction. *Psychol. Sci.* **12**, 450–456 (2001).

39. T. D. Wilson, D. T. Gilbert, in *Advances in Experimental Social Psychology*, M. P. Zanna, Ed. (Academic Press, San Diego, CA, 2003; http://www.sciencedirect.com/science/article/pii/S0065260103010062), vol. Volume 35, pp. 345–411.

40. M. G. Morgan, Use (and abuse) of expert elicitation in support of decision making for public policy. *Proc. Natl. Acad. Sci.* **111**, 7176–7184 (2014).

41. I. Grossmann, A. Dorfman, H. Oakes, H. C. Santos, K. D. Vohs, A. Scholer, Training for Wisdom: The Distanced Self-Reflection Diary Method. *Psychol. Sci.* (2021), doi:10.1177/09567976209691.

42. J. M. Ackerman, J. M. Tybur, A. D. Blackwell, What Role Does Pathogen-Avoidance Psychology Play in Pandemics? *Trends Cogn. Sci.* **25**, 177–186 (2021).

43. D. Kahneman, P. Slovic, A. Tversky, *Judgment under uncertainty: Heuristics and biases* (Cambridge University Press, New York, 1982).

44. L. Ross, in *Advances in experimental social psychology*, L. Berkowitz, Ed. (Academic Press, New York, NY, 1977), pp. 173–220.

45. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B*. **57**, 289–300 (1995).

46. Z. Dienes, Using Bayes to get the most out of non-significant results. *Front. Psychol.* **5** (2014), doi:10.3389/fpsyg.2014.00781.

47. E.-J. Wagenmakers, M. Marsman, T. Jamil, A. Ly, J. Verhagen, J. Love, R. Selker, Q. F. Gronau, M. Šmíra, S. Epskamp, D. Matzke, J. N. Rouder, R. D. Morey, Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychon. Bull. Rev.* **25**, 35–57 (2018).

48. M. D. Lee, E.-J. Wagenmakers, *Bayesian Cognitive MOdeling: A Practical Course* (Cambridge University Press, 2014).

49. M. Luhmann, Using Big Data to study subjective well-being. *Curr. Opin. Behav. Sci.* **18**, 28–33 (2017).

50. H. A. Schwartz, M. Sap, M. L. Kern, J. C. Eichstaedt, A. Kapelner, M. Agrawal, E. Blanco, L. Dziurzynski, G. Park, D. Stillwell, M. Kosinski, M. E. P. Seligman, L. H. Ungar, in *Biocomputing 2016* (World Scientific, 2016), pp. 516–527.

51. D. Witters, J. Harter, “In U.S., Life Ratings Plummet to 12-Year Low” (2020), (available at https://news.gallup.com/poll/308276/life-ratings-plummet-year-low.aspx).

52. L. Boxell, J. Conway, J. Druckman, M. Gentzkow, “Affective Polarization Did Not Increase During the Coronavirus Pandemic” (Cambridge, MA, 2020), , doi:10.3386/w28036.

53. E. Hehman, J. K. Flake, J. Calanchini, Disproportionate Use of Lethal Force in Policing Is Associated With Regional Racial Biases of Residents. *Soc. Psychol. Personal. Sci.* **9**, 393–401 (2018).

54. E. K. Ofosu, M. K. Chambers, J. M. Chen, E. Hehman, Same-sex marriage legalization associated with reduced implicit and explicit antigay bias. *Proc. Natl. Acad. Sci.* **116**, 8846–8851 (2019).

55. T. E. S. Charlesworth, M. R. Banaji, Patterns of Implicit and Explicit Attitudes: I. Long-Term Change and Stability From 2007 to 2016. *Psychol. Sci.* **30**, 174–192 (2019).

56. A. G. Greenwald, B. A. Nosek, M. R. Banaji, Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *J. Pers. Soc. Psychol.* **85**, 197–216 (2003).

57. N. Gross, E. Fosse, Why are professors liberal? *Theory Soc.* **41**, 127–168 (2012).

58. T. Talhelm, J. Haidt, S. Oishi, X. Zhang, F. F. Miao, S. Chen, Liberals Think More Analytically (More “WEIRD”) Than Conservatives. *Personal. Soc. Psychol. Bull.* **41**, 250–267 (2015).

59. R. E. Nisbett, K. Peng, I. Choi, A. Norenzayan, Culture and systems of thought: Holistic vs. analytic cognition. *Psychol. Rev.* **108**, 291–310 (2001).

**Supplementary Information**

**for**

**The pandemic fallacy: Inaccuracy of social scientists’ and lay judgments about COVID-19’s societal consequences in America**

Cendri A. Hutcherson1,2, Konstantyn Sharpinskyi3, Michael Varnum4, Amanda Rotella3, Alexandra Wormley4, Louis Tay5, & Igor Grossmann3

1 Department of Psychology, University of Toronto Scarborough, Toronto, ON M1C 1A4, Canada

2 Department of Marketing, Rotman School of Management, University of Toronto, Toronto, ON M5S 3E6, Canada

3 Department of Psychology, University of Waterloo, Waterloo, ON, N2L 3G1, Canada

4 Department of Psychology, Arizona State University, Tempe, AZ, 85287

5 Department of Psychological Sciences, Purdue University, West Lafayette, IN, 47907

**Author Note**

Correspondence should be addressed to Igor Grossmann, PAS 3047, University of Waterloo, Waterloo, ON N2L 3G1, [igrossma@uwaterloo.ca](mailto:igrossma@uwaterloo.ca).

Contents

[**Methods** 25](#_Toc64629912)

[***Confidence ratings*** 25](#_Toc64629913)

[***Demographics*** 25](#_Toc64629914)

[***Quantifying cultural distance*** 27](#_Toc64629915)

[***Quantifying open-ended responses in forecasting studies*** 27](#_Toc64629916)

[**Quantifying domain-specific expertise** 28](#_Toc64629917)

[**Accuracy** 30](#_Toc64629918)

[*Life satisfaction* 30](#_Toc64629919)

[*Additional markers of life satisfaction* 30](#_Toc64629920)

[*Loneliness* 31](#_Toc64629921)

[*Additional marker of loneliness* 31](#_Toc64629922)

[*Depression* 32](#_Toc64629923)

[*Additional markers of depression* 32](#_Toc64629924)

[*Additional marker of affective polarization* 34](#_Toc64629925)

[*Individualism* 34](#_Toc64629926)

[*Generalized Trust* 35](#_Toc64629927)

[*Traditionalism* 35](#_Toc64629928)

[*Additional marker of traditionalism* 35](#_Toc64629929)

[*Violence* 35](#_Toc64629930)

[*Attitudes toward climate change* 36](#_Toc64629931)

[*Charitable Giving* 36](#_Toc64629932)

[Supplementary benchmark indices 36](#_Toc64629933)

[**Results** 40](#_Toc64629934)

[**Deliberation check** 40](#_Toc64629935)

[**Description of predictions by domain** 40](#_Toc64629936)

[**Consensus among scientists’ predictions** 41](#_Toc64629937)

[**Changes in Predictions** 41](#_Toc64629938)

[**Social Scientists** 41](#_Toc64629939)

[**Retrospective Estimates** 52](#_Toc64629940)

[*Social scientists vs. lay people* 54](#_Toc64629941)

[*Domain expertise* 54](#_Toc64629942)

[*Vividness of memories and news exposure* 54](#_Toc64629943)

[*Prospective vs. retrospective estimates* 56](#_Toc64629944)

[**Accuracy** 57](#_Toc64629945)

[**Confidence and political orientation** 58](#_Toc64629946)

[**Confidence, accuracy, and personal experience** 59](#_Toc64629947)

[**Additional Supplementary Tables** 61](#_Toc64629948)

# Methods

Participants completed a series of questionnaires to assess their prospective (Studies 1 & 2) or retrospective (Study 3) predictions about how culture has changed in response to the COVID-19 pandemic. Participants were asked to estimate the amount of change across fifteen domains (for this list and their definitions, see Table S1). Additionally, they were asked to indicate the confidence of each of their estimate, followed by demographic questions.

## ***Confidence ratings***

In Study 1, we asked social scientists to provide an “estimate of the probability for this forecast being true (i.e., what is the likelihood that your estimate falls within 5% of the true value)?,” with responses on a scale from 0 to 100. As several social scientist participants pointed out to us, this question was not well understood. Therefore, we a priori chose not to analyze this initial question.

In Studies 2-3, we modified the question to a simpler question: “How confident are you in your prediction [Study 2]/estimate [Study 3]?’ We recorded responses on a 5-point scale, with exact anchor points: 1 = “not at all”, 2 = “slightly”, 3 = “moderately”, 4 = “highly”, 5 = “extremely”.

## ***Demographics***

Across Studies 1-3, participants reported organizational affiliation, organization size, ethnicity, annual total household income, political beliefs, gender, and age. We assessed organizational affiliation on a four-point scale. We assessed organizational size on a six-point scale. Participants selected their ethnicity from one of nine categories. Then, they indicated their total annual household income. See Table S1 for category labels.

We also measured political beliefs using a 7-point scale: 1 = Progressive, 4 = Neutral, 7 = Conservative and gender with these three choices: 1 = Woman, 2 = Man, 3 = Non-binary. Participants provided their biological age by typing a number into a textbox.

In addition to basic demographic information, we asked several questions that were only applicable to social scientists: country of origin, academic position, and field of research, additional fields of research and areas of expertise. Primary country of residence into a textbox and indicated their current position by selecting one of seven options: 1 = tenured faculty, 2 = nontenured faculty, 3 = adjunct professor, 4 = postdoc, 5 = graduate student, 6 = research scientist, 7 = other. They then indicated their main field of research by selection one from the following list: 1 = Psychology, 2 = Neuroscience, 3 = Medicine, 4 = Sociology, 5 = Political Science, 6 = Economics, 7 = Epidemiology, 8 = Biology, 9 = Computer science, 10 = Other.” In Studies 1-2 we asked participants to type in any additional fields of research they engaged in and to list any domain-relevant areas of expertise (e.g., prejudice, mental health, etc.). In Study 3, the open-ended domain-relevant area of expertise question was replaced with a 15-item multi-selection list where social scientists selected all domains they believed to have expertise in.

**Table S1** Descriptive Statistics

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Study 1 | Study 2  (Social Scientists) | Study 2  (Lay people) | Study 3  (Social Scientists) | Study 3  (Lay people) |
| Sample (*N*) |  | 401 | 316 | 394 | 270 | 411 |
| Age | *Mage* | 41 | 39 | 45 | 38 | 45 |
| Range | 22 – 88 | 19 – 87 | 18-78 | 22-76 | 18-78 |
| Gender | % Female | 45% | 63% | 52% | 72% | 50% |
| Household Income (*Md*) |  |  | $100,001 - $150,000 | $50,001 - $75,000 | $75,001 - $100,000 | $50,001 - $75,000 |
| Ethnicity | White |  | 234 (75%) | 269 (68%) | 212 (79%) | 276 (67%) |
| Asian |  | 31 (10%) | 29 (7%) | 22 (8%) | 33 (8%) |
| Hispanic |  | 13 (4%) | 18 (5%) | 5 (2%) | 21 (5%) |
| Black |  | 2 (1%) | 51 (13%) | 3 (1%) | 57 (14%) |
| Middle Eastern |  | 3 (1%) | 2 (1%) | 3 (1%) | 4 (1%) |
| East Indian |  | 8 (3%) | - | 3 (1%) | - |
| Aboriginal |  | - | 2 (1%) | 3 (1%) | - |
| Other |  | 21 (7%) | 23 (6%) | 18 (7%) | 18 (4%) |
| Country | USA | 195 (50%) | 194 (62%) | 394 (100%) | 196 (73%) | 411 (100%) |
|  | Canada | 93 (24%) | 57 (18%) |  | 47 (17%) |  |
|  | Germany | 17 (4%) | 7 (2%) |  | 5 (2%) |  |
|  | United Kingdom | 16 (4%) | 17 (5%) |  | 4 (1%) |  |
|  | Australia | 9 (2%) | 3 (1%) |  | - |  |
|  | Netherlands | 7 (2%) | 3 (1%) |  | 1 (< 1%) |  |
|  | Switzerland | 6 (2%) | 3 (1%) |  | 4 (1%) |  |
|  | China | 2 (< 1%) | 5 (2%) |  | - |  |
|  | Other | 48 (12%) | 23 (8%) |  | 13 (5%) |  |
| Research Field | Psychology | 330 (83%) | 237 (76%) |  |  |  |
| Neuroscience | 8 (2%) | 10 (3%) |  |  |  |
|  | Political Science | - | 21 (7%) |  |  |  |
|  | Economics | 11 (3%) | 8 (3%) |  |  |  |
|  | Computer Science | 1 (< 1%) | 5 (2%) |  |  |  |
|  | Sociology | 3 (1%) | 3 (1%) |  |  |  |
|  | Biology | - | 3 (1%) |  |  |  |
|  | Medicine and Epidemiology | 2 (<1%) | 3 (1%) |  |  |  |
|  | Other | 44 (11%) | 26 (8%) |  |  |  |
| Academic Position | Tenured Faculty | 137 (34%) | 82 (26%) |  |  |  |
| Non-Tenured Faculty | 73 (19%) | 37 (12%) |  |  |  |
|  | Postdoctoral Researchers | 40 (10%) | 32 (10%) |  |  |  |
|  | Graduate Students | 93 (23%) | 102 (33%) |  |  |  |
|  | Research Scientists | 28 (7%) | 29 (9%) |  |  |  |
|  | Other | 27 (7%) | 34 (11%) |  |  |  |
| Organization Type | College/University | 368 (92%) | 280 (89%) | 64 (16%) | 259 (96%) | 56 (14%) |
| Government | 10 (3%) | 10 (3%) | 23 (6%) | 2 (1%) | 21 (5%) |
| Private Company | 15 (4%) | 15 (5%) | 164 (42%) | 6 (2%) | 174 (43%) |
|  | Self-employed | 7 (2%) | 8 (3%) | 140 (36%) | 2 (1%) | 154 (38%) |
| Organization Size | < 10 | 6 (2%) | 10 (3%) | 156 (40%) | 3 (1%) | 154 (39%) |
| 11 - 100 | 11 (3%) | 7 (2%) | 54 (14%) | 5 (2%) | 49 (12%) |
|  | 101 - 1,000 | 17 (4%) | 19 (6%) | 55 (14%) | 7 (3%) | 79 (20%) |
|  | 1,001 – 10,000 | 86 (22%) | 52 (17%) | 53 (14%) | 43 (16%) | 57 (14%) |
|  | 10,001 – 50,000 | 201 (51%) | 163 (52%) | 39 (10%) | 139 (52%) | 43 (11%) |
|  | 50,000+ | 74 (19%) | 61 (20%) | 33 (8%) | 72 (27%) | 16 (4% |

*Notes.* Due to a technical error, information about the country of residence and research field was not collected in Study 3. To estimate country of residence for Study 3, we used each person’s IP address to geolocate their country of residence at the time of taking the survey (using the *rgeolocate* package in *R*).

## ***Quantifying cultural distance***

The Cultural Distance index data was retrieved from [http://culturaldistance.muth.io](http://culturaldistance.muth.io/) (Muthukrishna et al., 2020). Fourteen countries of residence identified in our study were not included in this dataset (Study 1 *n* = 9; Study 2 *n* = 7), in which case we imputed cultural distance values based on the countries close in geography, geopolitical and socio-historical proximity from the list. More specifically, data for the following countries were imputed to the closest correlate with the Cultural Distance dataset: Austria – Germany, Bolivia – Chile, Czech Republic - Serbia and Montenegro, Denmark - Sweden, Greece - Cyprus, Ireland - Great Britain, Kenya - Ethiopia, Luxembourg - Germany, Myanmar - Vietnam, Scotland - Great Britain, Slovakia - Serbia and Montenegro, Oman - Qatar, Portugal - Spain. We could not estimate cultural distance for a handful of participants from Israel, because the Cultural Distance Index data did not include good proxy for this country. Results are close to identical if omitting scores from participants from these countries.

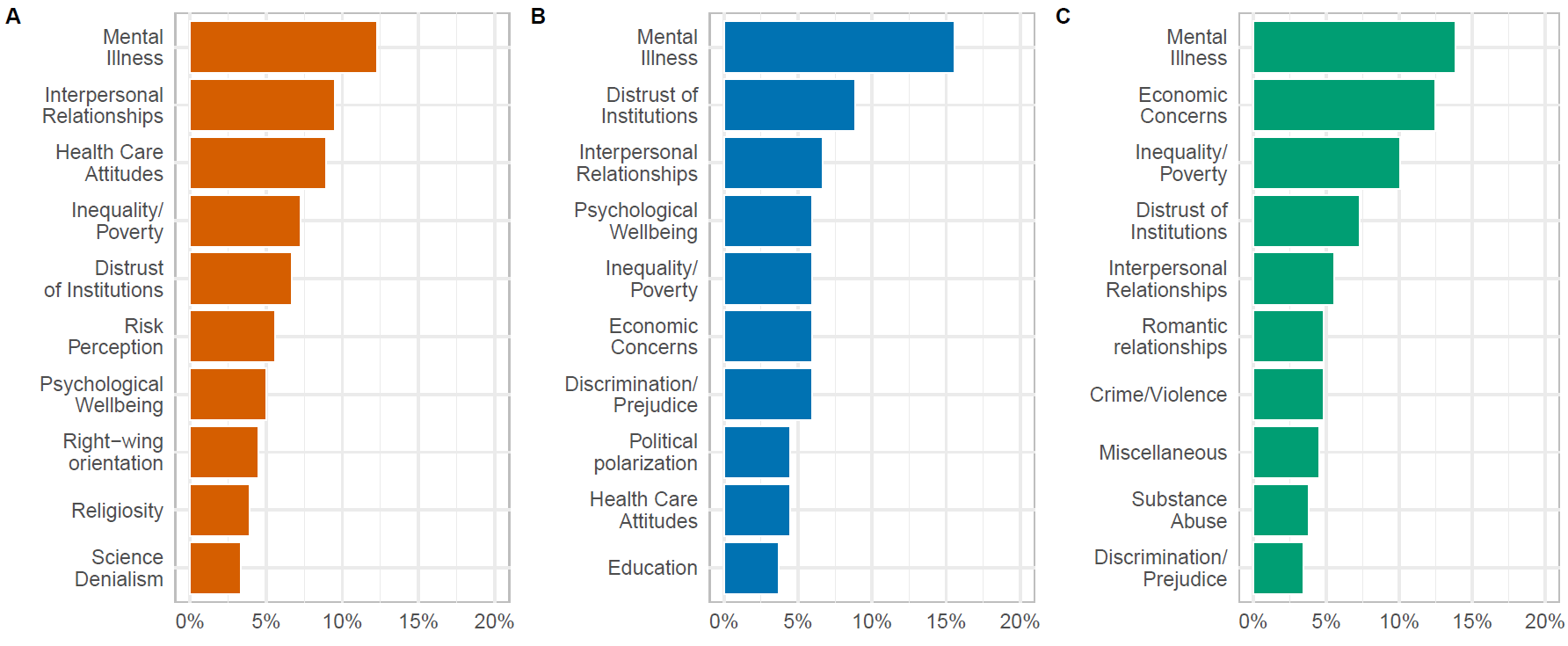
## ***Quantifying open-ended responses in forecasting studies***

Participants were asked to complete numerous open-ended responses, including the other domains of culture in which they expected to see change and domain of the expertise.

Using a grounded approach, two coders categorized open-ended responses for domains of cultural change from Studies 1-2 into a set of distinct categories. If participants mentioned more than one domain of change, coders were instructed to select the most salient domain. Iteratively, first, second, third, and senior authors reduced the set of categories to 35 groups: Ageism, Animal Welfare, Child Development, Civic-Mindedness, Civil Liberties, Consumer Behavior/Attitudes, Crime/Violence, Discrimination/Prejudice, (Dis)Trust in Institutions, Economic Concerns, Education, Environmentalism, Health Behaviors, Healthcare Politics, Immigration, Inequality/Poverty, Interpersonal Relationships, Isolationism, Medical, Mental Illness, Misinformation, Personality, Political Polarization, Psychological Well-being, Religiosity, Right-Wing Orientation, Risk Perception, Romantic Relationships, Science Denialism, Sexism, Social Norms, Social Welfare, Substance Ab/use, Technology Reliance, and Miscellaneous.

