

Perceiving Change in Uncertain Times: Public Accuracy and its Individual Differences in  
Estimating Past Societal Shifts During COVID-19

by

Peter Diep

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Arts

in

Psychology

Waterloo, Ontario, Canada, 2025

© Peter Diep 2025

### **Author's Declaration**

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Abstract

Understanding where society is headed requires a clear grasp of where it has been, to make sense of what factors shape direction of societal change. Yet during periods of heightened uncertainty, such as the COVID-19 pandemic, most people struggle to accurately perceive the direction of societal shifts. In this thesis, I used the first year of the COVID-19 pandemic as a naturalistic context of uncertainty to investigate how the public estimated societal change during such volatile periods, what domains may have been more accurately perceived, and what individual factors may have shaped this accuracy. U.S. participants ( $N = 644$ ) estimated societal change across thirteen domains over six-month (April–October 2020) and one-year periods (April 2020–April 2021), either via free-text or on a -50% to +50% slider, providing their confidence for their estimates per domain. Next, they completed measures assessing general knowledge, confidence in their knowledge judgments, and metacognitive engagement (e.g., reflecting on limits of one’s knowledge when discussing social issues). I further assessed deliberation-related engagement by tracking time spent on each estimation. Results showed that Americans held a largely pessimistic view of societal changes over the pandemic, especially when considering actual change in the domains of depression rates, mortality, violent crimes, unemployment, and charitable donations. Out of all domains, the majority of the sample correctly estimated the direction of societal change for depression rates, life satisfaction, explicit prejudice, charitable giving, and religiosity. However, for other domains participants were either at chance or got the direction of change wrong. Participants were also more accurate in estimation of change for shorter (vs. longer) time frames, and when using open-ended response options (vs. percentage-based slider).

Moreover, individuals showing greater task deliberation, metacognitive engagement, and confidence in their estimates, but not greater general knowledge or calibration in confidence and accuracy of their knowledge, were more accurate. Additionally, effects of knowledge calibration and deliberation varied as a function of domain, either improving estimates or increasing error or bias. These findings suggest that how people engage with information matters more than simply what they know. The thesis concludes by discussing implications for understanding public perceptions of pandemic-related societal changes and identifies factors that may help align these perceptions with actual social trends.

## **Acknowledgements**

I express my sincere gratitude to my supervisor, Dr. Igor Grossmann, for his support, accountability, guidance, and leadership throughout this project and the beginning of my career into psychology. Although I am still far from being an ideal researcher in my own eyes, his insights and wisdom have been integral to my development and without his support, I may never would have started this journey in the first place.

I'd also am grateful to Dr. Cendri Hutcherson, who supported development of this project during the COVID-19 pandemic, data collection, and initial analyzes that would lead to the fruits of this thesis.

I would also like to thank both my thesis reviewers, Dr. Clara Colombatto and Dr. Johnathan Fugelsang. Their efforts and feedback have been highly appreciated for this project. Beyond their formal reviews, their insights and thought-provoking comments, whether in class or during cognitive-area seminars, have been a great source of inspiration for deeper reflection and encouraging me to explore directions I might not have considered on my own.

I extend my gratitude to my fellow students in the Social area, whose guidance and support have been critical for my development and perseverance during this first year. I would especially like to thank my office- and lab-mate Niyati Kachhiyapatel, who has been a great mentor and friend since I've started the program. I'd also like to thank the Wisdom and Culture Lab's post-doc, Dr. Maksim Rudnev, whose been a great guide to learning complex statistics and academia, and our many RAs over the year. I deeply want to thank Dr. Hilary Bergsieker, for her statistical enthusiasm and guidance, and Dr. Neil Hester, for his insights into psychological theory and academia. They have been central to my professional development and to my greater appreciation of the discipline at large.

## **Dedication**

To my late cat, Bella, who passed on October 31<sup>st</sup>, 2024, early into my journey into graduate school. Although I only shared six years of her fourteen, I would never have gone this far without her and continue to miss her each and every day.

To my first and living cat, Nala, who has been a source of comfort throughout this journey. I hope to enjoy every single day with her.

## Table of Contents

Author's Declaration.....	ii
Abstract.....	iii
Acknowledgements.....	iii
Dedication.....	vi
List of Figures.....	ix
List of Tables.....	x
Introduction.....	1
Theoretical and empirical background.....	2
How accurately do individuals assess societal and individual change?.....	2
Who is more likely to accurately estimate societal change?.....	6
Present research.....	11
Methods.....	13
Procedure.....	15
Estimating actual societal change.....	17
Measures.....	18
Directional accuracy.....	18
Deliberation-related task engagement.....	19
General knowledge, confidence in one's knowledge & confidence-knowledge calibration.....	19
Metacognitive engagement in reflections on social interactions.....	20
Demographics.....	21
Analytical procedure.....	22

Results.....	24
What direction did people estimate change in each societal domain? .....	24
How good are people at estimating the direction of societal change and in what domains? .....	26
Impact of response modality and timeframe of estimation .....	28
What individual differences are associated with better estimators of past societal change? .....	30
Discussion .....	36
Limitations .....	46
Conclusion.....	47
References.....	49

## List of Figures

<b>Figure 1.</b> <i>Estimates of societal change over six-month and one-year timeframes compared to actual change in domain</i> .....	25
<b>Figure 2.</b> <i>Estimated marginal probability of directional accuracy in estimates of societal change by domain and timeframe</i> .....	27
<b>Figure 3.</b> <i>Predicted probabilities of directional accuracy by confidence</i> .....	32
<b>Figure 4.</b> <i>Predicted probabilities of directional accuracy by metacognitive engagement</i> ....	32
<b>Figure 5.</b> <i>Predicted probabilities of directional accuracy by deliberative task-engagement</i>	33
<b>Figure 6.</b> <i>Predicted probabilities of directional accuracy by knowledge calibration</i> .....	33
<b>Figure 7.</b> <i>Predicted probabilities of directional accuracy for domain by deliberative engagement and knowledge calibration</i> .....	34

## List of Tables

<b>Table 1.</b> <i>Sociodemographic characteristics of sample</i> .....	14
<b>Table 2.</b> <i>Ground truth values of societal change across domains as percentages</i> .....	18
<b>Table 3.</b> <i>Descriptive statistics and correlations of key variables of interest</i> .....	30
<b>Table 4.</b> <i>Analysis of deviation table (Type 3 Wald <math>\chi^2</math>) of main effects of variables of interests and interactions with domain on estimations of societal change</i> .....	31

## Introduction

Preparing for the future requires having an accurate understanding of the past. This entails not only tracking historical trends but also evaluating the long-term effects of disruptions on such trends. As a result, substantial research and investment has focused on developing indicators to monitor key dimensions of societal change to best prepare for the future (Hák et al., 2016; Land & Michalos, 2018; Trewin & Hall, 2010). Yet despite increasing availability of data, many individuals within the public still misjudge the extent or even the direction of societal change, whether in overestimating decline or underestimating improvement. The public's assessments of where one's society has been heading — such as in perceptions of moral decline (Mastroianni & Gilbert, 2023) or in comparisons between past and present generations (Protzko & Schooler, 2019) — are often inaccurate and susceptible to “rosy retrospection” (Mitchell et al., 1997). In addition, the public's perceptions may be strongly influenced by surrounding media in the present. For example, media coverage of COVID-19 was predominantly negative, with many individuals even preferring to focus on such news over positive changes (Sacerdote et al., 2020). Even experts, who are held with high regard for policy guidance, perform no better than chance when predicting societal trends and may rely on the most salient media or memories to make predictions (Grossmann et al., 2024a; Hutcherson et al., 2023). Together, such misjudgments may be particularly problematic during periods of volatility when the ability to accurately assess societal change becomes especially critical for navigating uncertainty.