Subsequently, two new raters independently classified participants’ responses into one of the 35 categories. Inter-rater agreement across 35 categories was over 95% agreement, with minor disagreement resolved via discussion with the senior author.

Furthermore, we selected 20 of the most frequent categories from Study 1 & Study 2 (frequency > 2; see Figure S1 for the top ten) which yielded 31 unique categories). Using open-ended coding by two coders, we identified a parsimonious subset of six: health and well-being, interconnectedness, social justice, economics, (child) development and political discord and (mis)trust in institutions. We excluded seven initially identified categories because they did not fit into higher order meta categories: risk perception, religiosity, technology reliance, crime/violence, environmentalism, animal welfare and consumer behavior/attitudes (and miscellaneous).

*Figure S1*. Bar graphs depicting top ten most significant societal issues social scientists (Panels A: Study 1; and Panel B: Study 2) and lay people (Panel C) forecasted change for beyond the domains in the survey.

### **Quantifying domain-specific expertise**

To examine whether domain-specific expertise qualifies forecasts and/or retrospective accuracy, in Studies 1-2 we examined participants’ self-reported research areas, quantifying them in terms of applicability for each of the forecasted domains. Third and fourth authors independently categorized each of the listed research areas and subjects of study. Two coders used a grounded and iterative approach with input from the authors to code domain of expertise. Participants were asked to report their area of expertise. First, coders decided what category of social science this expertise fell into: social/personality psychology, cognitive psychology and neuroscience, clinical psychology, developmental, and other. Second, coders decided which domain of change each participant may have an expertise in. For example, a participant who said their expertise was in “prejudice” would be coded as a “social/personality psychology,” specifically with expertise in “implicit prejudice” and “explicit prejudice” within the domains focused on within this study. If participants mentioned more than one area of expertise, coders were instructed to select as many domains as applied. Inter-rater reliability was high (90% agreement). Disagreements were minor (< 5%) and resolved via discussion with the senior author.

To assess domain-specific expertise in Study 3, participants indicated in a check-box survey at the end of the study whether they had received graduate training/education (i.e., taking psychology classes or researching these or related topics) in any of the fifteen domains for which they provided estimates.

*Table S2.* Variable descriptions provided to participants for the fifteen domains.

|  |  |
| --- | --- |
| Variable | Description Provided to Participant |
| Generalized Trust | For each time period below, use the sliders to indicate how much generalized trust in other people in the United States will change (in %) from where it stands right now. |
| Life Satisfaction | Consider a person’s life satisfaction in the United States, an overall assessment of how content a person is with their life overall, and measured by endorsement of statements like “The conditions of my life are excellent.” |
| Clinical Depression | Consider clinical depression in the United States, as diagnosed by criteria listed in the Diagnostic and Statistical Manual, is characterized by feeling sad, losing interest in activities once enjoyed, and a loss of energy over a prolonged period of time, and is measured by agreement with statements like "I am sad all the time and I can't snap out of it." |
| Political Affective Polarization | Consider political affective polarization in the United States, defined as the degree of dislike and distrust towards those from the opposing political party. |
| Individualism | Consider people’s concern for individualism in the United States, defined as values that emphasize the uniqueness, autonomy and individual goal pursuit. |
| Traditionalism | Consider people’s concern for traditionalism in the United States, defined as a concern for adherence to traditional beliefs and practices, and measured by endorsement of items like, “the ‘old-fashioned ways’ and ‘old-fashioned values’ still show the best way to live.” |
| Delay of Gratification | Consider the extent to which people delay their gratification in the United States, defined as resistance to the temptation of an immediate reward in preference for a later, larger reward. |
| Explicit Prejudice | Consider endorsement of general levels of explicit prejudice toward ethnic minorities in the United States, defined in this case as consciously holding negative attitudes toward ethnic or racial minorities, and measured by endorsement of items such as “Over the past few years, ethnic/racial minorities have gotten more economically than they deserve.” |
| Implicit Prejudice | Consider measures of implicit prejudice toward ethnic minorities in the United States, defined here as negative feelings and/or beliefs about an ethnic group that people hold without being aware of it, and which is thought to operate automatically, with little intention or control on the part of the person. |
| Concern for Climate Change | Consider Americans' concern about climate change in the United States, as measured by endorsement of the idea that global warming is personally worrisome and a problem that is of pressing concern. |
| Birth Rates | Consider birth rates in the United States, defined as the total number of live births per 1000 women in the total population. |
| Charitable Giving | Consider charitable giving in the United States, defined as donations made by individuals to non-profit organizations, charities, or private foundations. |
| Violent Crimes | Consider violent crimes in the United States, defined as any violation of the law that is committed with physical force and is measured by rates of murder, manslaughter, rape, robbery, and aggravated assault. |
| Religiosity | Consider religiosity in the United States, defined as belief in a higher power, and measured by questions like "How certain are you of the existence of a higher power?" |
| Loneliness | Consider loneliness in the United States, defined as a person's subjective feeling of loneliness, and measured by agreement with statements like “A lot of times I feel lonely.” and "I often feel left out of things." |
| Other Variable | Is there another key psychological or social issue in the United States we have not mentioned that you think would change? If so, please identify ONE key issue that you think is most important. |

# **Accuracy**

To gauge how much each dimension changed between April/May and October at a national level we searched for representative surveys that tracked constructs of interest over time. We were able to identify 16 such surveys covering nine domains (four options for depression, three for life satisfaction, two for loneliness, and political polarization and one for the rest) from 11 sources.

## *Life satisfaction*

Our primary marker of life satisfaction, along with several secondary markers describe below, relied on Gallup Panel data – COVID-19 Survey. The COVID-19 web survey began fielding on March 13, 2020 with daily random samples of U.S. adults, aged 18 and older who are members of the Gallup Panel. Approximately 1,200 daily completes were collected from March 13 through April 26, 2020. From April 27 to August 16, 2020 approximately 500 daily completes are being collected. Starting August 17, 2020, the survey moved from daily surveying to a survey conducted one time per month over a two-week field period (typically the last two weeks of the month). The Gallup Panel is a probability-based, nationally representative panel of U.S. adults. Members are randomly selected using random-digit-dial phone interviews that cover landline and cellphones and address-based sampling methods.

Gallup weights the obtained samples each day to adjust for the probability of selection and to correct for nonresponse bias. Nonresponse adjustments are made by adjusting the sample to match the national demographics of gender, age, race, Hispanic ethnicity, education and region.  
Demographic weighting targets are based on the most recent Current Population Survey figures  
for the aged-18-and-older U.S. population.

As a benchmark for life satisfaction change, we used the same time period as estimated by participants in our forecasting Study 2 (April 23 – May 5, 2020) and retrospective Study 3 (October 14 – October 26, 2020). Gallup panel provided the closest match in the definition of life satisfaction provided to our Study 2-3 participants, as it was assessed with classic Cantril ladder question “Please imagine a ladder with steps numbered from 0 at the bottom to ten at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?” The response scale ranged from 10 – best possible to 0 – worst possible. We obtained average estimates across each time point. Subsequently, we calculated percentage change between April/May and October estimates as our primary marker of % societal change in life satisfaction.

### *Additional markers of life satisfaction*

#### Twitter-based estimates

We considered Twitter based estimates, because work suggests that national estimates obtained via social media language can reliably track subjective well-being (*49*). For each month, we used previously validated predictive models of well-being, as measured by life satisfaction scales (*50*). Life satisfaction was estimated using a ridge regression model trained on *latent Dirichlet allocation* topics, selected using univariate feature selection and dimensionally reduced using randomized principal component analysis, to predict Cantril ladder life satisfaction scores. Such Twitter-based estimates tend to follow nationally representative polls (*51*). We applied the respective models to Twitter data from late April/early May 2020 and late October 2020 to obtain estimates of life satisfaction via language on social media. Estimates obtained this way were 1% off from the Gallup estimates we selected and vastly different from the forecasted and retrospective estimates provided by our participants.

#### ICL/YouGov

Imperial College London partnered with YouGov to track behaviors over the period of coronavirus pandemic, and included measures of life satisfaction over time (<https://github.com/YouGov-Data/covid-19-tracker>). The Imperial College London YouGov Covid 19 Behaviour Tracker is a multinational COVID tracker of behavior, which included surveys for life satisfaction in the US. Unfortunately, the surveys for April and September were much smaller in scope compared to Gallup Polls and the US-based survey did not include estimates for October. Thus, we focused our analyses on the Gallup data described above.

## *Loneliness*

We used USC's Center for Economic and Social Research (CESR) Understanding America Study (UAS) – a probability-based Internet panel. Since April, 2020, the Understanding America Study included a coronavirus tracking poll. Each panel member was randomized to respond on a pre-assigned day of the week, distributed so that a full sample is invited to participate over a 14-day period. Respondents have 14 days to complete the survey but receive an extra monetary incentive for completing the survey on the day they are invited to participate. Data for the full sample is thus final after a 28-day period. Approximately 7100 adult residents of the U.S. have been participating in the ongoing surveys; roughly 550 per day. Full information on the survey, including full methodology, is available at <https://covid19pulse.usc.edu>.

The survey included the question “In the past 7 days, how often have you felt lonely?,” with response options 1 = “not at all or less than 1 day,” 2 – “1-2 days,” 3 – “3-4 days,” 4 – “5-7 days.” We used weighted average responses for this question for each of two time periods roughly corresponding to the time windows we gathered prospective survey data (Study 2) – April 15-May 15 and retrospective data (Study 3) – September 15-October 15. We subsequently calculated % change between these average points.

We chose the Understanding America Study marker of loneliness because it had a more balanced sample size across both time windows compared to other metrics (See Table S3 below) and because it more closely matches the definition provided to participants.

### *Additional marker of loneliness*

We used Gallup Panel data as a secondary marker of loneliness, assessed with a question “Did you experience the following feelings during A LOT OF THE DAY yesterday?” Response options included a range of feelings, from enjoyment, and happiness, to worry and loneliness. We examined weighted % of participants reported experiencing loneliness yesterday as a secondary marker of loneliness. We used the same time period as estimated by participants in our forecasting Study 2 (April 23 – May 5, 2020) and retrospective Study 3 (October 14 – October 26, 2020). However, the Gallup panel score for loneliness was not an ideal match for the definition of loneliness provided to our Study 2-3 participants (see Table S1), because Gallup data explicitly focused on the feeling of loneliness from the previous day (presence/absence) rather its magnitude (e.g., “A lot of times I feel lonely,” “I often feel left out of things”). Thus, we focus our primary analyses on the Understanding America Study described above. Estimates for societal change using Gallup Panel loneliness were very similar to the Understanding America Study estimates (see Table S3).

## *Depression*

Our primary source for rate of depression came from the US Centers for Disease Control and Prevention Household Pulse Survey. The Household Pulse Survey (HPS) was launched on April 23, 2020 as a joint effort between National Center for Health Statistics (NCHS) and Census Bureau to monitor changes in mental health throughout the COVID-19 pandemic. It concerned a 20-minute survey, aiming to assess frequency of anxiety and depression symptoms. The questions were modified versions of the two-item Patient Health Questionnaire (PHQ-2) and the two-item Generalized Anxiety Disorder (GAD-2) scale on the Household Pulse Survey, collecting information on symptoms over the last 7 days. Full information on the survey and methodology can be found online at <https://www.cdc.gov/nchs/covid19/pulse/mental-health.htm>. We focused on weighted depression scores from this survey, targeting the dates most closely matching our forecasted and retrospective Studies 2-3: April 23-May 5, 2020 and October 14 – October 26, 2020. We chose the CDC Household Pulse Survey, because it included the most extensive survey of clinical measures of depression and best matched questions we asked participants in Studies 1-2.

### *Additional markers of depression*

#### Gallup Panel

We used Gallup Panel data as a secondary marker of depression, assessed with a question “Did you experience the following feelings during A LOT OF THE DAY yesterday?” Response options included a range of feelings, from enjoyment, and happiness, to worry and depression. We examined weighted % of participants who reported experiencing depression yesterday as a secondary marker of depression. We used the same time period as estimated by participants in our forecasting Study 2 (April 23 – May 5, 2020) and retrospective Study 3 (October 14 – October 26, 2020). Gallup panel score for depression was not an ideal match for the definition of depression provided to our Study 2-3 participants (see Table S1), because Gallup data explicitly focused on the feeling of depression from the previous day (presence/absence) rather than a clinical definition of depression we provided to participants in the forecasting and retrospective studies (characterized by feeling sad, losing interest in activities once enjoyed, and a loss of energy over a prolonged period of time, and is measured by agreement with statements like “I am sad all the time and I can't snap out of it”).

#### Understanding America Study

We used USC's *Understanding America Study* described above as a third benchmark of depression. Like for loneliness, it concerned a response to the question “In the past 7 days, how often have you felt depressed?,” with response options 1 = “not at all or less than 1 day,” 2 – “1-2 days,” 3 – “3-4 days,” 4 – “5-7 days.” We used weighted average responses for this question for each of two time periods roughly corresponding to the time windows we gathered prospective survey data (Study 2) – April 15-May 15 and retrospective data (Study 3) – September 15-October 15. We subsequently calculated % change between these average points. We chose not to use this index as a primary marker because it is smaller and less representative of clinical depression compared to the HPS above.

*Affective polarization*

Our primary marker of affective polarization, as well as several other indices below, relied on nationally representative data from Nationscape (Tausanovitsch, Vavreck, Reny, Hayes, Rudkin, 2020; Tausanovitch & Vavreck, 2020). Nationscape is a survey conducting 500,000 interviews of Americans from July 2019 through December 2020, covering the 2020 campaign and election. The survey was in the field since July 10, 2019, and it included interviews with roughly 6,250 people per week. Nationscape samples were provided by Lucid, a market research platform that runs an online exchange for survey respondents. The samples drawn from this exchange match a set of demographic quotas on age, gender, ethnicity, region, income, and education. Respondents were sent from Lucid directly to survey software operated by the Nationscape team. All respondents took the survey online and completed an attention check before taking the survey. The survey was conducted in English. The Nationscape data are weighted to be representative of the American population. The weights are generated using a simple raking technique, as there is little benefit to more complicated approaches (Mercer et al. 2018). Nationscape generated a set of weights for each week’s survey. The targets to which Nationscape is weighted were derived from the adult population of the 2017 American Community Survey of the U.S. Census Bureau. The one exception was the 2016 vote, which was derived from the official election results released by the Federal Election Commission. The Nationscape team weighted the data on the following factors: gender, the four major census regions, race, Hispanic ethnicity, household income, education, age, language spoken at home, nativity (U.S.- or foreign-born), 2016 presidential vote, and the urban-rural mix of the respondent’s ZIP code. Data were also weighted on the following interactions: Hispanic ethnicity by language spoken at home, education by gender, gender by race, race by Hispanic origin, race by education, and Hispanic origin by education. More information on the survey can be found at [www.voterstudygroup.org](http://www.voterstudygroup.org).

Following Boxell, Conway, Druckman, and Gentzkow (*52*), we conceptualized affective polarization via responses to the question stating, “Here are the names of some groups that are in the news from time to time. How favorable is your impression of each group or haven’t you heard enough to say?” and containing responses for “Very favorable,” “Somewhat favorable,” “Somewhat unfavorable,” “Very unfavorable,” and “Haven’t heard enough.” The survey then goes on to ask about the groups: “Republicans” and “Democrats.” We code “Very unfavorable” through “Very favorable” from 0 to 3 respectively and we exclude respondents with other responses. Affective polarization at time t was then defined as:



where Nt is the set of respondents in period t identifying with either the Republican or Democratic party who have valid affect responses for both parties, *wi* is the survey weight, and *Nt* = ∑*i2Nt wi*. Affective polarization measures average feelings towards one’s own party minus average feelings towards the opposing party.

We use the periods overlapping with the time we gathered prospective data (Study 2) and retrospective data (Study 3)—April 23-May 6, 2020 and October 1-October 28, 2020. We restricted survey observations to respondents that give a valid state. We estimated change by examining % change of the October 2020 score relative to the April 2020 scores.

### *Additional marker of affective polarization*

We considered Gallup poll data of presidential approval ratings by party identification as an alternative marker (<https://news.gallup.com/poll/203198/presidential-approval-ratings-donald-trump.aspx>). We obtained a difference score in % of Republican versus Democrat approval ratings and estimated monthly averages for the time period of interest. We did not pursue this marker for primary analyses because it does not fully capture affective attitudes toward members of the other party and hence is not fully in sync with the definition provided to our participants. This said, respective estimates of societal change from this marker were within 1.5% of the estimate of the affective polarization marker from Nationscape.

## *Individualism*

We used the COVID-19 attitudes survey (Neuberg, Varnum, Becker, Ko, Pick, & Wormley, 2020) to estimate societal change in individualism. Using the Prolific survey collection platform, researchers collected two nationally representative samples of US residents to examine the effects of the coronavirus pandemic on a variety of behaviors. The dataset included information gathered at two time points close to the time we gathered prospective estimates (Study 2) and retrospective estimates (Study 3). Upon exclusion of incomplete responses or returned submissions, relevant waves from this project included substantial number of participants surveyed on April 22, 2020 (*N* = 1,510) and on September 23, 2020 (*N* = 805).

To assess individualism, participants rated their agreement with three statements (1 = Strongly Disagree, 9 = Strongly Agree) “It is better for me to follow my own ideas than to follow those of anyone else,” “I enjoy being unique and different from others in many respects,” and “My personal achievements and accomplishments are very important to who I am.” Items came from the established individualism scale (Kim, Sherman, Updegraff, 2016). Given the multi-item nature of the measure, we first inspected measurement invariance by comparing inter-item zero-order correlations at each time point. These preliminary results indicated a variable degree of inter-item association at time 1, .22< *r*s ≤ .35, and at time 2, .16< *r*s ≤ .36. Therefore, we selected two items with highest and largely comparable correlations at both time points, which concerned the first and the second items (time 1: *r =* .35; time 1: *r =* .36). We averaged these items, prior to performing a weighting procedure to ensure the responses represent US population. Like with Nationscape raking procedure, we weighted responses for race, gender, education, age, and political orientation, at each time point. Subsequently, we calculated percentage difference between April and late September estimates as a marker of societal change in individualism.

## *Generalized Trust*

We used the same database as for individualism described above (Neuberg et al., 2020). Researchers measured generalized trust with a single item measure from prior research (Aarøe, Osmundsen, & Petersen, 2016): “Generally speaking, would you say that most people can be trusted or that you can’t be too careful in dealing with people?” with 1 indicating “You can’t be too careful” and 9 indicating “Most people can be trusted.” We applied the raking weighting procedure as described above, and calculated percentage difference between April and late September estimates as a marker of societal change in generalized trust.

## *Traditionalism*

Using the *Nationscape* data described above, we examined weighted % of people choosing the response option “The government should promote traditional family values in our society” instead of option “The government should not promote traditional family values in our society.” We calculated the difference in percent participants agreeing to this question in surveys conducted in April 2020 (April 23-May 6 – same period as Study 2) and October 2020 (October 1-28, 2020 – same period as Study 3) as a marker of change in endorsement of traditionalist values.

### *Additional marker of traditionalism*

To obtain another marker of traditionalism, we used data from an on-going project on societal attitudes during COVID-19 described above (Neuberg et al., 2020).

To assess traditionalism, participants rated their agreement with three statements (1 = Strongly Disagree, 9 = Strongly Agree): “Traditions interfere with progress,”\* “People should respect social norms,” and “Traditions are the foundation of a healthy society and should be respected” (\* = reverse coded). Items came from the established traditionalism scale (Dunwoody & Funke, 2016). Given the multi-item nature of the measure, we first inspected measurement invariance by comparing inter-item zero-order correlations at each time point. There preliminary results indicated variable degree of association at time 1, .33< *r*s ≤ .58, and at time 2, .42< *r*s ≤ .63. Therefore, we selected two items with the highest and largely comparable correlations at both time points, which concerned the first and last items (time 1: *r =* .55; time 1: *r =* .66). We averaged these items, prior to performing a weighting procedure to ensure the responses represent US population. Like with Nationscape raking, we weighted responses for race, gender, education, age, and political orientation, at each time point. Subsequently, we calculated percentage difference between April and late September estimates. Because the time frame for this estimate was a month shorter than for Nationscape data, we chose to treat this estimate as a secondary benchmark. We note that the estimates of societal change in traditionalism were very similar across both primary and secondary markers (within 5% change).