Given the high stakes, it is crucial to understand who among the public may accurately perceive societal change and what dispositional factors might explain their ability. Most individuals are disposed to act as “cognitive misers” (Fiske & Taylor, 2020), relying on heuristics and simple intuitions guided by general beliefs and their current understanding of the

world. Yet often perceived as a limitation of the mind, simple models of the world may actually be more accurate, such as by restricting noise (Gigerenzer & Gaissmaier, 2011; Green & Armstrong, 2015; Grossmann et al., 2023; Hertwig & Herzog, 2009).

Others might be more statistically-inclined, deliberate, flexible, and abstract in their thinking; a necessary ability to engage with trends and consider broader perspectives. Such active cognitive engagement has been regarded as highly valuable when making forecasts of future societal change (Mellers, Stone, Murray, et al., 2015). Regardless of approach, the effectiveness of either modes of thinking may also vary on individual characteristics, such as general intelligence, cognitive effort, motivational or affective cues, and metacognitive insight into limits and calibration of one's knowledge. This variability raises some key questions: how accurate are public perceptions of societal change? How does accuracy differ across domains, temporal horizons and estimation modalities? Finally, what individual differences might help identify better estimators of societal change?

### **Theoretical and empirical background**

*How accurately do individuals assess societal and individual change?*

Current research indicates that public perceptions of societal change are often misaligned with objective trends across various domains. Perhaps most notable of this position, Pinker (2018) argues that prevailing societal pessimism misrepresents substantial improvements in areas that have historically challenged human well-being, such as economic inequality and public health. While contemporary threats and emerging crises are undeniably pressing, Pinker (2018) emphasizes that certain historical hardships, including infant mortality, have dramatically declined since the Enlightenment, largely due to scientific progress and rationalism.

Importantly, these misperceptions are not confined to distant historical trends. Even assessments of present-day conditions, such as national income distribution or levels of economic inequality, are frequently inaccurate (Hauser & Norton, 2017). Similarly, perceptions of widespread moral decline of one's society are not reflected in respondents' moral judgments of those around them, suggesting a generalized belief rather than actual observation (Mastroianni & Gilbert, 2023). These distortions extend to beliefs about public attitude as well. For instance, individuals tend to report personal attitude shifts based on beliefs of broader societal attitude change, rather than if any change has actually occurred (Mastroianni & Dana, 2022). If individuals struggle to accurately perceive the present, it is perhaps unsurprising that their understanding of the past may also be skewed.

This tendency to misperceive change is not limited to judgments about society at large; it may even manifest when assessing one's journey. For example, adults' retrospective reports of their childhood health often conflict with objective hospital records (Henry et al., 1994). Likewise, self-assessments of behavioral changes during the COVID-19 pandemic were found to be inconsistent with measured differences across timepoints (Dupuis et al., 2023). It is worth noting that this research focused on substance use and addictive behaviors, which may not fully generalize to broader behavioral domains. Yet, flashbulb memories of one's personal experience during historical events (e.g., 9/11 terrorism attack) are similarly unstable and inconsistent (Hirst et al., 2009). Overall, this research suggests that whether individuals are reflecting on collective trends or personal history, they are prone to systematic distortions in how they perceive and interpret change.

One of the most common expressions of this distorted perception of change is reflected in generational comparisons, framed through nostalgic references to how things were "back in the

day.” Individuals often idealize the past by viewing their own generation as morally or culturally superior, using it as a reference point which younger generations are judged (Protzko & Schooler, 2019). For example, many in the U.S. romanticize the latter half of the 20th century, overlooking substantial challenges that marked the era such as ongoing racial segregation and the beginning of long-term economic stagnation that disproportionately affected the Generation X cohort. This intergenerational bias often is central to (or acts proxy to) broader narratives of societal decline (Protzko & Schooler, 2019).

Some may suggest that these biases may only strongly emerge when making retrospections across longer timeframes, such as in generational comparisons. Accessing information between extremely distant timepoints is likely fragmented or may lack clear continuity between Time A and Time B. Subsequently, information to inform such judgments may be anchored to most accessible, temporally local timepoint. In other words, evaluations tend to be disproportionately shaped by more recent events (Ashby & Rakow, 2014). However, anchoring effects may also occur over shorter intervals, such as between a single year (Godlonton et al., 2018). Given such biases towards the present, one may expect that longer timeframes are more susceptible to inaccuracy.

However, an alternative may be that assessments of longer intervals may yield more accurate estimates given they contain more information. Individuals with higher cognitive ability may be better able to retrieve and integrate this information when making evaluations (Ashby & Rakow, 2014), which may enhance the accuracy of their assessments. Construal level theory also suggests that greater temporal distance encourages more abstract thinking, which could allow individuals to focus on the most relevant features when making judgments rather than be susceptible to local noise (Fujita et al., 2008). However, these salient features may not always be

representative of actual change over time. Shorter intervals, by contrast, encourage more concrete thinking, which may lead to more precise assessments. At the same time, this concreteness also increases susceptibility to “noise” from recent events that overrepresents past trends. Given these opposing possibilities, it remains unclear whether longer or shorter retrospective intervals may yield more accurate estimations.

In addition to the temporal dimension, the way questions are framed may also influence retrospective accuracy. Anchoring effects arise simply through exposure to reference points presented in response scales (Frederick & Mochon, 2012; O’Muirheartaigh et al., 1995) or even occur when individuals are exposed to arbitrary numeric points (Tversky & Kahneman, 1974). Such anchors can influence the accuracy of subsequent estimates, persisting even when respondents are explicitly instructed to disregard the scale values (Berg & Moss, 2022). Given these challenges, free-response formats may improve accuracy given there is no exposure to any pre-existing values. Supporting this, numeric free-responses have been associated with greater precision and accuracy of estimates, as they reduce anchoring and encourage more deliberative answering (Mandel & Irwin, 2021; Mellers, Stone, Atanasov, et al., 2015a). Moreover, research in consumer studies has shown that slider-based responses tend to yield more extreme scores, whereas free responses generally produce more moderate estimates (Thomas & Kyung, 2019), perhaps due to slower deliberation of balances. Thus, shifting response modality may offer a simple yet effective means of enhancing retrospective assessment accuracy at the individual-level.