## *Violence*

To assess changes in violent crime between April and October of 2020, we relied on data from the Pandemic, Social Unrest, and Crime in U.S. Cities November 2020 report, prepared for the Council of Criminal Justice (Rosenfeld & Lopez, 2020; https://cdn.ymaws.com/counciloncj.org/resource/resmgr/covid\_commission/Crime\_in\_US\_Cities\_-\_October.pdf). It tracks ten criminal offenses since January 2017 for 28 U.S. cities, spanning a population of 866,000 people. Out of 10 only four crimes as these met the definition of violent crime provided to participants in Studies 1-3, as they specifically violent in nature: aggravated assault, homicide, gun assault, and domestic violence. We calculated a percentage difference score for each crime type and then created a composite by averaging all four.

## *Attitudes toward climate change*

Using *Nationscape* data described above, we examined weighted % of people agreeing to the question “We’d like to know whether you would cap carbon emissions to combat climate change,” with response options “agree,” “disagree,” “not sure.” We calculated the difference in percent participants agreeing to this question in surveys conducted in April 2020 (April 23-May 6 – same period as Study 2) and October 2020 (October 1-28, 2020 – same period as Study 3) as a marker of attitudes toward climate change.

## *Charitable Giving*

We obtained estimates from national-level surveys conducted by Giving Tuesday to estimate philanthropic sentiment (<http://data.givingtuesday.org/>), assessed as part of the Fundraising Effectiveness Project, as communicated to our team by the chief data officer of Giving Tuesday (Rosenbaum, 2020). The Fundraising Effectiveness Project and the Growth in Giving database (created in 2012) are administered by the Association of Fundraising Professionals. The Growth in Giving database is the world’s largest public record of donation activity, with more than 204 million donation transactions, and is continuously updated by leading fundraising software thought leaders (in alphabetical order) Bloomerang, DonorPerfect, and NeonCRM. Additional partners include the 7th Day Adventists, The Biedermann Group, DataLake Nonprofit Research, and DonorTrends (a division of EveryAction).

# Supplementary benchmark indices

We initially planned to include two additional markers concerning explicit and implicit prejudice toward minorities. The Project Implicit data source is not representative of the US population at large and relies on different on-line platforms through which participants are recruited. Because the topic concerned prejudice, and Black-Lives Matter protests in the summer let many outlets and diversity programs directing persons interested in learning about empathy and prejudice to the website, the representativeness of the data could be viewed as compromised. Due to the abundance of caution, we decided not to report this benchmark estimate in the main text. For the sake of transparency, we report all relevant analyses in this supplement.

This data came from the Project Implicit website (http://implicit.harvard.edu) which has collected continuous data concerning explicit stereotypes and implicit associations from a heterogeneous pool of volunteers (50,000 - 6,000 unique tests on each of these categories per month). Further details about the website and test materials are publicly available at <https://osf.io/t4bnj>. Recent work suggests that Project Implicit data can provide reliable societal estimates of consequential outcomes (*53*, *54*) and when studying cross-temporal societal shifts in U.S. attitudes (*55*). Despite the non-representative nature of the Project Implicit data, recent analyses suggest that bias scores captured by Project Implicit are highly correlated with nationally representative estimates of explicit bias, *r*  = .75, indicating that group aggregates of the bias data from Project Implicit can reliably approximate group-level estimates (*54*). To correct possible non-representativeness, we applied stratified weighting to the estimates, as described below.

Because of possible selection bias among the Project Implicit participants, we used a raking procedure similar to the one employed by Nationscape. We weighted monthly scores based on their representativeness of the demographic frequencies in the U.S. population (age, race, gender, education; estimated biannually by the U.S. Census Bureau; <https://www.census.gov/data/tables/time-series/demo/popest/2010s-national-detail.html>). Further, we adjusted weights based on political orientation (1 = “strongly conservative;” 2 = “moderately conservative;” 3 = “slightly conservative;” 4 = “neutral;” 5 = “slightly liberal;” 6 = “moderately liberal;” 7 = “strongly liberal”), using corresponding annual estimates from the General Social Survey. With the weighting values for each participant, we computed weighted monthly means for each attitude test. These procedures ensured that weighted monthly averages approximated the demographics in the U.S. population.

To correct for possible variability in monthly scores due to fluctuations in sources of participant recruitment, we further applied 30% loess smoothing function across monthly estimates from 2018 through 2020, prior to calculating % change scores between April (April 1 – 30) and October (October 1 – 31), 2020. This approach allows to correct for month-specific selection biases.

#### Explicit prejudice

For explicit attitude scores, participants provided ratings on feeling thermometers towards Asian-Americans and European Americans (to assess Asian-American bias), and White and Black Americans (to assess racial bias). We calculated relative explicit bias as the difference in responses to minority and majority groups on feeling thermometers (for Asian-American and Black Americans). The sample was further restricted to include only respondents from the United States to increase shared cultural understanding of attitude categories. The sample was also restricted to include only respondents with complete demographic information on age, gender, race/ethnicity, and political ideology. After raking and smoothing, we averaged responses across both estimates for an overall measure of bias toward ethnic minorities.

#### Implicit prejudice

Implicit attitude scores were computed using the revised scoring algorithm of the implicit association test (IAT) (*56*). The IAT is a computerized task comparing reaction times to categorize paired concepts (in this case, social groups, e.g., Black American vs. European American and Asian American vs. European American) and attributes (in this case, valence categories, e.g., good vs. bad). Average response latencies in correct categorizations were compared across two paired blocks in which participants categorized concepts and attributes with the same response keys. Faster responses in the paired blocks are assumed to reflect a stronger association between those paired concepts and attributes. In all tests, positive IAT *D* scores indicate a relative preference for the typically preferred group.

Respondents whose scores fell outside of the conditions specified in the scoring algorithm did not have a complete IAT *D* score and were therefore excluded from analyses. Restricting the analyses to only complete IAT *D* scores resulted in an average retention of 92% of the complete sessions across tests. The sample was further restricted to include only respondents from the United States to increase shared cultural understanding of attitude categories. The sample was restricted to include only respondents with complete demographic information on age, gender, race/ethnicity, and political ideology. We averaged responses across both estimates for an overall measure of bias toward ethnic minorities.

*Table S3.*Sources for benchmarking accuracy for estimates of societal change.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dimensions | Source | April/MayPeriod | OctoberPeriod | April/MayN | OctoberN | Observed Change |
| Life  Satisfaction | Gallup Panel – COVID-19 Survey | Apr 23 - May 5 | Oct 14 - Oct 26 | 10,058 | 2,667 | 1% |
| ICL/Yougov – Covid 19 Behaviour Tracker | Apr 27 – May 3 | Sep 14 – Sep 20 | 1,003 | 966 | -3.2% |
|  | Twitter-based estimates | Apr 01 - Apr 30 | Oct 01 - Oct 31 | - | - | -0.3% |
| Loneliness | Understanding America Study | Apr 23 - May 5 | Oct 14 - Oct 26 | 2,872 | 5,498 | -4% |
| Gallup Panel– COVID - 19 Survey | Apr 23 - May 5 | Oct 14 - Oct 26 | 10,058 | 2,667 | -3% |
| Depression | Household Pulse Survey | Apr 23 - May 5 | Oct 14 - Oct 26 | 69,316 | 76,034 | 3% |
|  | Understanding America Study | Apr 23 – May 05 | Oct 14 – Oct 26 | 5,600 | 5,498 | -6% |
|  | Gallup Panel - COVID-19 Survey | May 11- May 30 | Oct 14 – Oct 26 | 11,175 | 2,666 | 0% |
| Affective  Polarization | Nationscape | Apr 23 - May 6 | Oct 01 - Oct 28 | 8,108 | 26,000 | 4% |
| Gallup Poll | Apr 01 - Apr 30 | Oct 01 - Oct 31 | - | - | 5.5% |
| Individualism | COVID-19 Attitudes Survey | Apr 22 – Apr 24 | Sep 23 – 28 | 1,510 | 805 | 3.4% |
| Generalized Trust | COVID-19 Attitudes Survey | Apr 22 – Apr 24 | Sep 23 – 28 | 1,510 | 805 | -1.3% |
| Traditionalism | Nationscape | Apr 23 - May 6 | Oct 01 - Oct 28 | 13,058 | 26,222 | 1% |
| COVID-19 Attitudes Survey | Apr 22 – Apr 24 | Sep 23 – 28 | 1,510 | 805 | 5% |
| Violence | Pandemic, Social Unrest, and Crime in U.S. Cities (November 2020) | Apr 01 - Apr 30 | Oct 01 - Oct 31 | - | - | 24% |
| Climate Change | Nationscape | Apr 23 - May 6 | Oct 01 - Oct 28 | 13,069 | 26,321 | -1% |
| Charitable Giving | Giving Tuesday | Mid May | Mid November | Unknown | Unknown | 4% |
| *Explicit Prejudice* | *Project Implicit* | *Apr 01 – Apr 30* | *Oct 01 – Oct 31* | *70,700* | *239,740* | *- 51%* |
| *Implicit Prejudice* | *Project Implicit* | *Apr 01 – Apr 30* | *Oct 01 – Oct 31* | *70,700* | *239,740* | *- 6.2%* |

# **Results**

## **Deliberation check**

To examine whether participants relied on intuition or spent a substantial amount of time reflecting on predictions, we examined descriptive statistics for overall study completion time among participants who completed the whole survey. In Study 1, social scientists typically took 11 *min* (*Md*; M = 23.80; 95%*CI* [16.22, 31.38]) in total. In Study 2, they spent 14 *min* (*Md*; *M =* 20.72 95%*CI* [17.76, 23.67]) in total, whereas lay people spent 12 *min* (*Md*; *M* = 14.47; 95%*CI* [13.60, 15.33]). Consequently, participants typically spent less than a minute making predictions for each of the eleven (Study 1) / fifteen (Study 2) domains. This suggests that most of the participants’ predictions made were not the result of protracted reflection. In Study 3, we collected domain-specific times of completion, rather than simply completion time for the whole survey. By this measure, social scientists took on average between 10 and 20 *seconds* (*M* = 14.48; 9.05 < 95%*CI* ≤ 21.63) to read relevant descriptions and subsequently answer two questions per domain. In comparison, lay people took between 11 and 25 *seconds* per domain (*M* = 16.41; 10.20 < 95%*CI* ≤ 26.74). Thus, it appears that retrospective estimates were likewise not the result of protracted reflection, nor did social scientists deliberate for longer amounts of time.

## **Description of predictions by domain**

We analyzed this data for general predictions about change and whether those changes would return to baseline within the next two years. We also assessed whether consensus and heterogeneity of predictions.

We begin by focusing on Study 1, which took place at the beginning of April 2020. Figures S2-S3 indicate predictions for change across different domains. Social scientists predicted the largest changes for depression, political polarization, out-group prejudice, and life satisfaction (Figure S4, Panel A). Notably, for three of the eleven domains (traditionalism, generalized trust, delay of gratification) social scientists’ predictions for April 2022 were not statistically different from the baseline in April 2020, *ps* > .072; thus, for these domains social scientists predicted a full return to baseline. Social scientists predicted the remaining domains to remain significantly altered two years later, *ps* < .028 (see Table S8[[5]](#footnote-6)).

Did social scientists’ prospective intuitions shift over short periods of time? To assess this question, we turn to Study 2, conducted at the beginning of May 2020. As in Study 1, social scientists predicted a significant degree of societal change for each of these domains, 2.76 < *t*s ≤ 16.86, *p*s < .007 (see Table S9 and Figure S3 for comparison of April and May 2020 estimates by social scientists), except for delay of gratification (replicating Study 1), *p* = .100. Social scientists predicted that traditionalism, birth rate, delay of gratification, and charitable donations would return to May 2020 baseline in two years, while for the remaining domains social scientists expected a significant difference two years from May 2020, *ps* < .033 (see Table S12 and analyses in the online supplement). Notably, a comparison of estimates across Studies 1 and 2 revealed highly similar patterns of forecasts, both for temporal trends (see Figs. S2-S3), and rank-order of trends (compare Panels A-B in Figure S4), Bayes Factor (null / alternative, BF01) > 6 (see Table S10 for frequentist results). We observed only four exceptions: different linear trends for individualism, birth rate and political polarization and quadratic trends for birth rates and life satisfaction, Bayes factor < 2.

## **Consensus among scientists’ predictions**

Social scientists were remarkably consistent in their predictions: for each domain, 95 out of 100 estimates were within 3%-3.8% points of the mean (see Figures S2-S3, and Tables S5 and S11), with little evidence of a bimodal distribution of responses (see Figure S5). Predictions also did not significantly differ in their estimates as a function of academic discipline (psychology vs. other sciences), *p*s > .131 (see Table S6 and S11), country of residence (US vs. non-US; see Tables S7 and S11), or degree of cultural distance from the US (see Figures S6-S7). We also found little evidence that degree of scientific experience (tenured faculty vs. non-tenured faculty vs. graduate students/others; Figure S8) or broad field of expertise (mental health vs. social psychology vs. others) significantly impacted predictions (see Figure S9), *ps* > .999. We observed two exceptions: Graduate students/Post-docs predicted more extreme political polarization across all time points compared to faculty, 3.78 < z ≤ 5.58, *ps* < .008, and greater change in charitable giving compared to tenured faculty in six months, z = 4.58, *p* < .001.

Finally, predictions were not significantly different for domains in which social scientists indicated they had domain-related expertise, Domain × Expertise × Time interactions, -3.36 < *t*s ≤ 2.02, *p*s > .06, *ns,* except for delay of gratification and violence, where experts predicted less change than non-experts, *ps*adj < .040. Domain-specific expertise also did not play a role for overall magnitude of forecasted change, 0.03 < *t*s ≤ 2.82, *p*adj = .076.

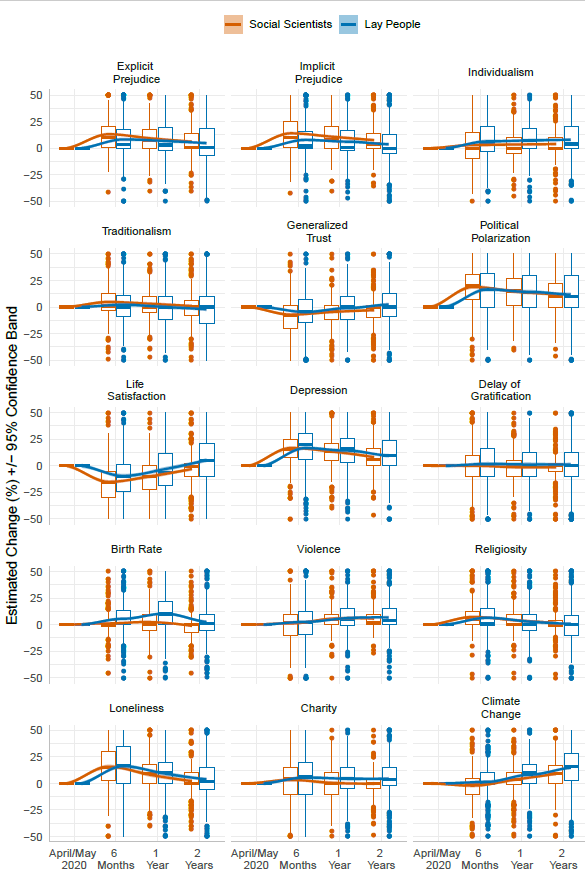
## **Changes in Predictions**

To determine whether forecasted trends were significantly different from April and May 2020 baseline levels, we first identified the time point (6, 12 or 24 months) with the largest absolute change for each domain. Next, we ran single sample t-tests with *mu* set to 0 and adjusted *p*-values using Benjamini-Hochberg correction (Table S4 & S9). Participants predicted significant changes across all domains except delay of gratification for both April and May samples.

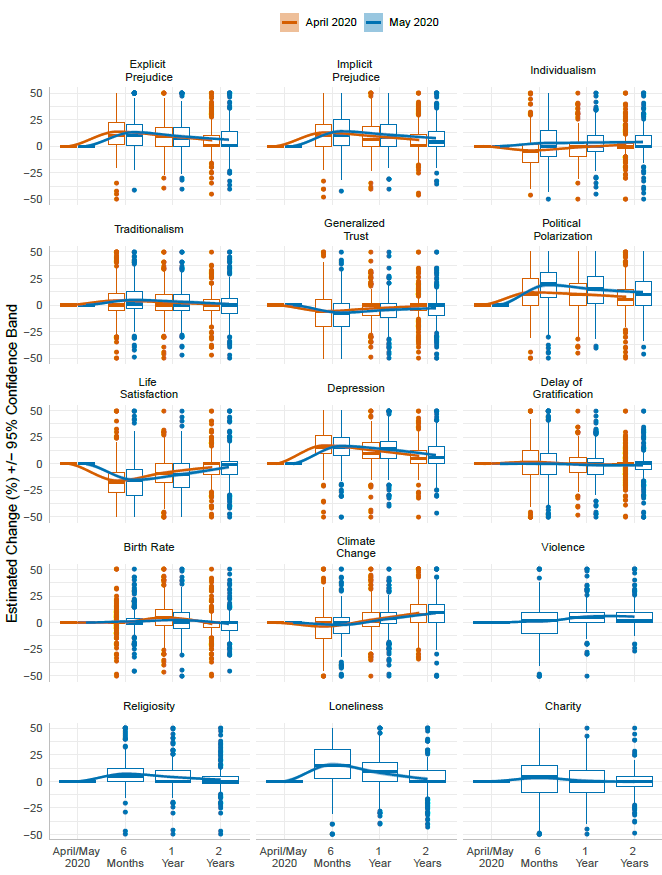
To assess whether participants believed changes would be long-lasting, we estimated significant differences between April and May 2022 estimates and April and May 2020 baseline separately for each domain by running a single sample *t*-test with *mu* set to zero and adjusted *p*-values using Benjamini-Hochberg correction (Tables S8-S12). Out of 11 three domains (traditionalism, generalized trust, and delay of gratification) were not significantly different in April 2022 from baseline. The remaining eight (implicit and explicit prejudice, individualism, political polarization, life satisfaction, depression, birth rate and climate change), however, were statistically different from April 2020 levels. Similarly, four domains (charity, traditionalism, delay of gratification, birth rate) were not statistically different from baseline in May 2022, while the remining 11 (depression, implicit and explicit prejudice, violence, individualism, generalized trust, life satisfaction, religiosity, loneliness, political polarization, climate change) were.

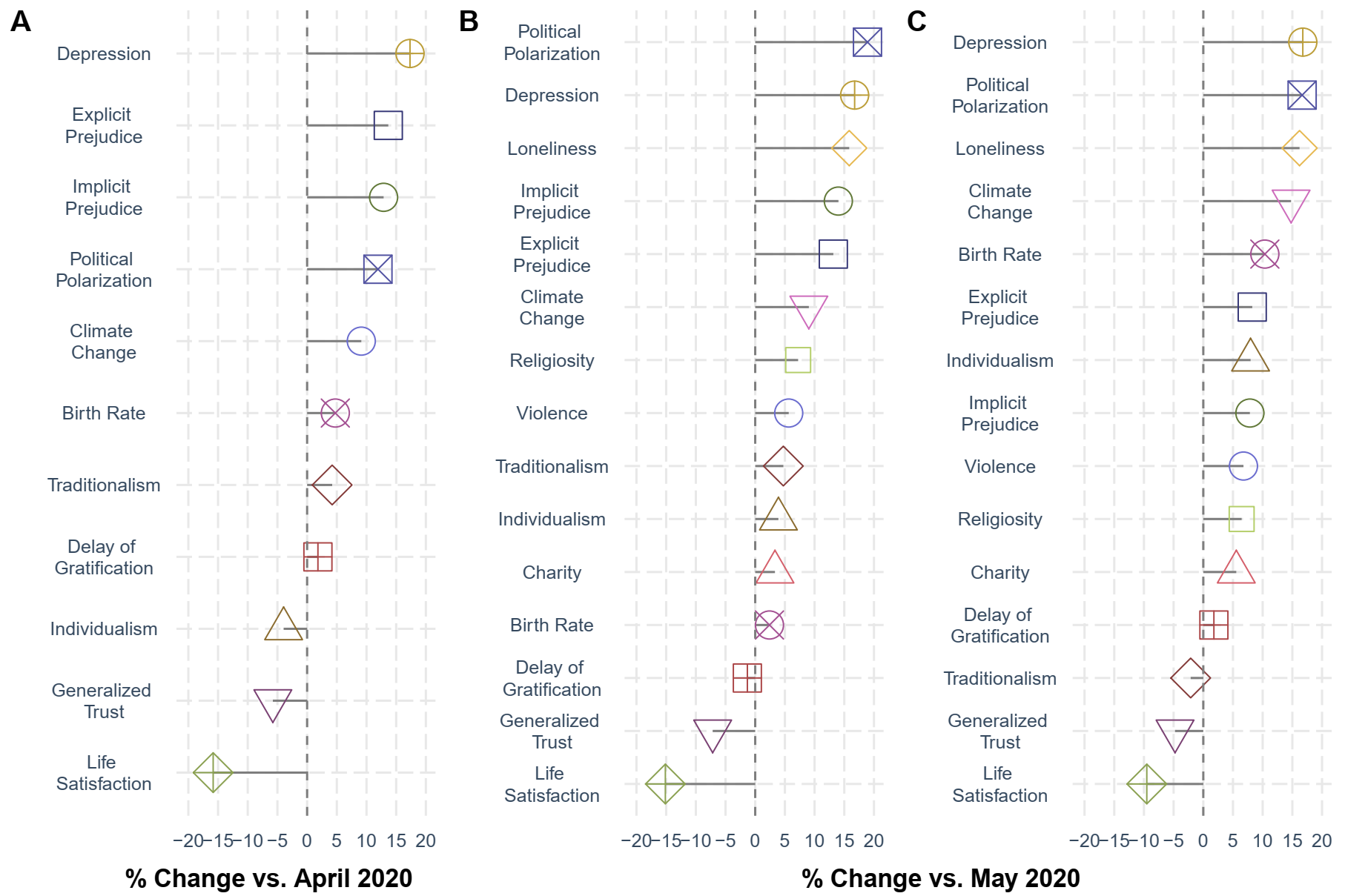
### **Social Scientists**

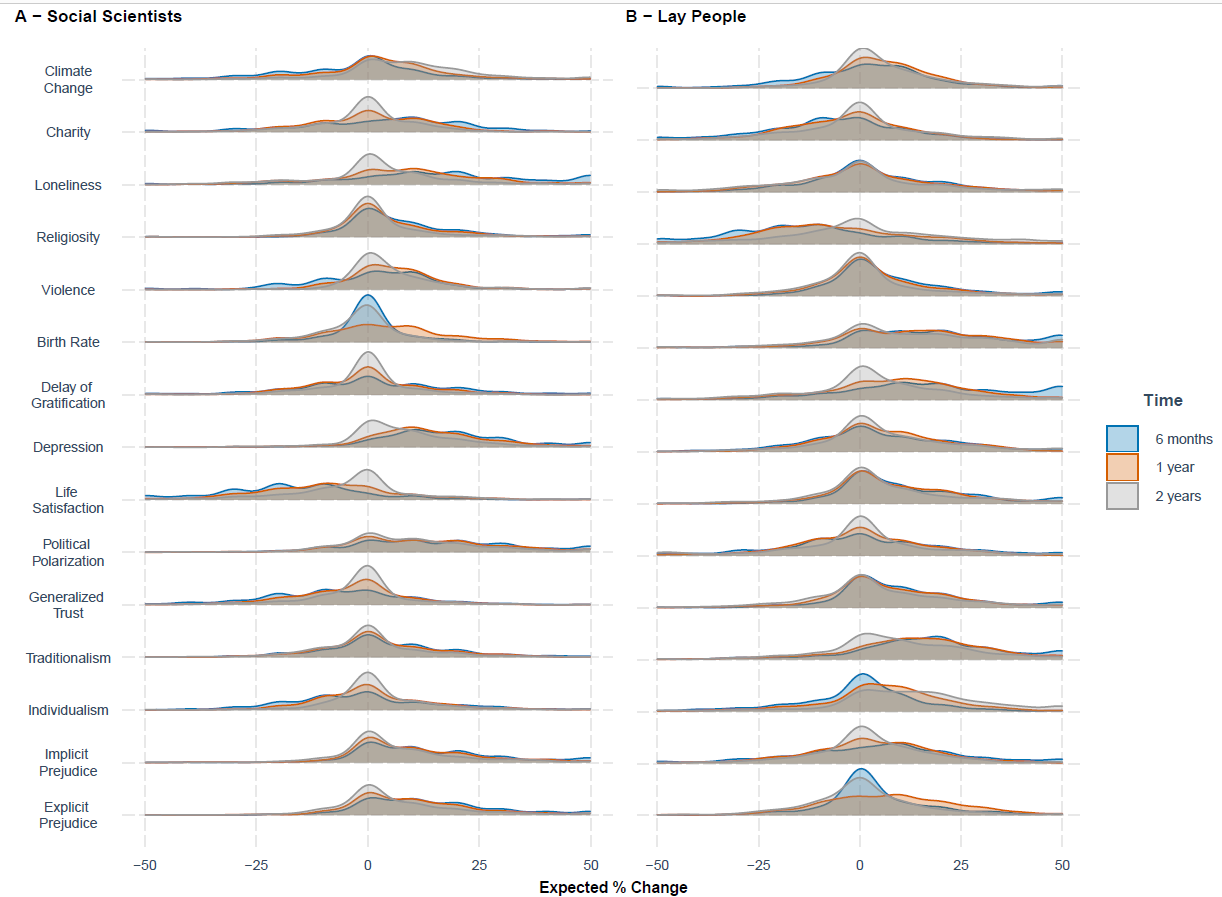
To examine social scientists’ forecasts, we calculated means and confidence intervals for April and May 2020 by first fitting a 2-way linear mixed model with time and domain as level 1 predictors along with a 2-way interaction term for them. Observations were nested in participants. We extracted means, standard errors, and 95% confidence intervals from the model (Tables S5-S11). Descriptive means showed substantial heterogeneity by domain across the two samples, not only in terms of magnitude but also direction. At the same time, confidence intervals around domain-specific estimates were surprisingly small, showing little variation by domain with 95% of the scores falling within three (April) and four (May) percent of the mean. Moreover, as Figure S5 shows, results indicate no indication of bimodal distribution at any time point.

**

*Figure S2*. Predictions for change across 15 societal domains from social scientists and lay people. Graphs indicate boxplots for a given time-point forecast (half a year, year, two years from April/May 2020) and loess line of best fit across time points with 95% confidence bands around the estimate.

*Figure S3*. Social scientists predicting societal change in April vs. May, 2020. Graphs indicate boxplots for a given time-point forecast (half a year, year, two years from April/May 2020) and loess line of best fit across time points with 95% confidence bands around the estimate. Positive numbers refer to positive change, whereas negative—a negative change.

 *Figure S4.* Ranking of domains based on magnitude and direction of predicted societal change, as estimated by social scientists over two years from April 2020 (Panels A) and May 2020 (Panel B), as well as lay people’s predictions over the same time period (Panel C).

**

*Figure S5.* Density distribution of predictions of societal changes by prediction timepoint (6 months, 1 year and 2 years). Panels show results of forecasts provided by social scientists (left) and lay people (right) in May 2020. Results indicate no indication of bimodal distribution at any time point. Positive numbers refer to positive change, whereas negative—a negative change.

#### Role of cultural distance from the US

To examine the effect of culture on predictions of societal change in the US, we estimated social scientists’ “cultural distance” from the US based on their country of residence. Figures S6 (April 2020 sample) & S7 (May 2020 sample) show results of the estimates from a 3-way linear mixed model using *lme4* package and obtaining p-values via *jtools* package in *R*. We set Time and Dimension as level 1 predictors and Cultural Distance as level 2 predictor, while also including age, gender (men vs. women dummy variable), and faculty type as covariates, nesting observations in participants. Results indicate no significant difference in predictions from the US vs. elsewhere after corrections (Tables S7-S11).

*Figure S6*.April 2020 predictions by social scientists for change across 11 societal domains in standard deviation units of Cultural Distance from the US. Graphs indicate means and 95% confidence intervals for given time-point forecast (half a year, year, two years from April 2020). Positive numbers refer to positive change, whereas negative—a negative change.

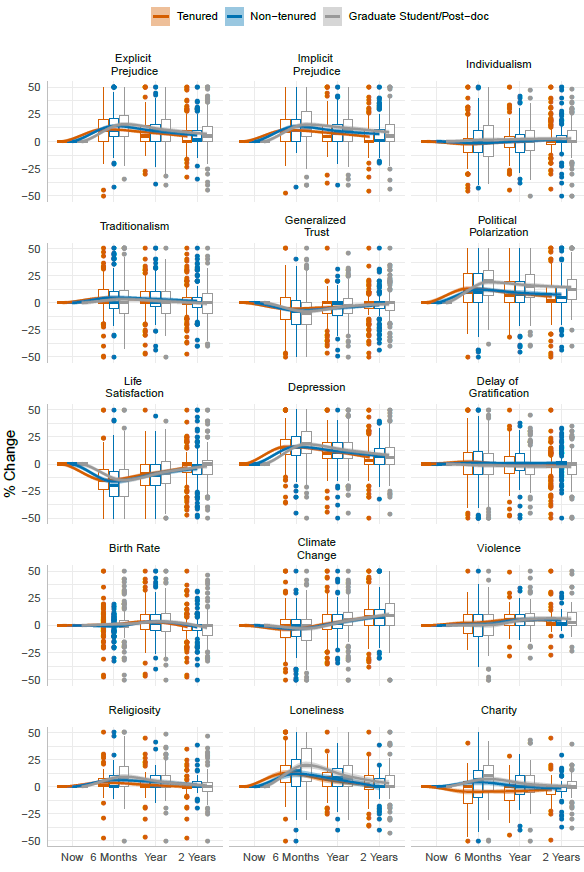


Figure S7. May 2020 predictions by social scientists for change across 15 societal domains in standard deviation units of Cultural Distance from the US. Graphs indicate means and 95% confidence intervals for given time-point forecast (half a year, year, two years from May 2020). Positive numbers refer to positive change, whereas negative—a negative change.

#### Role of expertise

We were interested in the effects of *expertise level* in social sciences on predictions. We operationalized expertise level by categorizing our sample into three clear-cut categories: a) tenured; b) non-tenured; c) graduate students/post-doc. Tenured faculty consists of social scientists who chose the “tenured faculty” option as their current position at university/college. Non-tenured faculty consists of those who chose “nontenured faculty” and “adjunct professor” as their current position. Finally, the Grad Students/Post-doc group is comprised of those who selected “graduate students” and “post-docs” as their current position respectively. We fit a 3-way MLM with expertise level, time, and domain as well as all two- and three-way interaction terms predicting participants’ forecasts. In addition, we controlled for age, university size, affiliate organization type, gender and country of residence to rule out the possibility that demographic variability between the two samples could be responsible for observed results. Observations were nested in participants. In addition, to control for number of comparisons we used the Benjamini-Hochberg correction. We observed no significant differences in predictions between the three groups of experts, *ps* > .058 except for two domains. Tenured faculty forecasted lower levels of change at 6 months for charity, compared to non-tenured and graduate students/post-docs. Similarly, graduate students/post-docs forecasted greater political polarization then both tenured and non-tenured faculty for all time points.

Participants listed their areas of expertise, which we sorted into one of three categories: Social/Personality Psychology, Mental Health and Other, with Other encompassing all other areas of psychology and other social and life sciences. We fit a 3-way MLM model with Time and Dimension as level 1 predictors and Academic Discipline as level 2 predictor (a factor decomposed into two dummy variables) as well as all 2 and 3-way interaction terms. The model also included the following socio-demographic covariates: age, university size, organization, gender, and country of residence. Fitted means and confidence intervals were extracted from the model and were then used in plotting (Figure S9). To test for differences between expertise categories, we conducted pairwise comparisons in *R* using *emmeans* and controlled for number of tests using the Benjamini-Hochberg correction. The differences between areas of expertise at each time point and domain were not statistically significant, *ps* > .55.



*Figure S8.* Social scientists’ predictions for change across 15 societal domains by Faculty Type. Data is pooled across Studies 1-2. Graphs indicate boxplots for a given time-point forecast (half a year, year, two years from April/May 2020) and loess line of best fit across time points with 95% confidence bands around the estimate. Positive numbers refer to positive change, whereas negative—a negative change.



Figure S9. Experts’ predictions for change across 15 societal domains by area of expertise from April/May 2020. Graphs indicate means and 95% confidence intervals for a given time-point forecast (half a year, year, two years from April/May 2020). Positive numbers refer to positive change, whereas negative—a negative change

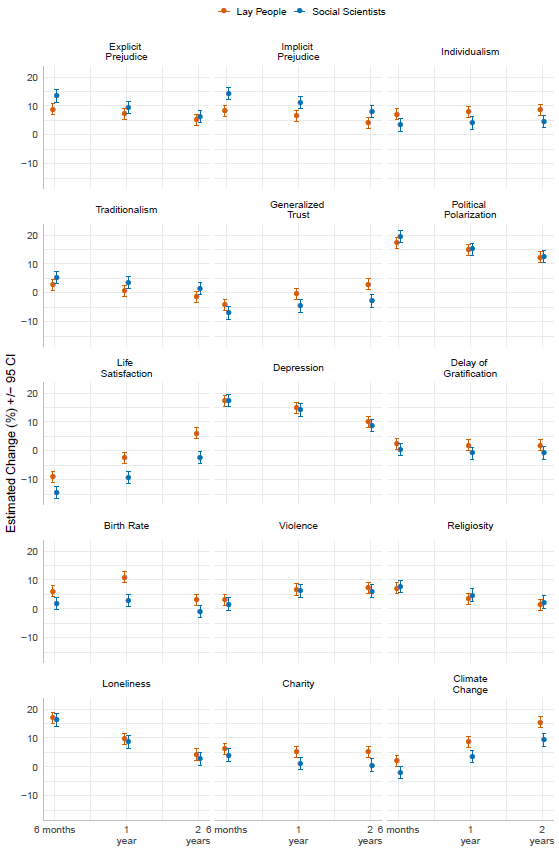
Psychology group included experts who selected either “Psychology” or “Neuroscience” as their main field of research. We combined the remaining choices into “Other disciplines.” Then, we calculated differences between the two conditions for April 2020 and May 2020 forecasts by fitting a 2-way interaction MLM with time and domain as level 1 predictors and field of research as level 2 along with all 2 and 3-way interaction terms. We included the following covariates in the model: age, gender, and level of expertise. We then performed pairwise comparisons between psychology and other disciplines in *R* using *emmeans* package for each domain and each time point (Table S6 & Table S11). The results indicated no significant difference in the predictions made by psychologists versus those in other disciplines, after applying Benjamini-Hochberg correction to control for false discovery rate, *ps* > .10.

#### Lay versus Academic Predictions

To test the difference between linear and quadratic slopes of social scientists’ forecasts in April and May 2020, we fit a 2-way linear mixed model with time and time squared as level 1 predictors and sample dummy variable (April vs. May) as a level 2 predictor, along with the time × sample and time2 × sample interaction terms. We treated Time as a continuous variable and mean centered before generating a quadratic term to avoid multicollinearity. We obtained *t* and *p* values for the interaction term from these models, employing Benjamini-Hochberg correction to adjust for number of pairwise tests between samples for each domain. We did not find any significant differences between slopes, *p*s > .150, except for individualism (linear), political polarization (linear) and birth rate (quadratic; Table S10). We also fit a second model using the same procedure as for the first, except we compared lay people’s and social scientists’ forecasts in May 2020. The results indicated no significant difference in lay people’s versus social scientists’ prediction except for explicit prejudice (linear trend) and birth rate (quadratic trend; Table S13).

**Retrospective Estimates**

In October and November of 2020, we asked participants to look back and estimate how much they thought certain domains had changed in the last six months. To test whether a lack of differences between social scientists and lay people can be attributable to demographic differences, we fit a 2-way linear mixed model with sample (lay people vs. academics), domain, and their interaction as predictors of estimates, while controlling for ethnicity, political affiliation, age, gender, and income, and nesting observations in participants. Figure S15 present estimates from these models, showing close to identical results for models with and without covariates.



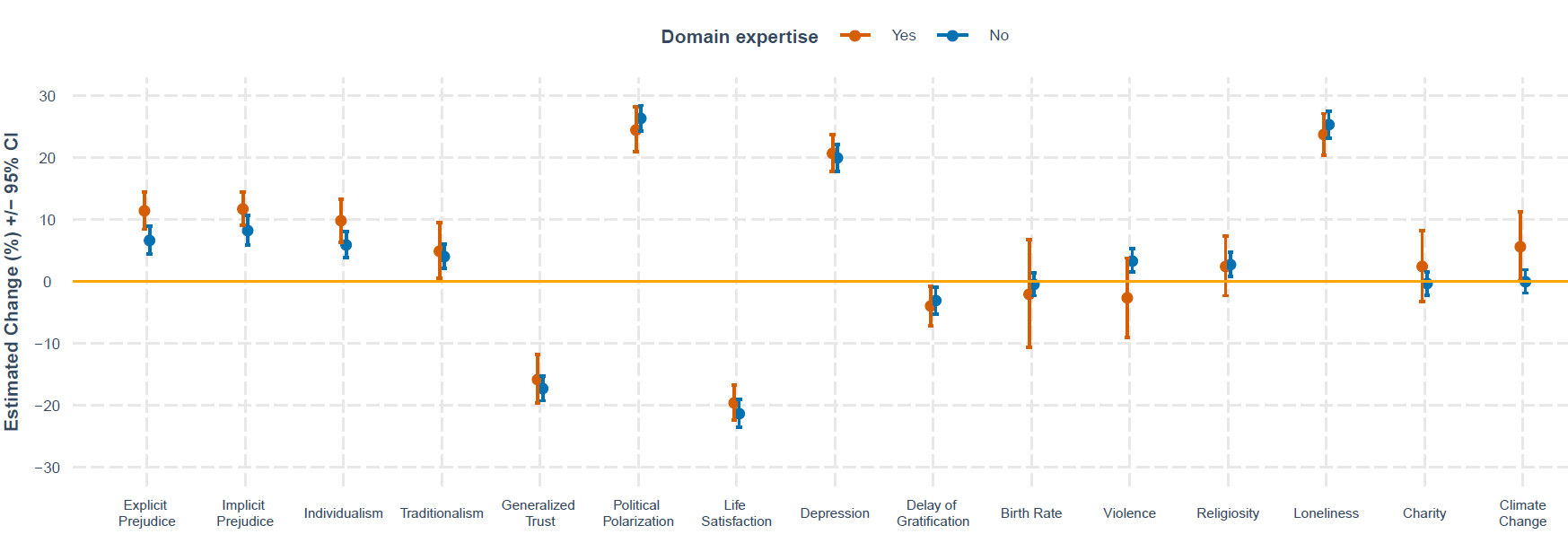
*Figure S10*. Predictions for change across 15 societal domains from social scientists and lay people controlling for age, gender, household income, and political orientation. Graphs indicate means and 95% CI for a given timepoint forecast (half a year, year, two years from May 2020).

### *Social scientists vs. lay people*

To test the difference between lay people’s and social scientists’ retrospective estimates in October/November 2020, we fit a 2-way linear mixed model with domain as level 1 predictor and sample dummy variable (lay people vs. social scientists) as level 2 predictor, along with the domain x sample interaction term. Then, we performed pairwise comparisons between lay people and social scientists in *R* using *emmeans* package for each domain, and subsequently used Benjamini-Hochberg method for false discovery rate correction to account for number of tests (Table S14). Lay people and social scientist only significantly disagreed in their estimates in five out of fifteen domains: generalized trust, delay of gratification, violence rates, individualism and depression.