Together, such patterns suggest that distorted perceptions may not be merely random errors but instead reflect deeper processes and tendencies. Perceptions of one’s past or of one’s society’s may arise not just from cultural narratives or collective memory, but from enduring

psychological traits that shape how individuals interpret change over time. Understanding which individual factors contribute to accurate assessments of change may therefore help explain variation in the public's understanding of societal trends.

*Who is more likely to accurately estimate societal change?*

Although little research has directly examined which individual traits are associated with accurately estimating past societal change, an extensive literature has identified predictors of accuracy in forecasting such trends. For instance, one of the most immediately apparent differences between top and bottom forecasters reflects their cognitive style, such as their tendency to rely on intuition versus deliberative judgment (Mellers, Stone, Murray, et al., 2015). Some individuals prefer to make decisions quickly and spontaneously, while others engage in more effortful and systematic analysis before reaching a conclusion. This is not to say that these modes of thinking are mutually exclusive; rather, they often complement one another (De Neys, 2025; Gigerenzer & Gaissmaier, 2011) and may shift depending on context or task demands (Grossmann et al., 2025; Inbar et al., 2010; Pachur & Spaar, 2015). However, some factors may contribute to persistence of cognitive style across situations, requiring conditions to be met prior to engaging in other modes of thinking (De Neys, 2025; Grossmann et al., 2025). For instance, “superforecasters” — individuals who consistently make accurate predictions across a range of social and political domains — score higher on measures of need for cognition, reflecting enjoyment of and commitment to effortful thinking (Himmelstein et al., 2021). By contrast, those lower on this trait are more likely to default to superficial or intuitive judgments when confronted with complex information. They also may require additional incentives or motivational cues to encourage greater cognitive engagement when given difficult tasks.

Building on these findings, research on forecasting has examined ways to actively promote deliberate analysis, particularly in governmental and intelligence contexts where accuracy of predictions is critical. Interventions include structured forecasting tournaments that encourage effortful engagement and regular practice among intelligence analysts (Mandel & Irwin, 2021). Even relatively simple procedural adjustments, such as requiring forecasters to provide numeric estimates rather than qualitative descriptions, can foster greater precision and reduce interpretive ambiguity to improve performance (Mandel & Irwin, 2021; Mellers, Stone, Atanasov, et al., 2015a).

Yet when such strategies are not available, even experts may default to intuitive judgments, relying on heuristics despite having the knowledge and cognitive capacity to perform more deliberative evaluations, which may result in reduced performance (Hutcherson et al., 2023). However, defaulting to heuristics may not always lead to inadequate outcomes. Superforecasters, for instance, still outperform others even when leaning more heavily on intuition (Mellers, Stone, Murray, et al., 2015). Similarly, skillful intuitions by experts can be distinguished from the overconfidence and bias that often accompany intuitive reasoning in less knowledgeable individuals (Kahneman & Klein, 2009). Indeed, the relative advantages of intuition versus deliberation may even vary with expertise: individuals with moderate knowledge may benefit more from trusting their intuitive judgments as they lack the formal training to effectively apply deliberate strategies to their knowledge (Dijkstra et al., 2013).

If deliberation versus intuition does not necessarily differentiate (at least fully) between high and low performers, what else might? Deliberative cognition might be most valuable or accessible when individuals already have high cognitive ability or high knowledge.

Superforecasters often score highly on measures of domain-general cognitive abilities, novel

problem-solving, and cognitive flexibility (Mellers, Stone, Murray, et al., 2015). Fluid intelligence in particular, referring to the capacity for abstract and active cognitive processing (Cattell, 1963), has repeatedly been identified as a key distinguishing factor among high-performing forecasters, necessary for the integration of new information with existing knowledge and adaptation to unfamiliar domains (Chang et al., 2016; Himmelstein et al., 2021; Mellers, Stone, Atanasov, et al., 2015b).

However, the extent to which general intelligence is necessary for assessing past trends, relative to making predictions, remains unclear. On one hand, retrospective assessments may be comparatively passive, given retrospection focuses on events that have already occurred rather than integrating novel information about future possibilities. Such judgments may therefore rely more heavily on crystallized intelligence, as in the accumulated knowledge and memory of past events (Cattell, 1963), rather than processes based on fluid intelligence. This type of knowledge is necessary for individuals to recall and contextualize historical events, recognize patterns of change, and relate those patterns to current conditions. On the other hand, accurate estimation of past trends may still require the active integration of multiple lines of evidence, which may at times be discrepant or even contradictory, into a coherent and accurate representation of the past. Crucially, integration depends not only on the possession of relevant knowledge and reasoning ability but also on the confidence to evaluate evidence, weight competing interpretations, and filter high-quality information from low-quality or misleading sources.

Confidence, in this sense, may function as both a signal and a potential source of bias in retrospective judgment. When appropriately calibrated, it may reflect genuine mastery of the material and enable more decisive, precise estimations (Friedman et al., 2018). Over time, individuals may even become better calibrated, aligning their confidence with actual accuracy as

they gain experience and learn to identify accurate cues of confidence-accuracy calibration (Moore et al., 2017). Greater confidence in predictions is also associated with higher directional accuracy in forecasts, though this relationship tends to be weaker in experts who may be more cautious in their own judgments given their broader knowledge (Hutcherson et al., 2023).

Yet confidence is not always a reliable indicator of accuracy. It may instead stem from task familiarity or the apparent clarity of a trend rather than genuine insight. In such cases, high confidence can foster overconfidence, narrowing attention to confirmatory evidence and reducing openness to contradictory information, which experts and superforecasters may be protected against due to their knowledge or cognitive strategies (Hutcherson et al., 2023; Mellers, Stone, Murray, et al., 2015). The role of information availability in shaping this relationship in laypeople is mixed, however. Some research suggests that having more information can temper overconfidence by prompting greater deliberation or hesitation before judgment (Peterson & Pitz, 1986). Other studies, however, find that confidence can increase as more information is gathered, even when accuracy remains unchanged (Fleisig, 2011; Tsai et al., 2008). These conflicting findings underscore that confidence alone cannot be taken as a straightforward proxy for judgmental accuracy.

Given these limitations, neither cognitive style nor knowledge alone may be sufficient to ensure accurate judgment, especially when prior knowledge and biases complicate the task. This points to the importance of a deeper layer of processing: metacognition. Metacognition often involves critical engagement with one's cognition, such as recognizing the limits of one's knowledge, questioning one's assumptions, and remaining open to alternative interpretations (Grossmann et al., 2020; Grossmann & Johnson, 2025). Such processes have been identified as crucial for actively engaging with complexity in forecasting (Mandel & Irwin, 2021; Mellers,

Stone, Murray, et al., 2015; Tetlock, 1999). For instance, superforecasters often score higher on metacognitive traits such as open-mindedness, which facilitates the consideration of new information, alternative perspectives, and counterfactuals (Mellers, Stone, Atanasov, et al., 2015a; Mellers, Stone, Murray, et al., 2015). Notably, experts, who may already possess requisite knowledge, often identify these metacognitive dispositions as critical to their success when evaluating uncertain or evolving societal conditions (Grossmann et al., 2022).