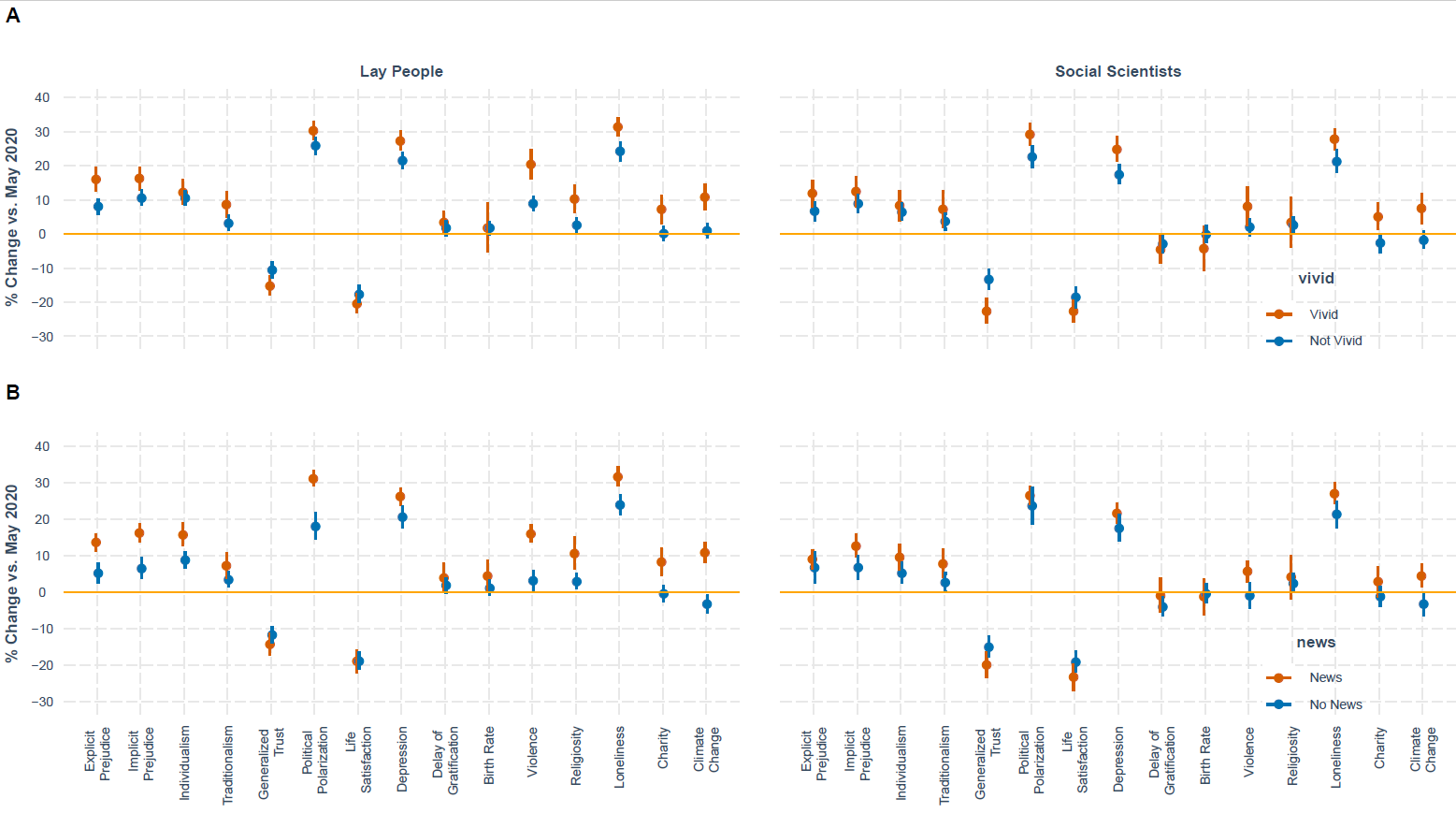
In addition to this frequentist approach, we conducted a Bayesian analysis to test for statistical equivalence between social scientists’ and lay people’s retrospective estimates. The models were fit using *stan\_glm* function from *rstanarm* package in *R*. For each domain, we fit a Bayesian linear mixed model with sample as the sole predictor and observations nested in participants, with normally distributed priors, *N* ~ (0, 5), for predictor and intercept. Bayes Factor was computed in favor of the null hypothesis, such that there is no difference between lay and expert retrospective estimates (BF01; Table S15).

### *Domain expertise*

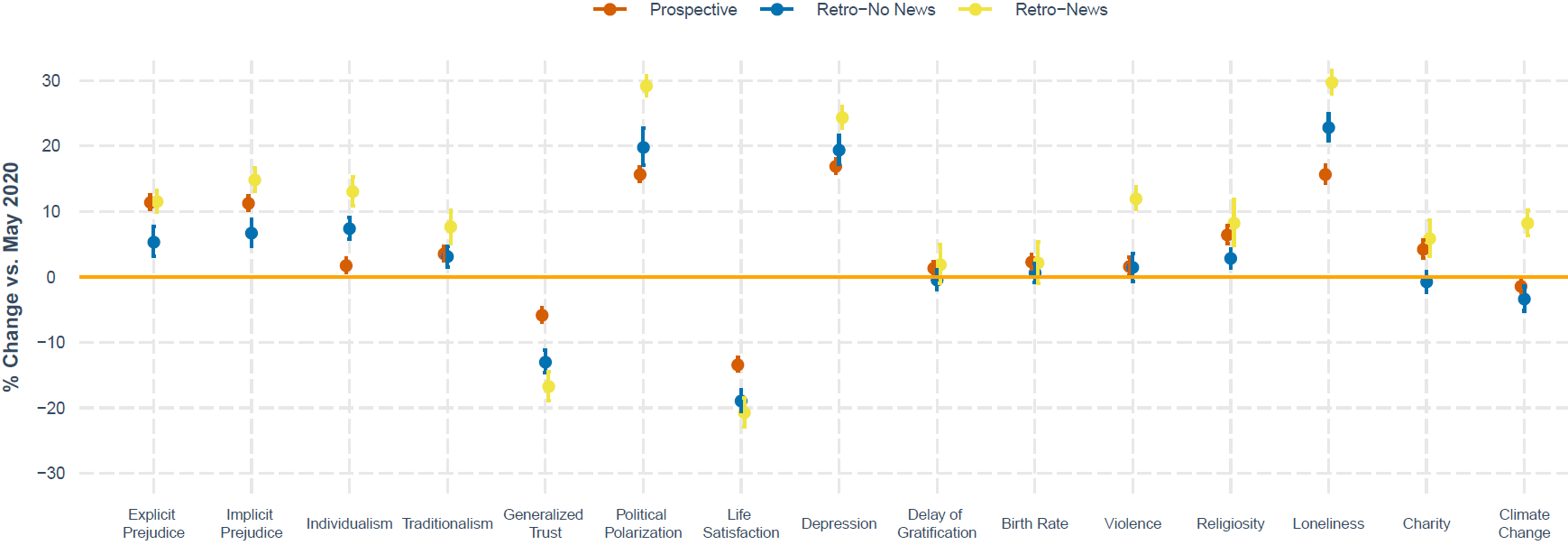
To test whether domain expertise affected social scientists’ retrospective estimates we fit a 2-way linear mixed model with domain expertise (yes/no) × domain interaction predicting retrospective estimates. Observations were nested in participants. Fitted means and CIs were then extracted from the model and plotted (Figure S11). Having domain expertise had no significant impact upon retrospective estimates of change (see Table S16).

*Figure S11*. Retrospective mean estimates of change across 15 societal domains for social scientists by domain expertise.

### *Vividness of memories and news exposure*

We examined how two characteristics: i) vividness of memories; ii) news exposure impacted retrospective assessments of scientists and lay people. We fit two linear mixed models to test whether vivid memories and news reports affected retrospective estimates for both social scientists and lay people by domain. In both models we included main effects of sample (social scientists vs. lay) and either vividness (vivid/not vivid) or news exposure (news/no news) and a sample × vividness / news exposure interaction term. Observations were nested in participants. Figure S12 shows fitted means and 95% CIs. We performed pairwise comparisons (news/no news or vivid/not vivid) with a subsequent Benjamini-Hochberg false discovery rate correction. As Table S17 shows, we observed systematic effects of vividness and news exposure resulting in more extreme estimates among both groups. Figure S13-S14 provides a comparison of prospective and retrospective mean estimates of change across 15 societal domains for social scientists, separately for scientists who recalled vs. did not recall news reports and who reported vs. did not report vivid memories on the topic of the respective forecasting domain.

*Figure S12*. Retrospective mean estimates of change across 15 societal domains for social scientists by memory vividness (Panel A; Vivid/Not Vivid) and news exposure (Panel B; News/No News). Error bars indicate 95% confidence interval.

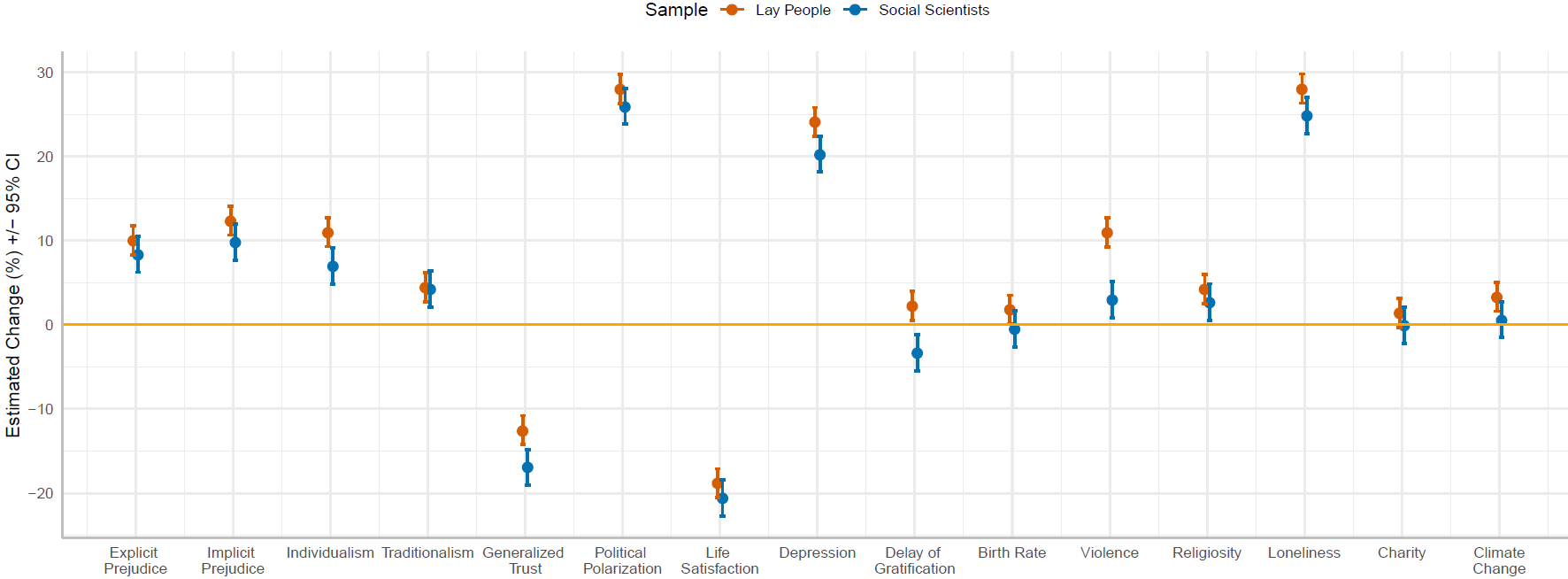


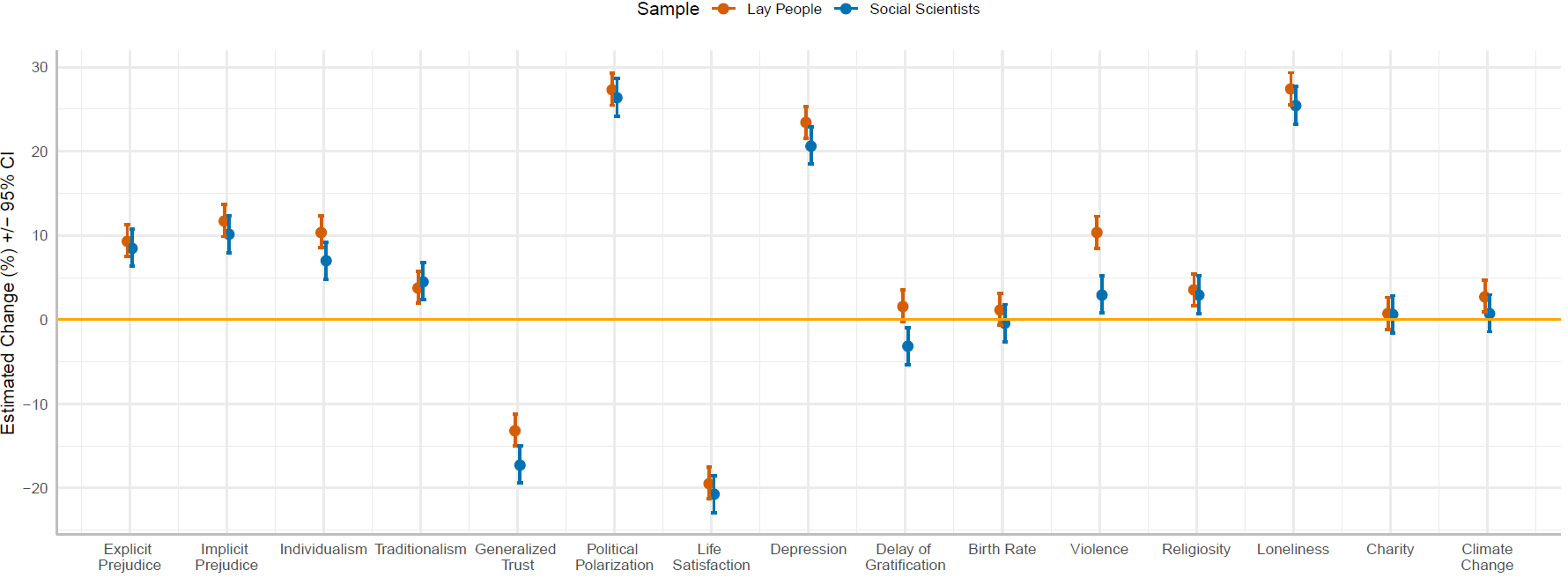
*Figure S13*. Comparison of prospective and retrospective mean estimates of change across 15 societal domains for social scientists by news. Error bars indicate 95% confidence interval.

### *Prospective vs. retrospective estimates*

To test whether retrospective estimates were less extreme than prospective we fit two linear mixed models: one comparing retrospective against prospective estimates for social scientists and the other for lay people. In both models we included estimate type (prospective vs. retrospective), domain, and estimate type × domain interaction, while nesting observations in participants. Then, we performed domain-wise pairwise comparisons between prospective and retrospective estimates, applying Benjamini-Hochberg method for false discovery rate correction. Figures S15 and Table S18 show estimates of accuracy and 95% CIs. Figure S15 & Table S19 shows results of equivalent analyses with demographic covariates (ethnicity, political affiliation, age, gender, and income), suggesting that including covariates leads to largely identical results to those without covariates.

*Figure S14*. Comparison of prospective and retrospective mean estimates of change across 15 societal domains for social scientists by memory vividness. Error bars indicate 95% confidence interval.





*Figure S15*. Retrospective mean estimates of social change (April vs. October/November 2020). Top: without covariates; Bottom: controlling for ethnicity, political affiliation, age, gender, and income. Error bars show 95% CI.

## **Accuracy**

To test the accuracy of social scientists’ and lay people’s prospective and retrospective estimates we ran a series of one sample t-tests with *mu*set to the accuracy benchmark level retrieved from nationally representative samples by domain, with a subsequent Benjamini-Hochberg correction. As Table S20 shows, for most domains prospective and retrospective estimates of change were significantly different from actual changes. As Figure S16 below shows, for most estimates the majority of prospective and retrospective estimates were inaccurate in the direction of estimated societal change (< 40%), and only for a few domains the majority was accurate in estimating direction of societal change (depression, trust, political polarization).

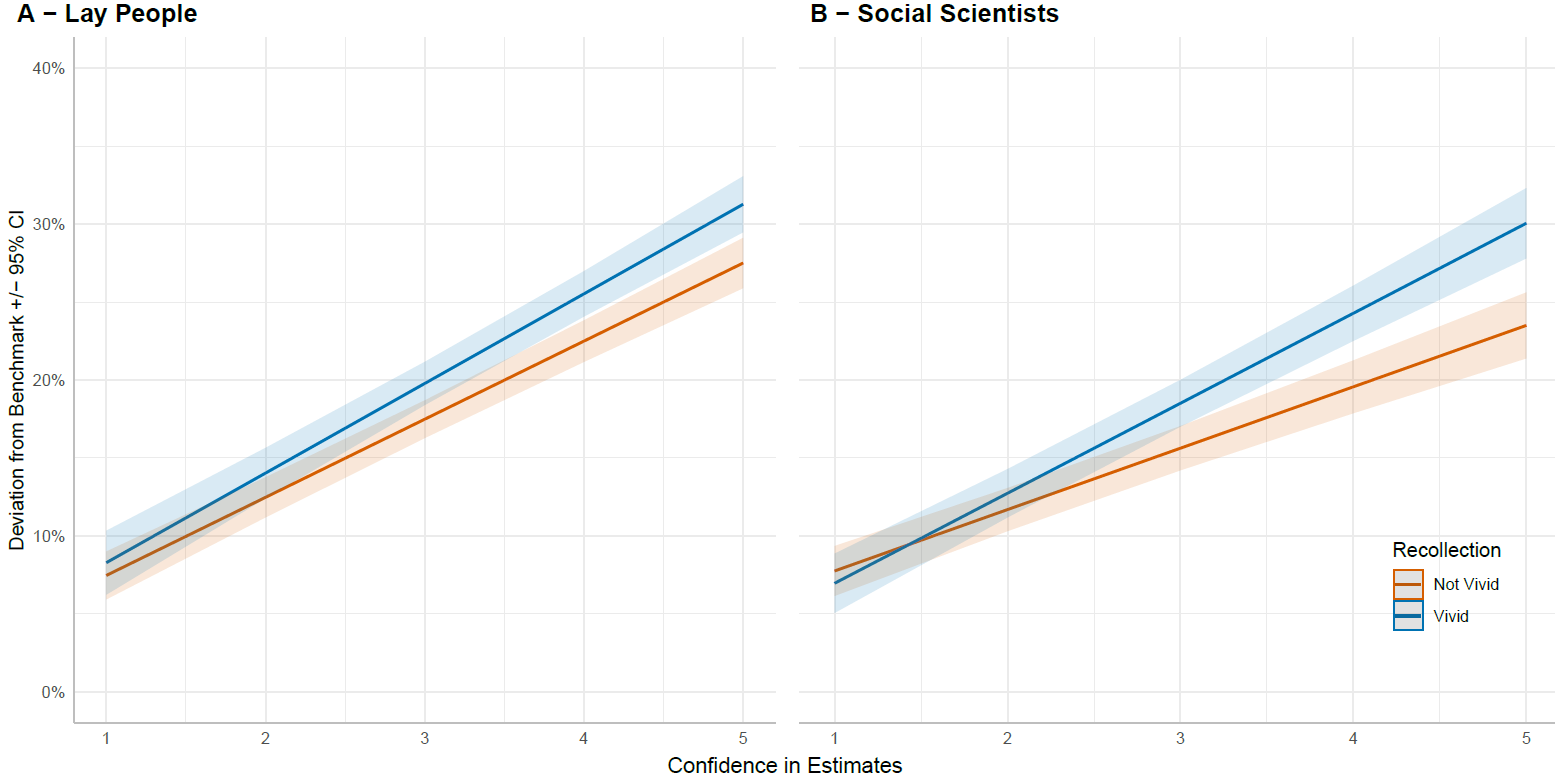
*Figure S16.* Percentage of participants whose direction of estimates (i.e., whether there would be an increase or decrease in that domain) matched objective data. Gray shaded area between 40% and 60% indicated 10% above and below chance levels, to help visualize if most of the sample was accurate or inaccurate in predicting the direction of change (not the amount of change).

## **Confidence and political orientation**

Cultural psychological research provides further reasons to anticipate differences among social scientists and lay people when predicting societal change. On average, social scientists are more liberal in their political orientation than the US population (*57*). In turn, liberal political orientation is associated with a more analytic and less holistic cognitive style (*58*). One aspect of analytic (vs. holistic) cognitive style concerns a tendency to reason about social change in a linear (vs. dialectical) fashion (*59*). To the extent that social scientists are more liberal in their political orientation, it is possible they would predict different trajectories of social change compared to lay people.

To determine whether demographic differences might explain differences in confidence between social scientists and lay people, we examined several potential moderators of confidence. While few covariates correlated with confidence, we did observe that in both samples, more conservative political orientation was associated with greater confidence in forecasts, scientists: *r* = .167, *p* < .001, lay people: *r* = .092, *p* = .014, and retrospective estimates, scientists: *r* = .077, *p* = .060, lay people: *r* = .125, *p* < .001. Moreover, in each case the indirect effect of group (social/behavioral scientist vs. lay people) on confidence via political orientation differences was significant, prospective: *B* = .09, 95% CI [.05, .13]*, p* <.001; retrospective: *B* = .09, 95% CI [.05, .13]*, p* <.001¸ accounting for 16% prospective / 19% retrospective of the variance in group effect on confidence.

## **Confidence, accuracy, and personal experience**

To follow-up on the negative association between confidence and accuracy of estimates, we explored whether retrospective estimates for which participants reported recalling vivid personal experiences qualified the confidence-inaccuracy association. As Figure S17 indicates, we observed significant confidence × vividness interactions for social scientists: χ2 (*df* = 1) = 17.52, *p* = .001 and a trend in the same direction for lay people: χ2 (*df* = 1) = 3.63, *p* = .057, with a steeper negative association for domains in which participants recalled vivid personal experiences.

*Figure S17*. Effect of confidence recall of personal experience for given domain of forecast on forecasting accuracy. Y-axis represents inaccuracy, quantified in absolute deviations from benchmarks across ten domains for lay people and social scientists. Confidence bands indicate 95% confidence band.

**Confidence, and accuracy for prejudice estimates**

Confidence was associated with a significantly greater inaccuracy among lay people, Explicit Prejudice: *B* = 4.79, *SE* = 0.65, 95%*CI* [3.52, 6.07]. Implicit Prejudice: *B* = 5.07, *SE* = 0.50, 95%*CI* [4.08, 6.05], and social scientists, Explicit Prejudice: *B* = 5.69, *SE* = 0.66, 95%*CI* [4.40, 6.99]. Implicit Prejudice: *B* = 5.89, *SE* = 0.59, 95%*CI* [4.73, 7.05]. Furthermore, the association between confidence in prejudice estimates and accuracy was moderated by vividness of memories, lay peoples: χ2 (*df* = 1) = 7.46, *p* = .006, social scientists: χ2 (*df* = 1) = 2.70, *p* = .010. Vivid memories were associated with a greater negative association between confidence and accuracy compared to non-vivid.

**Role of extremity in estimates for understanding the confidence-inaccuracy association**

As Figure 4 in the main manuscript shows, greater confidence in one’s estimates was negatively associated with accuracy of one’s estimates. We examined one possible reason for this association, which deals with the extremity of forecasts: Forecasting confidence may be associated with more extreme estimates. Because for most domains of societal change, actual benchmarks showed relatively modest degree of change, most extreme forecasts would be more inaccurate forecasts. To address this possibility, we obtained absolute values of the estimates and examined multi-level correlations (with responses nested within participants, using *correlation* package in *R*) between extremity of estimates (as indicated by greater absolute values), accuracy estimates, and confidence. Extremity was associated with greater confidence, *r* = .39, and greater inaccuracy, *r* = .85.

Moreover, when including absolute values of the estimates as covariates in linear mixed models with confidence predicting accuracy, the overall effect of confidence on inaccuracy changed from significant positive association, *B =* 5.25, *SE* = 0.13, *t*(10,956.98) = 39.52, *p* < .001, to a weaker negative association, *B*= - 0.26, *SE* = 0.06, *t*(13,535) = 4.37, *p* < .001, suggesting a significant indirect effect of confidence on inaccuracy via extremity, *B*= 5.72, 95% *CI* [5.48, 5.95], *p* < .001

|  |  |  |  |
| --- | --- | --- | --- |
| **Additional Supplementary Tables** *Table S4.* Significance Statistics for Maximum Absolute Change from the baseline in April 2020 for 11 Domains across the three time points: 6, 12 and 24 months. | | | |
| Dimension | *t* | *p* | *p\** |
| Explicit Prejudice | 16.00 | <.001 | <.001 |
| Implicit Prejudice | 14.91 | <.001 | <.001 |
| Individualism | -4.76 | <.001 | <.001 |
| Traditionalism | 4.66 | <.001 | <.001 |
| Generalized Trust | -6.12 | <.001 | <.001 |
| Political Polarization | 12.28 | <.001 | <.001 |
| Life Satisfaction | -17.38 | <.001 | <.001 |
| Depression | 22.17 | <.001 | <.001 |
| Delay of Gratification | 1.89 | .059 | .060 |
| Birth Rate | 5.89 | <.001 | <.001 |
| Climate Change | 11.87 | <.001 | <.001 |

*Note*. *P*-values in the rightmost column were adjusted using the Benjamini-Hochberg false discovery rate correction.

*Table S5.* Means and Confidence Intervals for Social Scientists’ Predictions in April 2020 by Domain and Time (6, 12 and 24 months)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Time (months)** | **Domain** | ***M*** | ***SE*** | ***95% CI\**** |
| 6 | Explicit Prejudice | 13.69 | .77 | 1.51 |
|  | Implicit Prejudice | 12.90 | .77 | 1.51 |
|  | Individualism | -3.92 | .76 | 1.50 |
|  | Traditionalism | 4.30 | .77 | 1.51 |
|  | Generalized Trust | -5.78 | .76 | 1.49 |
|  | Political Polarization | 12.02 | .76 | 1.50 |
|  | Life Satisfaction | -15.79 | .76 | 1.49 |
|  | Depression | 17.46 | .76 | 1.49 |
|  | Delay of Gratification | 1.83 | .77 | 1.50 |
|  | Birth Rate | -.04 | .76 | 1.50 |
|  | Climate Change | -3.57 | .76 | 1.48 |
| 12 | Explicit Prejudice | 9.91 | .77 | 1.51 |
|  | Implicit Prejudice | 9.60 | .77 | 1.51 |
|  | Individualism | -1.06 | .76 | 1.50 |
|  | Traditionalism | 2.74 | .77 | 1.51 |
|  | Generalized Trust | -3.27 | .76 | 1.49 |
|  | Political Polarization | 10.25 | .76 | 1.50 |
|  | Life Satisfaction | -8.30 | .76 | 1.49 |
|  | Depression | 12.90 | .76 | 1.49 |
|  | Delay of Gratification | -.23 | .77 | 1.50 |
|  | Birth Rate | 4.73 | .76 | 1.50 |
|  | Climate Change | 2.78 | .76 | 1.48 |
| 24 | Explicit Prejudice | 5.58 | .77 | 1.51 |
|  | Implicit Prejudice | 5.88 | .77 | 1.51 |
|  | Individualism | 1.43 | .76 | 1.50 |
|  | Traditionalism | .62 | .77 | 1.51 |
|  | Generalized Trust | -1.13 | .76 | 1.49 |
|  | Political Polarization | 7.74 | .76 | 1.50 |
|  | Life Satisfaction | -3.20 | .76 | 1.49 |
|  | Depression | 7.61 | .76 | 1.49 |
|  | Delay of Gratification | -.31 | .77 | 1.50 |
|  | Birth Rate | -1.41 | .76 | 1.50 |
|  | Climate Change | 9.21 | .76 | 1.48 |

*Note.* CI indicates 95% confidence distance from the mean.

*Table S6.* Mean Difference and Significance Statistics for Comparisons between Psychology and Other Disciplines by Time in April 2020.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Time (months) | Domain | *M*difference | *SE* | *z* | *p* | *p\** |
| 6 | Explicit Prejudice | -2.32 | 2.28 | -1.02 | .309 | .949 |
|  | Implicit Prejudice | -.17 | 2.23 | -.08 | .938 | .949 |
|  | Individualism | -.30 | 2.24 | -.14 | .893 | .949 |
|  | Traditionalism | 6.57 | 2.24 | 2.93 | .003 | .112 |
|  | Generalized Trust | -2.55 | 2.22 | -1.15 | .250 | .949 |
|  | Political Polarization | 4.07 | 2.22 | 1.83 | .067 | .949 |
|  | Life Satisfaction | -2.46 | 2.20 | -1.12 | .265 | .949 |
|  | Depression | 1.44 | 2.22 | .65 | .517 | .949 |
|  | Delay of Gratification | -.62 | 2.17 | -.29 | .775 | .949 |
|  | Birth Rate | -1.40 | 2.20 | -.64 | .524 | .949 |
|  | Climate Change | -.42 | 2.24 | -.19 | .850 | .949 |
| 12 | Explicit Prejudice | -.97 | 2.28 | -.43 | .670 | .949 |
|  | Implicit Prejudice | .73 | 2.23 | .33 | .742 | .949 |
|  | Individualism | -.35 | 2.24 | -.16 | .876 | .949 |
|  | Traditionalism | 2.91 | 2.24 | 1.30 | .195 | .949 |
|  | Generalized Trust | -2.71 | 2.24 | -1.21 | .227 | .949 |
|  | Political Polarization | 2.74 | 2.22 | 1.23 | .218 | .949 |
|  | Life Satisfaction | -1.11 | 2.20 | -.50 | .615 | .949 |
|  | Depression | -.76 | 2.22 | -.34 | .731 | .949 |
|  | Delay of Gratification | 2.03 | 2.17 | .94 | .350 | .949 |
|  | Birth Rate | .26 | 2.20 | .12 | .907 | .949 |
|  | Climate Change | 1.87 | 2.24 | .83 | .404 | .949 |
| 24 | Explicit Prejudice | -1.89 | 2.28 | -.83 | .406 | .949 |
|  | Implicit Prejudice | 1.63 | 2.23 | .73 | .465 | .949 |
|  | Individualism | -2.38 | 2.24 | -1.06 | .289 | .949 |
|  | Traditionalism | 1.23 | 2.24 | .55 | .583 | .949 |
|  | Generalized Trust | -1.56 | 2.24 | -.69 | .487 | .949 |
|  | Political Polarization | .59 | 2.24 | .26 | .792 | .949 |
|  | Life Satisfaction | 3.15 | 2.20 | 1.43 | .153 | .949 |
|  | Depression | .14 | 2.24 | .06 | .949 | .949 |
|  | Delay of Gratification | .20 | 2.17 | .09 | .927 | .949 |
|  | Birth Rate | 1.08 | 2.20 | .49 | .623 | .949 |
|  | Climate Change | 2.09 | 2.24 | .94 | .350 | .949 |

*Note*. *P*-values in the rightmost column were adjusted using the Benjamini-Hochberg false discovery rate correction. *M*difference refers to a difference score between mean of experts from psychology and mean of experts from other disciplines.

*Table S7.* Mean Difference and Significance Statistics for Comparisons between US and Other Countries by Time in April 2020.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Time (months) | Domain | *M*difference | *SE* | *z* | *p* | *p\** |
| 6 | Explicit Prejudice | -.09 | 1.54 | -.06 | .955 | .985 |
|  | Implicit Prejudice | 1.17 | 1.55 | .76 | .450 | .899 |
|  | Individualism | .45 | 1.54 | .29 | .769 | .974 |
|  | Traditionalism | -1.03 | 1.54 | -.67 | .503 | .899 |
|  | Generalized Trust | .69 | 1.53 | .45 | .654 | .899 |
|  | Political Polarization | 4.51 | 1.53 | 2.95 | .003 | .105 |
|  | Life Satisfaction | .69 | 1.53 | .45 | .653 | .899 |
|  | Depression | 2.92 | 1.53 | 1.92 | .056 | .458 |
|  | Delay of Gratification | 1.39 | 1.53 | .90 | .366 | .899 |
|  | Birth Rate | 1.10 | 1.53 | .72 | .474 | .899 |
|  | Climate Change | 4.05 | 1.52 | 2.66 | .008 | .129 |
| 12 | Explicit Prejudice | -.35 | 1.54 | -.22 | .823 | .974 |
|  | Implicit Prejudice | 1.15 | 1.55 | .74 | .457 | .899 |
|  | Individualism | .20 | 1.54 | .13 | .898 | .985 |
|  | Traditionalism | -.86 | 1.54 | -.56 | .575 | .899 |
|  | Generalized Trust | 1.62 | 1.53 | 1.06 | .290 | .899 |
|  | Political Polarization | 1.69 | 1.53 | 1.11 | .269 | .899 |
|  | Life Satisfaction | 2.70 | 1.53 | 1.76 | .078 | .513 |
|  | Depression | .99 | 1.53 | .65 | .519 | .899 |
|  | Delay of Gratification | .80 | 1.53 | .52 | .604 | .899 |
|  | Birth Rate | -.34 | 1.53 | -.22 | .826 | .974 |
|  | Climate Change | 2.45 | 1.52 | 1.61 | .107 | .590 |
| 24 | Explicit Prejudice | -.48 | 1.54 | -.31 | .758 | .974 |
|  | Implicit Prejudice | .70 | 1.55 | .45 | .654 | .899 |
|  | Individualism | -.02 | 1.54 | -.02 | .987 | .987 |
|  | Traditionalism | -.85 | 1.54 | -.55 | .580 | .899 |
|  | Generalized Trust | 1.78 | 1.53 | 1.16 | .245 | .899 |
|  | Political Polarization | .91 | 1.53 | .59 | .552 | .899 |
|  | Life Satisfaction | 1.72 | 1.53 | 1.12 | .262 | .899 |
|  | Depression | -2.99 | 1.53 | -1.96 | .050 | .458 |
|  | Delay of Gratification | 1.38 | 1.53 | .90 | .368 | .899 |
|  | Birth Rate | .23 | 1.53 | .15 | .881 | .985 |
|  | Climate Change | .14 | 1.52 | .09 | .925 | .985 |

*Note*. *P*-values in the rightmost column were adjusted using the Benjamini-Hochberg false discovery rate correction. *M*difference refers to a difference score between mean of experts from the U.S. and mean of experts from other countries.

*Table S8.* Significance testing of the difference between forecasted estimates for April 2022 and the baseline in April 2020.

|  |  |  |  |
| --- | --- | --- | --- |
| Dimension | *t* | *p* | *p \** |
| Explicit Prejudice | 7.96 | <.001 | <.001 |
| Implicit Prejudice | 8.50 | <.001 | <.001 |
| Individualism | 2.33 | .020 | .028 |
| Traditionalism | .81 | .417 | .459 |
| Generalized Trust | -1.89 | .059 | .072 |
| Political Polarization | 8.98 | <.001 | <.001 |
| Life Satisfaction | -4.51 | <.001 | <.001 |
| Depression | 12.11 | <.001 | <.001 |
| Delay of Gratification | -.58 | .566 | .567 |
| Birth Rate | -2.37 | .018 | .028 |
| Climate Change | 11.87 | <.001 | <.001 |

*Note:* *p*-values in the rightmost column were adjusted using the Benjamini-Hochberg false discovery rate correction.

*Table S9.* Significance Statistics for Maximum Absolute Change from the baseline in May 2020 for 15 Domains across the three time points: 6, 12 and 24 months.

|  |  |  |  |
| --- | --- | --- | --- |
| Dimension | *t* | *p* | *p \** |
| Explicit Prejudice | 14.78 | <.001 | <.001 |
| Implicit Prejudice | 14.63 | <.001 | <.001 |
| Individualism | 4.85 | <.001 | <.001 |
| Traditionalism | 5.11 | <.001 | <.001 |
| Generalized Trust | -7.25 | <.001 | <.001 |
| Political Polarization | 16.86 | <.001 | <.001 |
| Life Satisfaction | -13.69 | <.001 | <.001 |
| Depression | 16.41 | <.001 | <.001 |
| Delay of Gratification | -1.67 | .097 | .100 |
| Birth Rate | 2.76 | .006 | .007 |
| Climate Change | 9.82 | <.001 | <.001 |
| Violence | 7.41 | <.001 | <.001 |
| Religiosity | 8.34 | <.001 | <.001 |
| Loneliness | 12.19 | <.001 | <.001 |
| Charity | 2.82 | .005 | .006 |

*Note:* *p*-values in the rightmost column were adjusted using the Benjamini-Hochberg false discovery rate correction.

*Table S10.* Significance Testing Whether Forecasted Change for Each Domain (Linear and Quadratic Temporal Trends) Vary by Group of Sampled Social Scientists (April vs. May 2020).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Domain | Effect | *t* | *p* | *p\** |
| Explicit Prejudice | Linear | 1.11 | .266 | .568 |
|  | Quadratic | 0.50 | .618 | .849 |
| Implicit Prejudice | Linear | 0.64 | .525 | .825 |
|  | Quadratic | 0.17 | .869 | .893 |
| Individualism | Linear | -4.25 | <.001 | .001 |
|  | Quadratic | 1.02 | .310 | .568 |
| Traditionalism | Linear | -0.19 | .854 | .893 |
|  | Quadratic | 0.14 | .893 | .893 |
| Generalized Trust | Linear | -0.44 | .661 | .856 |
|  | Quadratic | -0.15 | .879 | .893 |
| Political Polarization | Linear | -2.75 | .006 | .045 |
|  | Quadratic | 1.68 | .093 | .342 |
| Life Satisfaction | Linear | -1.13 | .258 | .568 |
|  | Quadratic | 2.11 | .035 | .157 |
| Depression | Linear | 1.46 | .144 | .412 |
|  | Quadratic | -1.02 | .309 | .568 |
| Delay of Gratification | Linear | 0.76 | .448 | .758 |
|  | Quadratic | -0.50 | .616 | .849 |
| Birth Rate | Linear | -2.10 | .036 | .157 |
|  | Quadratic | 3.36 | .001 | .009 |
| Climate Change | Linear | -1.44 | .150 | .412 |
|  | Quadratic | 0.17 | .863 | .893 |

*Note:* Rightmost *p* value column was adjusted for number of tests using Benjamini-Hochberg false discovery rate correction. For 3 temporal estimates, estimation of change can be parsimoniously decomposed into overall degree of change (i.e., linear effect) and possible curve in the estimated trajectory (i.e., quadratic effect). Therefore, in our analyses we focused on estimation of sample-wise differences in linear and quadratic effects.