Metacognition also plays a key role in how individuals organize and update broader mental models of societal development, such as folk theories of social change (Kashima et al., 2009). These theories are culturally shared, intuitive models that guide one's understanding of societal evolution. These folk theories must integrate specific historical knowledge into broader present-day trends to mold one's expectations about future societal trajectories, similar to gauging the impact of disruptions on stable historical trends. Importantly, this integration requires reflective engagement with evidence, a willingness to question assumptions, and the flexibility to revise one's beliefs in response to new information.

In contrast, others might simply rely on "primal world beliefs" (Clifton, 2020) that reflect their general pessimism or optimism about society and its direction. Such beliefs are often reactive and emotionally-grounded, lacking critical perspective and resist awareness of changing times or world dynamics. These reflexive worldviews may lead to distorted interpretations of social change, particularly when individuals fail to integrate information contrary to their strongly-held beliefs. Thus, the capacity for metacognitive regulation may determine whether individuals hold rigid, affect-driven assumptions or develop more nuanced understandings of societal trends, necessary for accurate estimation.

Taken together, prior research suggests that accurate perceptions of societal change may be shaped by a combination of cognitive style, knowledge, and metacognitive capacities.

Whereas forecasting accuracy has been associated with traits such as intelligence, confidence, and reflective thinking, it remains unclear whether these same characteristics also influence the public's ability to estimate the direction of societal change *ex post* (after it occurred). Whereas *ex ante* forecasts may benefit from careful interpolation from past trends, consideration of base rates, variability in historical trends, *ex post* retrocasts may uniquely benefit from sensitivity to the changes in the world one has lived through. Among these processes, retrocasting may especially benefit from active cognitive engagement that explicitly incorporates metacognition, reflecting on and critically evaluating one's own knowledge and memories. This involves integrating and critically evaluating both personal and local experiences with broader collective narrative to construct a nuanced and accurate understanding of social change.

### **Present research**

In this thesis, I address key gaps in the literature by investigating the public's ability to accurately assess the direction of societal change during the first year of the COVID-19 pandemic. A critical step is to first understand how anchoring effects might influence retrospections of societal change, in relation to response modality (free-response versus sliding scale points) and retrospective timeframes. Addressing these methodological factors is essential before evaluating the accuracy of such retrospections. Next, given the widespread uncertainty and social disruption during this time, it remains unclear how the public perceived change, whether pessimistically or optimistically, during a period of isolation and distorted information. Moreover, how accurate was the public in tracking change when isolated and overloaded with largely negatively-framed information? Furthermore, it is unknown whether this accuracy was

consistent across different domains. I then examine which factors may be associated with more accurate perceptions of societal change. In particular, I examine whether cognitive style (preference for deliberation, indexed by time spent on estimates), general knowledge accuracy, and metacognitive characteristics (e.g., intellectual humility and consideration of different perspectives).

## Methods

This study was reviewed and received ethical clearance through the University of Waterloo Research Ethics Committee (ORE #42123). This study was preregistered on the Open Science Framework, which includes materials, methods, and data at <https://osf.io/q3cga>. The pre-registration included criteria for participant entry validation (see below). Research questions and analyses from this thesis were not included in the initial pre-registration.

### Participants

During April 2021, individuals from the U.S. public were randomly sampled using Prolific. An American sample was utilized given greater data availability in the U.S. of societal domains of interest to compare estimates against. Individuals were eligible for participation if they indicated that they currently lived in the U.S. and were native English speakers in Prolific pre-screening. Based on these criteria, a total of 1136 individuals were initially recruited, receiving £1.50 for study completion. From this pool, I excluded participants based on preregistration criteria: (i) noncompletion, (ii) if total duration of completion was under 3 *SDs* under mean completion time (indicating low effort responses, as per preregistration), (iii) failed initial comprehension check of mathematical understanding of percentages, (iv) provided nonsensical answers (e.g., age > 900), or (v) indicated having completed a similar survey previously.

After applying exclusion criteria, the final sample consisted of 644 lay individuals, ranging from ages 18 to 79 ( $M = 44.7$ ,  $SD = 16.3$ ; see Table 1 for other demographics). In terms of their politics (all scales ranging: 1 = *very liberal*, 7 = *very conservative*), participants averaged center-left in their overall views ( $M = 3.3$ ,  $SD = 1.8$ ), and similarly in their social ( $M = 3.0$ ,  $SD = 1.8$ ) and economic values ( $M = 3.5$ ,  $SD = 1.8$ ).

**Table 1***Sociodemographic characteristics of sample*

Sociodemographic variable	<i>n</i>	%
Gender		
Man	305	49.5
Women	319	47.4
Other/No response	20	3.1
Ethnicity		
Aboriginal/Native	1	< 1
Asian	47	7.3
Black	82	12.7
White	426	66.1
Middle Eastern	6	< 1
Hispanic	45	7.0
East Indian	4	< 1
Mixed race	25	3.9
Other/No response	8	1.2
Income		
Under \$15,000	48	7.5
\$15,001 - \$25,000	60	9.3
\$25,001 - \$35,000	67	10.4
\$35,001 - \$50,000	84	13.0
\$50,001 - \$75,000	133	20.7
\$75,001 - \$100,000	108	16.8
\$100,001 - \$150,000	82	12.7
Over \$150,000	55	8.5
Political agreement		
Democrats	378	58.7
Republicans	127	19.7
Neither/No response	139	21.6

Note: *N* = 644

## Procedure

The study design involved a 2 (between-person: *response modality*)  $\times$  2 (repeated measures: *timeframe*)  $\times$  13 (repeated measures: *domain*) mixed-design experiment. In a between-subject part of the experiment, participants were randomly-assigned into two response conditions: either providing their estimates openly (i.e., ‘free response’) or using a slider scale (i.e., ‘slider’) ranging from -50% to 50% change, but still allowing similar response granularity as free estimates. Although the role of response modality was not relevant to research questions of interest for this thesis, I assessed the impact of this experimental manipulation to control for later analyses. To test the role of timeframe in a repeated measures component, participants provided estimates of societal change for two periods of time in each of the thirteen domains: a six-month period (April 2020 to October 2020) and a one-year period (April 2020 to 2021). These timeframes were selected as data from this study was used to follow-up and make comparisons with findings from Hutcherson and colleagues (2023).

Participants were provided with instructions on how to make estimates of change using percentages, which were confirmed/corrected in a comprehension check. Specifically, participants were first provided with two examples of estimating an increase and decrease:

“For example, imagine estimating how the frequency of cooking at home changed after April 2020. If you think that people on average cooked their meals at home 4 nights a week in April, but in October/November they cooked at home 7 nights a week, this would represent an increase of 75% (i.e., 3 additional days/4 original days = +75%).

Similarly, imagine that you are asked to estimate the percent change in usage of public transportation. If you think that 1000 people used public transportation per day in April of 2020, at the beginning of the pandemic, and this dropped to 500 people per day in

October/November of 2020, this would represent a decrease of 50% (e.g., 500 fewer people/1000 people originally = -50%).”

Afterwards, participants then were asked to estimate the change in weekly social gatherings if they had estimated that this occurred twice per week in April 2020 but then reduced to zero in October/November. Participants who selected ‘a decrease of 100%’ out of the five available options successfully passed the comprehension check. As the societal change estimation task depended on participants’ full understanding how to make percentage estimates, it was imperative that participants included for analyses successfully completed this task.

Following this check, participants assessed thirteen domains of U.S. societal change since April 2020 in randomized order: all-cause mortality, birth rates, charitable giving, clinical depression rates, explicit prejudice towards minorities, generalized trust in others, life satisfaction, political polarization, public concern for climate change, religiosity, trust in U.S. institutions, unemployment rates, and violent crime rates. Although these domains are unlikely to reflect all areas relevant to societal change, these domains were chosen to encompass a broad range of societal concerns, possibly affected by the COVID-19 pandemic, and had greater data availability to identify objective estimates. Some domains were built on earlier scholarship from Hutcherson and colleagues (2023), allowing for comparative analyses between studies. Others sought to better capture domains of societal interest for which sociological, psychological and economic theories about the role of the pandemic on social and psychological processes exist (Grossmann et al., 2023; Hutcherson et al., 2023).

For each domain, participants also provided a confidence rating of the accuracy in their estimates (1 = *not at all* to 5 = *extremely*). After providing all estimates, participants completed additional measures, including social metacognitive engagement assessed by the Situated Wise

Reasoning Scale (SWiS; Brienza et al., 2018), general knowledge surrounding different facts about the world (e.g., history, geography, culture), along with confidence judgment into accuracy of one's knowledge. Finally, participants provided demographics and were debriefed.

### **Estimating actual societal change**

To assess whether participants' estimates of societal change were accurate to actual societal change, I collected ground truth estimates across domains during April 2020, October 2020, and April 2021 timepoints (see Table 2 for ground truth values). Out of the thirteen domains, I was able to identify reliable ground truth estimates of change for both timeframes for ten domains. One-year change estimates could only be obtained for charitable giving and religiosity, and no estimates of change could be obtained for generalized trust in others. Data chosen for each domain were selected based on reliability of the sources as used by Hutcherson and colleagues (2023). Moreover, sources that could provide estimates for all or near all timepoints (i.e., April 2020, October 2020, and April 2021) were prioritized.

Estimates for birth rates, depression rates, and all-cause mortality were collected from the Center of Disease Control (CDC). Data for charitable giving, life satisfaction, political polarization, and religiosity were collected from Gallup poll datasets. Trust in institution estimates were gathered from Pew Research, unemployment rates from the U.S. Bureau of Labor Statistics, violent crime rates from the Federal Bureau of Investigation, and concerns about climate change from the Yale Program on Climate Change Communication.

Estimates of explicit prejudice were calculated as the average of Black U.S. and Asian U.S. explicit bias scores from Project Implicit during the whole month of each time point. A Population-weighting based on specifications of the method from the American National Election Studies was applied based on age, sex, politics using *anesrake* R-package (Pasek,

2018), following similar transformations by Hutcherson and colleagues (2023) to better reflect American population demographics during 2020 and 2021. Participants under the age of eighteen were excluded from these calculations.

As Equation (1) shows, I followed Hutcherson and colleagues (2022) method for calculating the percentage of change between timepoints of interest (represented as  $x$ ) from April 2020: October 2020 for six-month periods and April 2021 for one-year periods.

$$\Delta Domain\ change = \left( \frac{x - \text{April 2020 estimate}}{\text{April 2020 estimate}} \right) \times 100 \quad (1)$$

**Table 2**  
*Ground truth values of societal change across domains as percentages*

Domain	Change in 6 months	Change in 1 year	Source
All-cause mortality	-15.0%	-20.3%	CDC
Birth rates	5.0%	0.9%	CDC
Charitable giving	—	457.2%	Gallup
Clinical depression	10.6%	4.2%	CDC
Explicit prejudice	51.5%	43.2%	Project Implicit
Life satisfaction	-0.3%	9.4%	Gallup
Political polarization	6.4%	-2.9%	Gallup
Public concern for climate change	0.1%	-2.1%	Yale Program on Climate Change Communication
Religiosity	—	4.1%	Gallup
Trust in institutions	-25.9%	-11.1%	Pew Research
Unemployment rates	-54.1%	-58.8%	U.S. Bureau of Labor Statistics
Violent crimes	26.6%	-11.8%	Federal Bureau of Investigation

*Note:* No estimates could be found for the ‘Generalized trust’ domain.

## Measures

### *Directional accuracy*

Consistent with Hutcherson and colleagues (2022), my primary criterion of accuracy concerned directional accuracy, as in whether participants had an accurate “gist” of societal change. The assumption behind this measure is that estimating the specific percentage magnitude of change or even point estimate may be challenging even for most erudite experts, whereas

general understanding of the direction of change may be more realistic (Hutcherson et al., 2022). Past research also indicated larger influences of overestimation when accessing magnitude, resulting in large errors that discouraged assessments of accuracy (Hutcherson et al., 2022). Directional accuracy was operationalized as whether the sign of the participant's estimate matched the direction of actual societal change, as determined by objective truth values. For example, if a participant estimated a 45% increase and the objective data showed a 20% increase, the participant was considered directionally accurate despite inaccurate in magnitude. Reliability alpha suggested high internal consistency for both free-response,  $\alpha = .80$ , and slider,  $\alpha = .92$ , versions of estimations.

#### *Deliberation-related task engagement*

I operationalized deliberation-related engagement on the task by estimating the duration of time (in seconds) taken to complete estimates for each domain. As I was interested only in the quantitative data for these analyses, I was restricted from using additional measures such as word count in qualitative responses. Given its nonnormal distribution (skew = 12.33, kurtosis = 289.33, following guidelines suggested by Kline, 2023), effort was log-transformed to produce a normal distribution necessary for regression analyses (skew = 0.36, kurtosis = 6.78). Log duration was then mean-centered.

#### *General knowledge, confidence in one's knowledge & confidence-knowledge calibration*

General knowledge was assessed using a 20-item inventory of common cultural facts (e.g., "which [planet] is bigger?"). For each question, participants selected from two options and indicated their confidence in the accuracy of their answer, ranging from 50% (i.e., at chance) to 100%. Accuracy of general knowledge was calculated as the percentage of correct answers per participant and mean-centered. Reliability alpha across general knowledge items revealed high

internal consistency,  $\alpha = .88$ . While this task does not represent the most rigorous assessment of intelligence compared to standardized fluid intelligence tests, it offers a practical and accessible means of evaluating participants' general world knowledge — an important facet of crystallized intelligence relevant to the broader public.