*Table S11.* Descriptives of Social Scientists’ May 2020 Forecasts for Societal Change over Time and Sub-group Comparisons.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | US vs. Non-US Scientists | | | | Psychology vs. non-Psychology | | | |
| Time | Domain | *M* | *SE* | *95% CI\** | *Mdifference* | *z* | *p* | *p\** | *Mdifference* | *z* | *p* | *p\** |
| 6 months | Explicit Prejudice | 13.04 | .91 | 1.78 | 1.41 | 0.74 | .461 | .777 | 0.45 | 0.20 | .842 | .975 |
| Implicit Prejudice | 13.83 | .91 | 1.79 | -0.17 | -0.09 | .931 | .974 | -1.89 | -0.84 | .404 | .699 |
| Individualism | 2.83 | .92 | 1.80 | 0.68 | 0.36 | .722 | .865 | 0.96 | 0.42 | .673 | .842 |
| Traditionalism | 4.64 | .92 | 1.80 | -2.18 | -1.14 | .257 | .777 | -6.10 | -2.67 | .008 | .131 |
| Generalized Trust | -7.19 | .92 | 1.80 | 0.11 | 0.06 | .956 | .977 | -4.91 | -2.16 | .031 | .230 |
| Political Polarization | 18.86 | .91 | 1.79 | 2.02 | 1.06 | .290 | .777 | 1.25 | 0.56 | .576 | .810 |
| Life Satisfaction | -15.20 | .91 | 1.79 | 2.77 | 1.45 | .146 | .659 | 3.38 | 1.48 | .138 | .444 |
| Depression | 16.72 | .92 | 1.80 | 2.31 | 1.21 | .227 | .777 | 0.77 | 0.34 | .733 | .892 |
| Delay of Gratification | .01 | .92 | 1.80 | 1.54 | 0.80 | .423 | .777 | -5.49 | -2.44 | .015 | .131 |
| Birth Rate | 1.04 | .92 | 1.80 | -0.25 | -0.13 | .895 | .974 | 0.09 | 0.04 | .969 | .991 |
| Violence | .96 | .97 | 1.89 | -3.39 | -1.67 | .094 | .659 | -5.82 | -2.45 | .014 | .131 |
| Religiosity | 7.13 | .96 | 1.88 | -0.65 | -0.32 | .749 | .865 | -1.19 | -0.50 | .617 | .817 |
| Loneliness | 15.72 | .95 | 1.87 | -1.46 | -0.73 | .466 | .777 | -4.84 | -2.04 | .041 | .265 |
| Charity | 3.21 | .95 | 1.87 | -0.94 | -0.47 | .638 | .836 | 0.33 | 0.14 | .889 | .975 |
| Climate Change | -2.20 | .92 | 1.80 | 0.90 | 0.47 | .638 | .836 | -1.81 | -0.80 | .426 | .711 |
| 12 months | Explicit Prejudice | 9.19 | .91 | 1.78 | -0.75 | -0.39 | .693 | .865 | -0.09 | -0.04 | .967 | .991 |
| Implicit Prejudice | 10.62 | .91 | 1.79 | -0.03 | -0.01 | .989 | .989 | -1.90 | -0.84 | .403 | .699 |
| Individualism | 3.36 | .92 | 1.80 | -0.61 | -0.32 | .750 | .865 | 0.36 | 0.16 | .874 | .975 |
| Traditionalism | 2.93 | .92 | 1.80 | -3.52 | -1.83 | .067 | .659 | -6.14 | -2.69 | .007 | .131 |
| Generalized Trust | -4.68 | .92 | 1.80 | 0.96 | 0.50 | .617 | .836 | -3.26 | -1.44 | .151 | .453 |
| Political Polarization | 14.58 | .91 | 1.79 | 1.65 | 0.87 | .387 | .777 | 1.89 | 0.84 | .399 | .699 |
| Life Satisfaction | -10.03 | .91 | 1.79 | 4.14 | 2.17 | .030 | .659 | 4.07 | 1.78 | .075 | .336 |
| Depression | 13.43 | .92 | 1.80 | 1.67 | 0.87 | .383 | .777 | -0.24 | -0.11 | .914 | .980 |
| Delay of Gratification | -1.30 | .92 | 1.80 | 0.87 | 0.45 | .650 | .836 | -4.07 | -1.81 | .070 | .336 |
| Birth Rate | 2.26 | .92 | 1.80 | -3.41 | -1.78 | .075 | .659 | 2.62 | 1.15 | .252 | .566 |
| Violence | 5.51 | .97 | 1.89 | -4.02 | -1.99 | .047 | .659 | -3.56 | -1.50 | .134 | .444 |
| Religiosity | 4.02 | .96 | 1.88 | -1.24 | -0.61 | .539 | .809 | -2.08 | -0.88 | .381 | .699 |
| Loneliness | 7.99 | .95 | 1.87 | -2.14 | -1.07 | .284 | .777 | -4.32 | -1.82 | .069 | .336 |
| Charity | .38 | .95 | 1.87 | -2.53 | -1.27 | .206 | .777 | 0.37 | 0.16 | .874 | .975 |
| Climate Change | 3.39 | .91 | 1.79 | 1.24 | 0.65 | .518 | .809 | -2.72 | -1.20 | .231 | .566 |
| 24 months | Explicit Prejudice | 6.03 | .91 | 1.78 | -1.76 | -0.92 | .357 | .777 | -1.01 | -0.44 | .657 | .842 |
| Implicit Prejudice | 7.40 | .91 | 1.79 | -0.24 | -0.13 | .899 | .974 | -1.74 | -0.77 | .443 | .712 |
| Individualism | 3.78 | .92 | 1.80 | -0.20 | -0.10 | .918 | .974 | 1.64 | 0.72 | .469 | .728 |
| Traditionalism | .83 | .92 | 1.81 | -1.89 | -0.98 | .326 | .777 | -5.63 | -2.46 | .014 | .131 |
| Generalized Trust | -3.05 | .92 | 1.80 | 3.08 | 1.60 | .109 | .659 | -2.46 | -1.09 | .278 | .595 |
| Political Polarization | 11.94 | .91 | 1.79 | -1.80 | -0.94 | .347 | .777 | 2.82 | 1.26 | .209 | .554 |
| Life Satisfaction | -3.07 | .91 | 1.79 | 4.98 | 2.61 | .009 | .406 | 3.41 | 1.50 | .134 | .444 |
| Depression | 8.09 | .92 | 1.80 | 2.00 | 1.05 | .295 | .777 | -1.52 | -0.67 | .503 | .755 |
| Delay of Gratification | -1.38 | .92 | 1.80 | 0.91 | 0.48 | .634 | .836 | -1.20 | -0.53 | .594 | .810 |
| Birth Rate | -1.38 | .92 | 1.80 | -1.43 | -0.75 | .454 | .777 | 1.24 | 0.54 | .588 | .810 |
| Violence | 5.44 | .97 | 1.89 | -2.95 | -1.46 | .146 | .659 | -2.74 | -1.15 | .249 | .566 |
| Religiosity | 1.61 | .96 | 1.88 | -1.68 | -0.83 | .405 | .777 | -2.98 | -1.26 | .209 | .554 |
| Loneliness | 2.06 | .95 | 1.87 | -2.98 | -1.49 | .135 | .659 | -4.06 | -1.71 | .087 | .355 |
| Charity | -.29 | .95 | 1.87 | -2.39 | -1.20 | .231 | .777 | -0.02 | -0.01 | .993 | .993 |
| Climate Change | 8.91 | .91 | 1.79 | 1.20 | 0.63 | .532 | .809 | -1.97 | -0.87 | .384 | .699 |

*Note.* *CI* indicates 95% confidence distance from the mean. Right column indicates *p*-values adjusted for number of tests using Benjamini-Hochberg false discovery rate procedure. *M*difference refers to a difference score between mean of experts in Psychology and other disciplines.

*Table S12.* Do Social Scientists Expect Significant Difference between May 2022 Forecasts and May 2020 Baseline?

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | Dimensions | *t* | *p* | *p\** | | Explicit Prejudice | 7.14 | <.001 | <.001 | | Implicit Prejudice | 8.94 | <.001 | <.001 | | Individualism | 4.85 | <.001 | <.001 | | Traditionalism | .96 | .338 | .361 | | Generalized Trust | -3.69 | <.001 | <.001 | | Political Polarization | 11.70 | <.001 | <.001 | | Life Satisfaction | -2.99 | .003 | .005 | | Depression | 9.48 | <.001 | <.001 | | Delay of Gratification | -1.67 | .097 | .112 | | Birth Rate | -1.82 | .069 | .087 | | Climate Change | 9.82 | <.001 | <.001 | | Violence | 7.94 | <.001 | <.001 | | Religiosity | 2.27 | .024 | .033 | | Loneliness | 2.30 | .022 | .033 | | Charity | -.23 | .816 | .816 |   *Note:* Right column indicates *p*-values adjusted for number of tests using Benjamini-Hochberg false discovery rate procedure. |

*Table S13*. Significance Testing Whether Forecasted Change for Each Domain (Linear and Quadratic Temporal Trends) Vary between Lay People vs. Social Scientists.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dimension | Effect | *t* | *p* | *p\** |
| Explicit Prejudice | Linear | -3.44 | .001 | .018 |
|  | Quadratic | 1.40 | .163 | .543 |
| Implicit Prejudice | Linear | -2.40 | .017 | .166 |
|  | Quadratic | .91 | .363 | .796 |
| Individualism | Linear | -.43 | .667 | .854 |
|  | Quadratic | .15 | .883 | .883 |
| Traditionalism | Linear | .43 | .668 | .854 |
|  | Quadratic | -.26 | .799 | .854 |
| Generalized Trust | Linear | -2.29 | .022 | .168 |
|  | Quadratic | .22 | .826 | .854 |
| Political Polarization | Linear | -1.89 | .059 | .300 |
|  | Quadratic | 1.16 | .248 | .707 |
| Life Satisfaction | Linear | -1.80 | .072 | .300 |
|  | Quadratic | .25 | .799 | .854 |
| Depression | Linear | -1.13 | .259 | .707 |
|  | Quadratic | .37 | .712 | .854 |
| Delay of Gratification | Linear | -.70 | .486 | .854 |
|  | Quadratic | .44 | .663 | .854 |
| Birth Rate | Linear | -.86 | .389 | .796 |
|  | Quadratic | 3.18 | .002 | .023 |
| Climate Change | Linear | -1.75 | .080 | .300 |
|  | Quadratic | .26 | .795 | .854 |
| Religiosity | Linear | .39 | .695 | .854 |
|  | Quadratic | -.55 | .579 | .854 |
| Charity | Linear | -1.77 | .076 | .300 |
|  | Quadratic | .80 | .425 | .796 |
| Violence | Linear | .55 | .579 | .854 |
|  | Quadratic | -.87 | .383 | .796 |
| Loneliness | Linear | -.84 | .400 | .796 |
|  | Quadratic | .26 | .798 | .854 |

*Note:* Rightmost *p* values were adjusted for number of tests using Benjamini-Hochberg false discovery rate correction. For 3 temporal estimates, estimation of change can be parsimoniously decomposed into overall degree of change (i.e., linear effect) and possible curve in the estimated trajectory (i.e., quadratic effect). Therefore, in our analyses we focused on estimation of sample-wise differences in linear and quadratic effects.

*Table S14.* Comparisons of Retrospective Estimates between Social Scientists and Lay People in October/November 2020 using Frequentist Methods.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dimension | *Mdifference* | *SE* | *z* | *p* | *p\** |
| Explicit Prejudice | -1.66 | 1.40 | -1.18 | .238 | .297 |
| Implicit Prejudice | -2.59 | 1.41 | -1.84 | .066 | .123 |
| Individualism | -4.02 | 1.40 | -2.87 | .004 | .015 |
| Traditionalism | -.25 | 1.40 | -.18 | .859 | .859 |
| Generalized Trust | -4.41 | 1.40 | -3.15 | .002 | .008 |
| Political Polarization | -2.05 | 1.40 | -1.46 | .144 | .216 |
| Life Satisfaction | -1.77 | 1.40 | -1.26 | .207 | .283 |
| Depression | -3.81 | 1.40 | -2.72 | .006 | .019 |
| Delay of Gratification | -5.56 | 1.40 | -3.97 | < .001 | .001 |
| Birth Rate | -2.28 | 1.40 | -1.63 | .104 | .173 |
| Violence | -8.01 | 1.41 | -5.66 | < .001 | < .001 |
| Religiosity | -1.50 | 1.41 | -1.06 | .289 | .332 |
| Loneliness | -3.17 | 1.41 | -2.24 | .025 | .062 |
| Charity | -1.44 | 1.41 | -1.02 | .310 | .332 |
| Climate Change | -2.73 | 1.40 | -1.95 | .051 | .110 |

*Note:* Rightmost *p* value column was adjusted for false discovery rate using Benjamini-Hochberg correction.

*Table S15.* Comparisons of Retrospective Estimates between Social Scientists and Lay People in October/November 2020 using Bayesian Statistics.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Dimension | *β* | *Estimated Error* | *95%CI (Lower)* | *95%CI (Upper)* | *CI Includes Zero* | *Bayes Factor* | *Strength of Evidence* |
| Traditionalism | .01 | .08 | -.14 | .17 | Yes | 62.61 | Very Strong (H0) |
| Charity | .07 | .08 | -.08 | .23 | Yes | 43.33 | Very Strong (H0) |
| Explicit Prejudice | .08 | .08 | -.07 | .24 | Yes | 36.21 | Very Strong (H0) |
| Life Satisfaction | .10 | .08 | -.06 | .25 | Yes | 31.00 | Very Strong (H0) |
| Religiosity | .10 | .08 | -.05 | .26 | Yes | 26.61 | Strong (H0) |
| Political Polarization | .11 | .08 | -.04 | .26 | Yes | 25.72 | Strong (H0) |
| Implicit Prejudice | .14 | .08 | -.02 | .29 | Yes | 13.48 | Strong (H0) |
| Climate Change | .15 | .08 | .00 | .31 | No | 9.48 | Moderate (H0) |
| Birth Rate | .18 | .08 | .03 | .34 | No | 4.29 | Moderate (H0) |
| Loneliness | .18 | .08 | .03 | .33 | No | 4.21 | Moderate (H0) |
| Individualism | .23 | .08 | .07 | .38 | No | 1.18 | Anecdotal (H0) |
| Generalized Trust | .24 | .08 | .08 | .39 | No | .80 | Anecdotal (H1) |
| Depression | .24 | .08 | .08 | .39 | No | .47 | Anecdotal (H1) |
| Delay of Gratification | .29 | .08 | .12 | .45 | No | .26 | Moderate (H1) |
| Violence | .43 | .08 | .27 | .59 | No | < 0.01 | Extreme (H1) |

*Note:* 95% CI refers to Bayesian credible interval and Bayes Factor denotes BF01. Bayes Factor interpretation is based on Lee & Wagenmakers, 2013.

*Table S16.* Comparisons of Social Scientists’ Retrospective Estimates in October/November 2020 by Self-Reported Expertise in Domain of Forecast.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dimension | *Mdifference* | *SE* | *z* | *p* | *p\** |
| Explicit Prejudice | 4.71 | 1.88 | 2.50 | .012 | .186 |
| Implicit Prejudice | 3.48 | 1.82 | 1.92 | .055 | .225 |
| Individualism | 3.86 | 2.05 | 1.88 | .060 | .225 |
| Traditionalism | .93 | 2.49 | .37 | .709 | .792 |
| Generalized Trust | 1.56 | 2.24 | .70 | .485 | .727 |
| Political Polarization | -1.79 | 2.09 | -.86 | .393 | .727 |
| Life Satisfaction | 1.70 | 1.84 | .92 | .356 | .727 |
| Depression | .81 | 1.88 | .43 | .666 | .792 |
| Delay of Gratification | -.86 | 1.94 | -.44 | .658 | .792 |
| Birth Rate | -1.51 | 4.52 | -.33 | .739 | .792 |
| Violence | -6.05 | 3.42 | -1.77 | .077 | .230 |
| Religiosity | -.21 | 2.64 | -.08 | .937 | .937 |
| Loneliness | -1.54 | 2.01 | -.77 | .442 | .727 |
| Charity | 2.82 | 3.09 | .91 | .361 | .727 |
| Climate Change | 5.65 | 2.98 | 1.90 | .058 | .225 |

*Note:* Rightmost *p* value column was adjusted for false discovery rate using Benjamini-Hochberg correction.

*Table S17.* Comparisons of the Effects of News Reports and Vivid Memories on Retrospective Estimates by Sample and Dimension.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sample | Dimension | Type | *Mdifference* | *SE* | *z* | *p* | *p\** | *Extremeness* |
| Lay People | Birth Rate | News | 3.27 | 2.24 | 1.46 | .144 | .188 | Unchanged |
|  | Vivid | .11 | 3.66 | .03 | .975 | .975 | Unchanged |
| Charity | News | 8.67 | 2.05 | 4.22 | <.001 | <.001 | Greater |
|  | Vivid | 7.15 | 2.18 | 3.28 | .001 | .003 | Greater |
| Climate Change | News | 14.09 | 1.68 | 8.39 | <.001 | <.001 | Greater |
|  | Vivid | 9.85 | 2.00 | 4.93 | <.001 | <.001 | Greater |
| Delay of Gratification | News | 1.95 | 2.13 | .92 | .360 | .400 | Unchanged |
|  | Vivid | 1.65 | 1.85 | .89 | .372 | .436 | Unchanged |
| Depression | News | 5.62 | 1.74 | 3.24 | .001 | .003 | Greater |
|  | Vivid | 5.82 | 1.70 | 3.42 | .001 | .002 | Greater |
| Explicit Prejudice | News | 8.28 | 1.70 | 4.87 | <.001 | <.001 | Greater |
|  | Vivid | 8.06 | 1.95 | 4.13 | <.001 | <.001 | Greater |
| Generalized Trust | News | -2.61 | 1.77 | -1.47 | .142 | .188 | Unchanged |
|  | Vivid | -4.54 | 1.70 | -2.67 | .008 | .015 | Unchanged |
| Implicit Prejudice | News | 9.66 | 1.71 | 5.63 | <.001 | <.001 | Greater |
|  | Vivid | 5.57 | 1.85 | 3.01 | .003 | .006 | Unchanged |
| Individualism | News | 6.97 | 1.79 | 3.89 | <.001 | <.001 | Greater |
|  | Vivid | 1.72 | 1.96 | .88 | .378 | .436 | Unchanged |
| Life Satisfaction | News | -.17 | 1.77 | -.10 | .924 | .924 | Unchanged |
|  | Vivid | -2.68 | 1.70 | -1.58 | .114 | .163 | Unchanged |
| Loneliness | News | 7.83 | 1.69 | 4.62 | <.001 | <.001 | Greater |
|  | Vivid | 7.21 | 1.70 | 4.24 | <.001 | <.001 | Greater |
| Political Polarization | News | 13.00 | 1.97 | 6.59 | <.001 | <.001 | Greater |
|  | Vivid | 4.43 | 1.70 | 2.61 | .009 | .016 | Unchanged |
| Religiosity | News | 7.59 | 2.31 | 3.28 | .001 | .003 | Greater |
|  | Vivid | 7.60 | 2.14 | 3.55 | <.001 | .002 | Greater |
| Traditionalism | News | 3.78 | 1.94 | 1.95 | .051 | .081 | Unchanged |
|  | Vivid | 5.42 | 2.05 | 2.64 | .008 | .015 | Unchanged |
| Violence | News | 12.95 | 1.73 | 7.49 | <.001 | <.001 | Greater |
|  | Vivid | 11.55 | 2.22 | 5.21 | <.001 | <.001 | Greater |
| Social Scientists | Birth Rate | News | -.93 | 2.66 | -.35 | .728 | .753 | Unchanged |
|  | Vivid | -4.14 | 3.36 | -1.23 | .218 | .275 | Unchanged |
| Charity | News | 3.89 | 2.34 | 1.66 | .096 | .138 | Unchanged |
|  | Vivid | 7.87 | 2.25 | 3.50 | <.001 | .002 | Greater |
| Climate Change | News | 7.87 | 2.07 | 3.80 | <.001 | <.001 | Greater |
|  | Vivid | 9.09 | 2.42 | 3.76 | <.001 | .001 | Greater |
| Delay of Gratification | News | 3.13 | 2.57 | 1.22 | .223 | .279 | Unchanged |
|  | Vivid | -1.44 | 2.37 | -.61 | .544 | .583 | Unchanged |
| Depression | News | 4.04 | 2.18 | 1.85 | .064 | .096 | Unchanged |
|  | Vivid | 7.32 | 2.16 | 3.39 | .001 | .002 | Greater |
| Explicit Prejudice | News | 2.32 | 2.39 | .97 | .331 | .382 | Unchanged |
|  | Vivid | 5.28 | 2.24 | 2.36 | .018 | .030 | Unchanged |
| Generalized Trust | News | -4.95 | 2.11 | -2.35 | .019 | .038 | Unchanged |
|  | Vivid | -9.13 | 2.14 | -4.27 | <.001 | <.001 | Greater |
| Implicit Prejudice | News | 5.99 | 2.09 | 2.87 | .004 | .009 | Unchanged |
|  | Vivid | 3.52 | 2.43 | 1.45 | .147 | .200 | Unchanged |
| Individualism | News | 4.25 | 2.12 | 2.00 | .045 | .080 | Unchanged |
|  | Vivid | 1.67 | 2.47 | .67 | .500 | .556 | Unchanged |
| Life Satisfaction | News | -4.26 | 2.15 | -1.98 | .048 | .080 | Unchanged |
|  | Vivid | -4.10 | 2.09 | -1.96 | .050 | .075 | Unchanged |
| Loneliness | News | 5.75 | 2.17 | 2.65 | .008 | .017 | Unchanged |
|  | Vivid | 6.45 | 2.12 | 3.05 | .002 | .005 | Unchanged |
| Political Polarization | News | 2.83 | 2.66 | 1.06 | .288 | .345 | Unchanged |
|  | Vivid | 6.56 | 2.09 | 3.14 | .002 | .004 | Greater |
| Religiosity | News | 1.62 | 3.12 | .52 | .604 | .647 | Unchanged |
|  | Vivid | .86 | 3.80 | .23 | .821 | .849 | Unchanged |
| Traditionalism | News | 5.21 | 2.24 | 2.33 | .020 | .038 | Unchanged |
|  | Vivid | 3.52 | 2.87 | 1.23 | .220 | .275 | Unchanged |
| Violence | News | 6.56 | 2.14 | 3.06 | .002 | .005 | Greater |
|  | Vivid | 6.11 | 2.93 | 2.08 | .037 | .059 | Unchanged |

*Note:* Rightmost *p* value column was adjusted for false discovery rate using Benjamini-Hochberg correction. *Mdifference* was computed by subtracting the mean of non-vivid from vivid and non-concrete news from concrete-news. Extremeness signifies whether vivid and concrete-news estimates were lower (absolute value of estimates closer to baseline), greater (absolute value of estimates farther away from baseline) or unchanged as compared to non-vivid and non-concrete-news estimates.

*Table S18*. Comparisons of Prospective and Retrospective Estimates by Sample and Dimension.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sample | Dimension | *Mdifference* | *SE* | *z* | *p* | *p\** | *Extremeness* |
| Lay People | Birth Rate | 3.73 | 1.39 | 2.69 | .007 | .011 | Unchanged |
|  | Charity | 4.18 | 1.39 | 3.00 | .003 | .005 | Lower |
|  | Climate Change | -1.82 | 1.39 | -1.32 | .188 | .211 | Unchanged |
|  | Delay of Gratification | -.46 | 1.39 | -.33 | .739 | .739 | Unchanged |
|  | Depression | -7.33 | 1.39 | -5.29 | < .001 | < .001 | Greater |
|  | Explicit Prejudice | -1.79 | 1.39 | -1.21 | .197 | .211 | Unchanged |
|  | Generalized Trust | 7.76 | 1.39 | 5.60 | < .001 | < .001 | Greater |
|  | Implicit Prejudice | -4.54 | 1.39 | -3.28 | .001 | .002 | Greater |
|  | Individualism | -4.54 | 1.39 | -3.27 | .001 | .002 | Greater |
|  | Life Satisfaction | 9.32 | 1.39 | 6.72 | < .001 | < .001 | Greater |
|  | Loneliness | -11.85 | 1.39 | -8.51 | < .001 | < .001 | Greater |
|  | Political Polarization | -11.36 | 1.39 | -8.18 | < .001 | < .001 | Greater |
|  | Religiosity | 2.28 | 1.39 | 1.63 | .102 | .128 | Unchanged |
|  | Traditionalism | -2.36 | 1.39 | -1.70 | .088 | .121 | Unchanged |
|  | Violence | -8.34 | 1.39 | -5.99 | < .001 | < .001 | Greater |
| Social Scientists | Birth Rate | .97 | 1.19 | .82 | .415 | .445 | Unchanged |
|  | Charity | 2.90 | 1.41 | 2.05 | .040 | .050 | Unchanged |
|  | Climate Change | -3.59 | 1.18 | -3.03 | .002 | .005 | Lower |
|  | Delay of Gratification | 4.34 | 1.19 | 3.65 | < .001 | .001 | Greater |
|  | Depression | -3.16 | 1.19 | -2.66 | .008 | .011 | Unchanged |
|  | Explicit Prejudice | 5.02 | 1.19 | 4.22 | < .001 | < .001 | Lower |
|  | Generalized Trust | 10.51 | 1.19 | 8.86 | < .001 | < .001 | Greater |
|  | Implicit Prejudice | 3.54 | 1.19 | 2.96 | .003 | .005 | Lower |
|  | Individualism | -7.92 | 1.19 | -6.67 | < .001 | < .001 | Greater |
|  | Life Satisfaction | 5.08 | 1.19 | 4.28 | < .001 | < .001 | Greater |
|  | Loneliness | -9.52 | 1.41 | -6.73 | < .001 | < .001 | Greater |
|  | Political Polarization | -10.88 | 1.19 | -9.16 | < .001 | < .001 | Greater |
|  | Religiosity | 4.06 | 1.42 | 2.86 | .004 | .006 | Unchanged |
|  | Traditionalism | .22 | 1.19 | .18 | .857 | .857 | Unchanged |
|  | Violence | -2.34 | 1.43 | -1.64 | .101 | .116 | Unchanged |

*Note:* Rightmost *p* value column was adjusted for false discovery rate using Benjamini-Hochberg correction. *Mdifference* was computed by subtracting the mean of retrospective from prospective estimates. Extremeness signifies whether retrospective estimates were: lower (absolute value of retrospective estimates closer to baseline than prospective), greater (absolute value of retrospective estimates farther away from baseline than prospective) or unchanged (no difference between the two types of estimates) as compared to prospective estimates.

*Table S19.* Comparisons of Prospective and Retrospective Estimates by Sample and Dimension Controlling for Demographics.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sample | Dimension | *Mdifference* | *SE* | *z* | *p* | *p\** | *Extremeness* |
| Lay People | Birth Rate | 3.68 | 1.39 | 2.65 | .008 | .012 | Unchanged |
|  | Charity | 4.18 | 1.39 | 3.00 | .003 | .004 | Lower |
|  | Climate Change | -1.91 | 1.39 | -1.38 | .169 | .195 | Unchanged |
|  | Delay of Gratification | -.49 | 1.39 | -.36 | .723 | .723 | Unchanged |
|  | Depression | -7.33 | 1.39 | -5.29 | < .001 | < .001 | Greater |
|  | Explicit Prejudice | -1.75 | 1.39 | -1.26 | .207 | .222 | Unchanged |
|  | Generalized Trust | 7.71 | 1.39 | 5.56 | < .001 | < .001 | Greater |
|  | Implicit Prejudice | -4.56 | 1.39 | -3.29 | .001 | .002 | Greater |
|  | Individualism | -4.59 | 1.39 | -3.31 | .001 | .002 | Greater |
|  | Life Satisfaction | 9.28 | 1.39 | 6.68 | < .001 | < .001 | Greater |
|  | Loneliness | -11.87 | 1.39 | -8.52 | < .001 | < .001 | Greater |
|  | Political Polarization | -11.35 | 1.39 | -8.17 | < .001 | < .001 | Greater |
|  | Religiosity | 2.30 | 1.40 | 1.65 | .099 | .124 | Unchanged |
|  | Traditionalism | -2.35 | 1.39 | -1.70 | .090 | .123 | Unchanged |
|  | Violence | -8.40 | 1.39 | -6.02 | < .001 | < .001 | Greater |
| Social Scientists | Birth Rate | 2.23 | 1.40 | 1.60 | .111 | .128 | Unchanged |
|  | Charity | 3.34 | 1.43 | 2.34 | .019 | .027 | Unchanged |
|  | Climate Change | -2.77 | 1.39 | -1.99 | .047 | .058 | Unchanged |
|  | Delay of Gratification | 3.65 | 1.40 | 2.62 | .009 | .017 | Unchanged |
|  | Depression | -3.41 | 1.39 | -2.45 | .014 | .021 | Unchanged |
|  | Explicit Prejudice | 4.75 | 1.39 | 3.41 | .001 | .002 | Lower |
|  | Generalized Trust | 10.29 | 1.39 | 7.39 | < .001 | < .001 | Greater |
|  | Implicit Prejudice | 4.10 | 1.40 | 2.94 | .003 | .007 | Lower |
|  | Individualism | -3.45 | 1.39 | -2.48 | .013 | .021 | Unchanged |
|  | Life Satisfaction | 6.19 | 1.39 | 4.45 | < .001 | < .001 | Greater |
|  | Loneliness | -9.40 | 1.43 | -6.58 | < .001 | < .001 | Greater |
|  | Political Polarization | -6.81 | 1.39 | -4.90 | < .001 | < .001 | Greater |
|  | Religiosity | 4.53 | 1.43 | 3.16 | .002 | .004 | Lower |
|  | Traditionalism | .73 | 1.40 | .52 | .601 | .601 | Unchanged |
|  | Violence | -1.57 | 1.44 | -1.09 | .276 | .296 | Unchanged |

*Note:* Rightmost *p* value column was adjusted for false discovery rate using Benjamini-Hochberg correction. *Mdifference* was computed by subtracting the mean of retrospective from prospective estimates. Extremeness signifies whether retrospective estimates were: lower (absolute value of retrospective estimates closer to baseline than prospective), greater (absolute value of retrospective estimates farther away from baseline than prospective) or unchanged (no difference between the two types of estimates) as compared to prospective estimates.

*Table S20.* Comparisons of Prospective and Retrospective Estimates against Actual Change between April 2020 and October 2020 by Sample and Dimension.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sample | Dimension | Estimate Type | Estimate | Actual Change | *Mdifference* | *t* | *p* | *p\** |
| Lay People | Depression | Prospective | 16.76 | 2.50 | 14.26 | 13.88 | <.001 | < .001 |
| Retrospective | 24.09 | 2.50 | 21.59 | 24.98 | <.001 | < .001 |
| Life Satisfaction | Prospective | -9.51 | 1.00 | 10.51 | -9.39 | <.001 | < .001 |
| Retrospective | -18.83 | 1.00 | 19.83 | -19.45 | <.001 | < .001 |
| Implicit Prejudice  Explicit Prejudice | Prospective | 7.85 | -6.20 | 14.35 | 16.24 | <.001 | < .001 |
| Retrospective | 12.37 | -6.20 | 18.57 | 19.26 | <.001 | < .001 |
| Prospective | 8.25 | -51.00 | 59.25 | 66.07 | <.001 | <.001 |
| Retrospective | 10.04 | -51.00 | 61.04 | 59.73 | <.001 | <.001 |
| Loneliness | Prospective | 16.21 | -4.10 | 20.31 | 16.95 | <.001 | < .001 |
| Retrospective | 28.04 | -4.10 | 32.14 | 34.60 | <.001 | < .001 |
| Political Polarization | Prospective | 16.65 | 4.00 | 12.65 | 11.45 | <.001 | < .001 |
| Retrospective | 28.00 | 4.00 | 24.00 | 24.06 | <.001 | < .001 |
| Climate Change | Prospective | 1.52 | -1.00 | 2.52 | 3.08 | .002 | .002 |
| Retrospective | 3.32 | -1.00 | 4.32 | 4.42 | <.001 | < .001 |
| Traditionalism | Prospective | 2.11 | 1.00 | 1.11 | 1.20 | .233 | .240 |
| Retrospective | 4.48 | 1.00 | 3.48 | 3.53 | <.001 | < .001 |
| Violence | Prospective | 2.61 | 24.00 | 21.40 | -21.84 | <.001 | < .001 |
| Retrospective | 10.93 | 24.00 | 13.07 | -13.01 | <.001 | < .001 |
| Charity | Prospective | 5.56 | 4.00 | 1.56 | 1.35 | .179 | .190 |
| Retrospective | 1.37 | 4.00 | 2.63 | -2.51 | .013 | .015 |
|  | Generalized Trust | Prospective | -4.76 | -1.30 | 3.46 | -3.53 | <.001 | < .001 |
|  | Retrospective | -12.54 | -1.30 | 11.24 | -11.03 | <.001 | < .001 |
|  | Individualism | Prospective | 6.48 | 3.40 | 3.08 | 3.22 | .001 | .001 |
|  | Retrospective | 11.00 | 3.40 | 7.61 | 7.85 | <.001 | < .001 |
| Social Scientists | Depression | Prospective | 16.75 | 2.50 | 14.25 | 13.96 | <.001 | < .001 |
| Retrospective | 20.26 | 2.50 | 17.76 | 20.62 | <.001 | < .001 |
| Life Satisfaction | Prospective | -15.13 | 1.00 | 16.13 | -14.60 | <.001 | < .001 |
| Retrospective | -20.59 | 1.00 | 21.59 | -22.24 | <.001 | < .001 |
| Implicit Prejudice | Prospective | 14.00 | -6.2 | 20.20 | 21.10 | <.001 | < .001 |
| Retrospective | 9.82 | -6.2 | 16.02 | 16.40 | <.001 | < .001 |
| Explicit Prejudice | Prospective | 13.16 | -51.00 | 64.16 | 72.06 | <.001 | < .001 |
| Retrospective | 8.38 | -51.00 | 59.38 | 55.44 | <.001 | < .001 |
| Loneliness | Prospective | 15.83 | -4.10 | 19.93 | 15.35 | <.001 | < .001 |
| Retrospective | 24.87 | -4.10 | 28.97 | 30.49 | <.001 | < .001 |
| Political Polarization | Prospective | 18.91 | 4.00 | 14.91 | 13.29 | <.001 | < .001 |
| Retrospective | 25.92 | 4.00 | 21.92 | 21.38 | <.001 | < .001 |
| Climate Change | Prospective | -2.10 | -1.00 | 1.10 | -1.18 | .024 | .025 |
| Retrospective | .59 | -1.00 | 1.59 | 1.91 | .057 | .064 |
| Traditionalism | Prospective | 4.74 | 1.00 | 3.74 | 4.03 | <.001 | < .001 |
| Retrospective | 4.21 | 1.00 | 3.21 | 3.29 | .001 | .001 |
| Violence | Prospective | 1.09 | 24.00 | 22.91 | -22.58 | <.001 | < .001 |
| Retrospective | 2.94 | 24.00 | 21.07 | -23.14 | <.001 | < .001 |
| Charity | Prospective | 3.33 | 4.00 | .67 | -0.57 | .569 | .581 |
| Retrospective | -.05 | 4.00 | 4.05 | -3.83 | <.001 | < .001 |
|  | Generalized Trust | Prospective | -7.17 | -1.30 | 5.87 | -5.94 | <.001 | < .001 |
|  | Retrospective | -16.92 | -1.30 | 15.62 | -17.64 | <.001 | < .001 |
|  | Individualism | Prospective | 2.96 | 3.40 | .44 | -0.41 | .684 | .684 |
|  |  | Retrospective | 6.99 | 3.40 | 3.59 | 4.14 | <.001 | < .001 |

*Note:* Rightmost *p* value column was adjusted for false discovery rate using Benjamini-Hochberg correction.

*Table S21.* Comparisons of Prospective and Retrospective Estimates against Actual Change between January 2020 and October 2020 by Sample and Dimension.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sample | Dimension | Estimate Type | Estimate | Actual Change | *Mdifference* | *t* | *p* | *p\** |
| Lay People | Implicit Prejudice | Prospective | 7.85 | -12.3 | 20.15 | 22.80 | < .001 | < .001 |
|  |  | Retrospective | 12.37 | -12.3 | 24.67 | 25.58 | < .001 | < .001 |
|  | Political Polarization | Prospective | 16.65 | 1.3 | 15.35 | 13.90 | < .001 | < .001 |
|  |  | Retrospective | 28.00 | 1.3 | 26.70 | 26.76 | < .001 | < .001 |
|  | Climate Change | Prospective | 1.52 | -1 | 2.52 | 3.08 | .002 | .002 |
|  |  | Retrospective | 3.32 | -1 | 4.32 | 4.42 | < .001 | < .001 |
|  | Traditionalism | Prospective | 2.11 | 0 | 2.11 | 2.27 | .024 | .025 |
|  |  | Retrospective | 4.48 | 0 | 4.48 | 4.55 | < .001 | < .001 |
|  | Violence | Prospective | 2.61 | 27 | 24.40 | -24.90 | < .001 | < .001 |
|  |  | Retrospective | 10.93 | 27 | 16.07 | -16.00 | < .001 | < .001 |
| Social Scientists | Implicit Prejudice | Prospective | 13.39 | -12.3 | 25.69 | 40.07 | < .001 | < .001 |
|  |  | Retrospective | 9.82 | -12.3 | 22.12 | 22.64 | < .001 | < .001 |
|  | Political Polarization | Prospective | 15.04 | 1.3 | 13.74 | 18.41 | < .001 | < .001 |
|  |  | Retrospective | 25.92 | 1.3 | 24.62 | 24.02 | < .001 | < .001 |
|  | Climate Change | Prospective | -2.97 | -1 | 1.97 | -3.09 | .002 | .002 |
|  |  | Retrospective | 0.60 | -1 | 1.59 | 1.91 | .057 | .057 |
|  | Traditionalism | Prospective | 4.45 | 0 | 4.45 | 6.85 | < .001 | < .001 |
|  |  | Retrospective | 4.21 | 0 | 4.21 | 4.31 | < .001 | < .001 |
|  | Violence | Prospective | 1.09 | 27 | 25.91 | -25.53 | < .001 | < .001 |
|  |  | Retrospective | 2.94 | 27 | 24.07 | -26.44 | < .001 | < .001 |

*Note:* Rightmost *p* value column was adjusted for false discovery rate using Benjamini-Hochberg correction.

*Table S22*. Role of Domain-Related Expertise for Confidence in Prospective and Retrospective Estimates among Social Scientists.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Survey | Domain | *Mdifference* | *SE* | *z* | *p* | *p\** |
| Prospective | Explicit Prejudice | .09 | .07 | 1.32 | .188 | .764 |
|  | Implicit Prejudice | -.06 | .07 | -.89 | .376 | .764 |
|  | Individualism | -.06 | .07 | -.85 | .398 | .764 |
|  | Traditionalism | .06 | .07 | .83 | .408 | .764 |
|  | Generalized Trust | .06 | .07 | .87 | .385 | .764 |
|  | Political Polarization | -.02 | .07 | -.35 | .724 | .984 |
|  | Life Satisfaction | .02 | .07 | .28 | .778 | .984 |
|  | Depression | -.14 | .07 | -1.95 | .051 | .764 |
|  | Delay of Gratification | -.01 | .07 | -.10 | .919 | .984 |
|  | Birth Rate | -.01 | .07 | -.18 | .856 | .984 |
|  | Violence | .00 | .07 | -.01 | .991 | .991 |
|  | Religiosity | .02 | .07 | .27 | .784 | .984 |
|  | Loneliness | .06 | .07 | .84 | .403 | .764 |
|  | Charity | .06 | .07 | .89 | .375 | .764 |
|  | Climate Change | -.01 | .07 | -.16 | .876 | .984 |
| Retrospective | Explicit Prejudice | -.15 | .09 | -1.59 | .111 | .208 |
|  | Implicit Prejudice | -.17 | .09 | -1.85 | .065 | .162 |
|  | Individualism | -.29 | .10 | -2.83 | .005 | .067 |
|  | Traditionalism | .05 | .12 | .44 | .660 | .762 |
|  | Generalized Trust | -.18 | .11 | -1.59 | .111 | .208 |
|  | Political Polarization | -.20 | .10 | -1.96 | .050 | .151 |
|  | Life Satisfaction | .19 | .09 | 2.13 | .033 | .125 |
|  | Depression | -.14 | .09 | -1.49 | .137 | .228 |
|  | Delay of Gratification | -.05 | .10 | -.55 | .582 | .728 |
|  | Birth Rate | -.52 | .22 | -2.30 | .022 | .108 |
|  | Violence | -.13 | .17 | -.77 | .444 | .605 |
|  | Religiosity | -.34 | .13 | -2.61 | .009 | .067 |
|  | Loneliness | -.08 | .10 | -.81 | .415 | .605 |
|  | Charity | .01 | .15 | .04 | .965 | .988 |
|  | Climate Change | .00 | .15 | -.01 | .988 | .988 |

*Note:* Rightmost *p* value column was adjusted for false discovery rate using Benjamini-Hochberg correction.

1. When referring to social scientists, we chiefly focus on behavioral, psychological, and neuroscientists, though our samples also include political scientists, sociologists, and computational social scientists. [↑](#footnote-ref-2)
2. Is it possible that similarity in forecasts between our samples is driven by ad-hoc selection of domains or questionnaire-format response-biases? Supplementary analyses of open-ended top additional domains of societal change suggest it is unlikely. As Figure S1 shows, we observed substantial similarity between social scientists in Studies 1-2 and the lay people in Study 2. [↑](#footnote-ref-3)
3. Analyses with demographic-level covariates (political orientation, ethnicity, age, gender and income) yield close to identical results (see Table S19 & Figure S15). [↑](#footnote-ref-4)
4. The effect also held when controlling for political affiliation, ethnicity, age, gender and income, 3.052 < *Z*s ≤ 7.24, .002 < *p*s ≤ .001. [↑](#footnote-ref-5)
5. Beyond the domains provided in a questionnaire format, analyses of open-ended responses revealed that social scientists identified health and well-being (mental illness, psychological and physical well-being), interconnectedness (romantic relationships, social norms), economics (economic concerns, health care attitudes), social justice (inequality, poverty), child development (education, child development), political discord and mistrust in institutions (science denialism, right-wing orientation) as key domains of pandemic-related societal change (see online supplement). [↑](#footnote-ref-6)