In addition, I also assessed participants' overall confidence when assessing their general knowledge accuracy. I took the average of participants' confidence when answering general knowledge questions, which ranged from 50% (i.e., just a guess) to 100% (i.e., extremely confident). Reliability alpha indicated high internal consistency across items,  $\alpha = .91$ .

I was also interested in individual differences in knowledge-confidence calibration — i.e., the extent to which participants' confidence in the accuracy of their answer on the general knowledge questionnaire tracked their accuracy. To this end, I ran a logistic generalized mixed-effects models assessing accuracy ( $1 = correct$ ,  $0 = incorrect$ ) with confidence as a fixed effect and participant as a random intercept. I then estimated participants' random slopes as a metric of calibration between their accuracy and their confidence<sup>1</sup>.

#### *Metacognitive engagement in reflections on social interactions*

Metacognitive engagement was measured using a short 10-item version of the Situated Wise Reasoning Scale (SWiS; Brienza et al., 2018). For this instrument, participants were asked to recall an interaction with another person whose opinion differed from the participant's opinion on the usage of proof of vaccination documents for accessing public places (e.g., airports, restaurants, concert venues). Following the event reconstruction to avoid memory recall bias

---

<sup>1</sup> I used this approach instead of typical signal detection methods (e.g., meta- $d'$ ) for metacognitive sensitivity due to the absence of direct stimuli and low trial count, well below the 100–200 trials typical, which can result in unstable or even below-chance estimates (Maniscalco et al., 2016; Rahnev, 2025).

(Kahneman et al., 2004; Schwarz et al., 2009), participants reported the location, their actions, their thoughts, and their feelings during the recalled or imagined encounter.

Participants then reported their engagement in metacognitive behaviours during this situation. These included the extent they recognized change or alternatives, engaged with intellectual humility, sought compromise, considered the others' perspective, and considered an outsiders' perspective of the interaction (1 = *not at all* to 5 = *very much*). The SWiS was employed to address limitations observed in prior measures of metacognitive engagement (or "wise reasoning"). Unlike trait-based approaches, the SWiS adopts a situated perspective, grounding responses to specific issues in their social context, rather than abstract self-assessments (Brienza et al., 2018).

Participants who demonstrated 'flatlined' responses (i.e., using the same response option across all items) were excluded from these analyses due to response bias. This resulted in  $n = 591$  for all SWIS-related analyses. Scores across items showed very high internal consistency, reliability  $\alpha = .89$ . Thus, I averaged these scores into a single index of domain-independent metacognitive engagement for further analysis<sup>2</sup>.

### *Demographics*

At the end of the study, participants provided their demographic information, including their age, gender (1 = *women*, 2 = *men*, 3 = *other*), and political orientation. To assess political orientation, participants responses to questions concerning their general political agreement between Democrats versus Republicans ("With respect to politics, I tend to agree more often with positions taken by:" 1 = *Democrats*, 2 = *Republicans*, 3 = *neither*), their overall political

---

<sup>2</sup> CFA for a three-factor model for metacognition, perspective-taking, and conflict resolution indicated poor fit given the data. EFAs suggesting two-four factor models also indicated inadequate fit and produced incoherent dimensions.

orientation (“If I had to describe my political views overall, I would say that I am:” 1 = *very liberal*, 7 = *very conservative*), and their economic and social values (“With respect to [economic/social] matters, I consider myself:” 1 = *very liberal*, 7 = *very conservative*).

Excluding general political agreement, responses to the latter rating-scale questions were highly correlated ( $r_s \geq .81$ ) and thus averaged into an index of political orientation (excluding party leaning from this index), with higher scores reflecting greater conservatism.

### **Analytical procedure**

I conducted all analyses in R (version 4.4.0, R Core Team, 2024) through RStudio (Posit Team, 2025). Because of the binary nature of the criterion variable (1 = *correct direction*, 0 = *incorrect direction*), I relied on generalized linear mixed-effects models (logit family) with the *lme4* R-package (Bates et al., 2015). I calculated estimated marginal probabilities and post hoc comparisons for domain comparisons using the *emmeans* R-package (Lenth, 2025). In addition, I visualized findings using *ggplot2* R-package (Wilkinson, 2011). I also assessed variables for normality and mean-centered independent variables for interactions, excluding directional accuracy, which follows a binomial distribution.

In my analysis, I first examined how participants perceived change across domains of societal interest to get a sense of whether the public’s pessimism or optimism in each domain. To do this, I calculated the median rating of participants’ estimates which I then plotted for each period, contrasting estimates with the objective values.

After analyzing the public’s views of societal change, I determined participants’ overall directional accuracy in making their societal estimates. To this end, I fit a null model, only accounting for participant variability as a random intercept. Next, I fitted a model examining the impact of timeframe of estimates and response condition as covariates, as well as domain-

specific interactions with these predictors. To evaluate how individual differences related to accuracy in perceptions of societal change, I extended the model to include all predictors of interest, along with domain-specific interactions. Whenever I observed a significant domain-specific interaction, I estimated marginal probabilities of directional accuracy for each domain and transformed them into percentages for interpretability, using *emmeans* and *emtrends* packages in R (Lenth, 2025), with Tukey-adjusted significance estimates from pairwise comparisons of accuracy across domains.

Although effect sizes should be interpreted based on previous findings of specific relationships within the literature, I followed a few guidelines as a basis for initial interpretations (Flora, 2020; Funder & Ozer, 2019). In the context of individual differences research, Gignac and Szodorai (2016) have found that  $r = .11$ ,  $r = .19$ , and  $r = .29$  reflected correlations at the 25th, 50th, and 75th percentiles across 708 correlations, which I will interpret as small, medium, and large effects respectively.

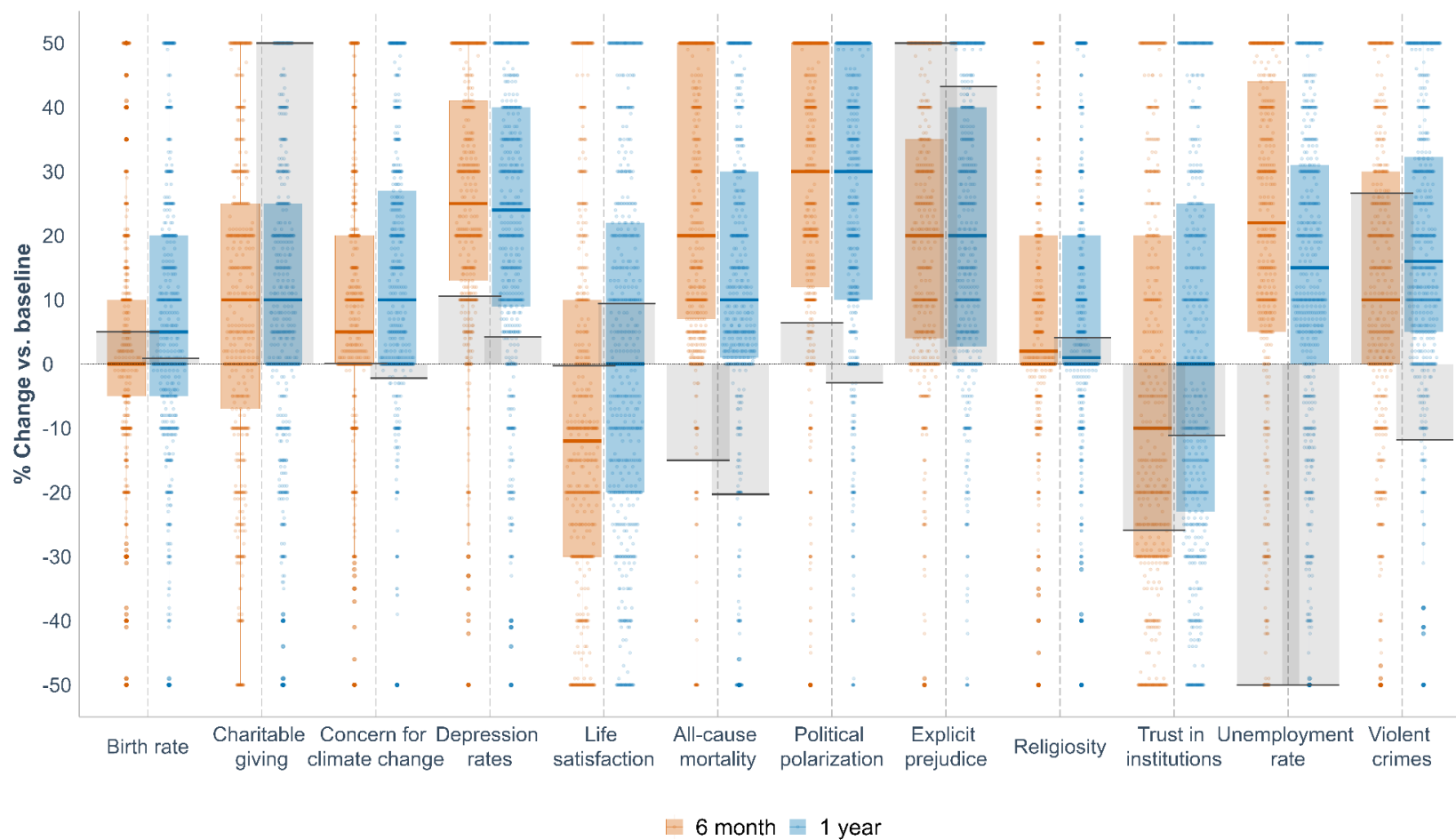
## Results

### What direction did people estimate change in each societal domain?

Out of the twelve domains, only three domains (i.e., birth rate, concern for climate change, and religiosity) lacked a clear *a priori* direction of pessimism/optimism. Consequently, I did not use these domains to examine the public's general attitudes/beliefs on societal change. As shown in Figure 1, participants held generally negative perceptions about the direction of societal change. Participants tended to perceive substantial increases for depression rates, all-cause mortality, political polarization, unemployment, and one-year estimates of violent crime. However, actual change in these domains tended to be either relatively minor (e.g., depression rates) or occurred in the opposite direction (e.g., all-cause mortality). In addition, participants also perceived less positive changes in life satisfaction and charitable giving compared to actual change. The only domains which suggested some optimism in the public's perceptions are explicit prejudice (perceived increasing much less than actual) trust in institutions (perceived as decreasing much less than actual), and six-month estimates of violent crime rates.

**Figure 1**

*Estimates of societal change over six-month and one-year timeframes compared to actual change in domain*



*Note:* Bars represent objective change for each period if available, with grey reflecting difference from 0%. Objective change values greater than +/-50 limited to scale limits. Box plots represent median and 25/75% CIs

## How good are people at estimating the direction of societal change and in what domains?

To assess general directional accuracy across all tested domains, I assessed a model which only included participants as a random intercept. The intercept-only model estimated participants' baseline probability of directional accuracy at 49.4%,  $SE = 0.4%$ ,  $z = -1.42$ ,  $p = .156$ . In other words, there was null evidence that participants' estimates were any better than chance. This interpretation is further supported by the lack of ICCs, suggesting that there was no systematicity in responses across domain. At the same time, as Figure 2 demonstrates, accuracy varied by domain,  $\chi^2(11, N = 644) = 2148.84$ ,  $p < .001$ . Participants showed above chance accuracy in estimating societal change for five out of twelve domains — i.e., for depression rates, explicit prejudice, charitable giving, religiosity, and life satisfaction. On the other hand, estimates of change for political polarization, trust in institutions, and birth rates were at chance, with estimates of all-cause mortality, violent crimes, unemployment rates, and concern for climate change trailing below chance<sup>3</sup>.

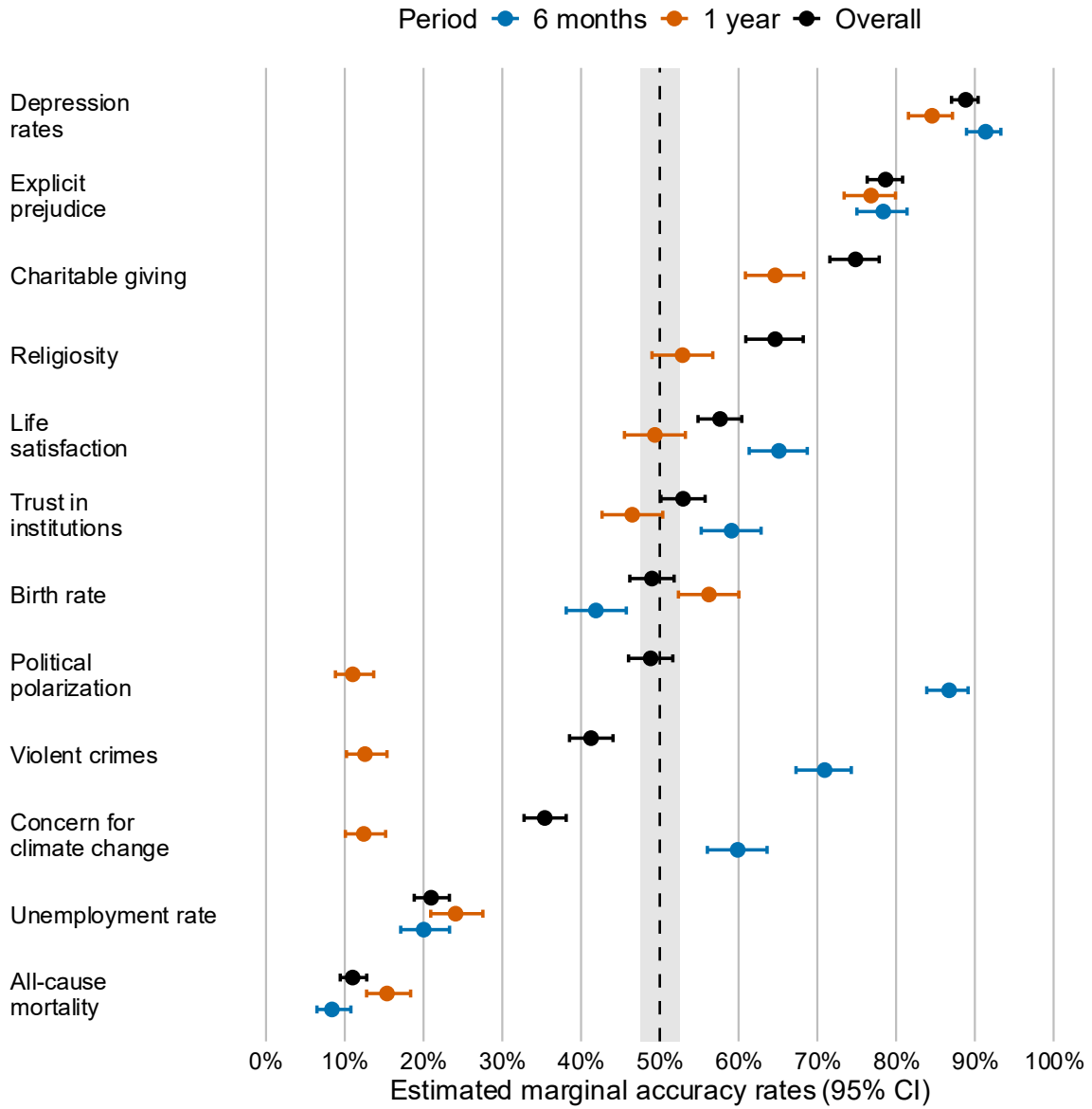
Post-hoc Tukey tests corroborated descriptive observations from Figure 2,  $r$ s ranging from  $-.68$  (i.e., all-cause mortality to explicit prejudice) to  $.75$  (i.e., depression rates vs. mortality),  $ps \leq .010$ . Again, there was strong evidence that depression and explicit prejudice domains were most accurate while all-cause mortality and unemployment were least accurate.

---

<sup>3</sup> We also found differences in domain estimates in terms of timeframe. When evaluating accuracy across timeframes, both depression rates and explicit prejudice demonstrated consistently high accuracy, with directional estimates significantly above chance at both six-month and one-year intervals. In contrast, estimates for all-cause mortality and unemployment rates consistently fall below chance levels, regardless of the timeframe.

**Figure 2**

*Estimated marginal probability of directional accuracy in estimates of societal change by domain and timeframe*



### **Impact of response modality and timeframe of estimation**

In addition to domain, I also assessed whether experimental conditions had any significant effect on their estimates. Both response condition,  $\chi^2(1, N = 644) = 4.21, p = .040$ , and timeframe for estimation,  $\chi^2(1, N = 644) = 573.77, p < .001$ , influenced directional accuracy of estimates. Specifically, slider estimates (51.2%,  $SE = 0.1\%$ ), were less likely to be directionally accurate compared to free-response estimates (53.1%,  $SE = 0.1\%$ ),  $r = .02, SE = .01, z = 2.05, p = .040$ . On the other hand, shorter periods for estimation (64%,  $SE = 0.1\%$ ) were more likely to be accurate compared to longer periods (40.1%,  $SE = 0.1\%$ ),  $r = .26, SE = 0.01, z = 23.95, p < .001$ . Consequently, I included these conditions as controls in the following model including individual differences.

### **Descriptives and correlations between key variables**

As Table 3 shows, participants held moderate confidence in their estimates (1 = *not at all* to 5 = *extremely*), spending 18.52 seconds ( $SD = 29.00$ ) on average when making estimates for each domain. Moreover, participants performed better than chance on average across general knowledge questions, also showing moderate confidence in these questions, and had little miscalibration when answering. Participants also showed moderate metacognitive engagement when considering interactions relevant to relevant societal changes (1 = *not at all* to 5 = *very much*).

Following conventional benchmarks for effect sizes in individual differences research (Gignac & Szodorai, 2016), confidence in one's estimates of societal change was highly negatively associated with knowledge calibration,  $r = -.28$ . In contrast, confidence in estimates demonstrated a moderate positive association with metacognitive engagement measured by the SWiS,  $r = .15$ . Knowledge calibration and metacognitive engagement within the social domain

were moderately and negatively associated with one another,  $r = -.19$ . This suggests that while both may be related, such measures may not capture evaluative processes in the same way.

General knowledge was moderately and negatively associated with confidence in estimates,  $r = -.17$ , which suggests that confidence in societal estimates was more closely linked to subjective processes (such as metacognitive engagement) than to objective cognitive ability. Interestingly, calibration general knowledge was highly positively associated with general knowledge accuracy,  $r = .86$ , whereas metacognition was only weakly and negatively associated with general knowledge,  $r = -.13$ . Again, this may suggest greater domain specificity between the two measures of active, evaluative processing. Knowledge calibration was also weak but positively associated with deliberation on estimates,  $r = .13$ , whereas metacognition was unrelated,  $r < .01$ . Knowledge confidence was weakly but also positively associated with confidence in estimates,  $r = .13$ . This may be due to differences in task difficulty, as estimating societal change is likely a much more complex task compared to a two-choice test of general knowledge.

**Table 3***Descriptive statistics and correlations of key variables of interest*

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6
1. General knowledge	0.69	0.16	—					
2. Deliberation on task (log[time])	18.52	29.00	<.01	—				
3. Confidence in estimates	2.68	1.02	-.17	<-.01	—			
4. Knowledge confidence	64.14	26.49	.26	.02	.13	—		
5. Knowledge calibration	0.16	0.04	.86	.13	-.28	.05	—	
6. Metacognition (SWiS)	2.96	0.88	-.13	<.01	.15	.07	-.19	—

*Note:* All variables were mean-centered, *Ms* = raw scores prior to centering. *M* of Log(duration) also reflects duration prior to any log-transformations.

### What individual differences are associated with better estimators of past societal change?

As findings between models are fairly consistent (see Table 4), reported results will be based on findings post demographic controls unless specified. First, there was moderate evidence of a positive association of greater confidence in estimates with directional accuracy,  $\chi^2(1, N = 583) = 4.16, p = .035$  (see Figure 3). Similarly, there was strong evidence of a positive relationship between metacognitive engagement and directional accuracy,  $\chi^2(1, N = 583) = 8.76, p = .004$  (see Figure 4). Deliberative task-engagement was also found to be positively associated with directional accuracy,  $\chi^2(1, N = 583) = 13.48, p < .001$  (see Figure 5). Conversely, there was only marginal evidence that knowledge calibration was inversely related to directional accuracy,  $\chi^2(1, N = 583) = 3.82, p = .051$  (Figure 6). This discrepancy between different forms of metacognitive engagement may suggest limited generalizability of between domains<sup>4</sup>.

<sup>4</sup> There was also moderate evidence of an inverse association between age and directional accuracy, suggesting that younger cohorts outperformed older cohorts at estimating prior societal change.

**Table 4**

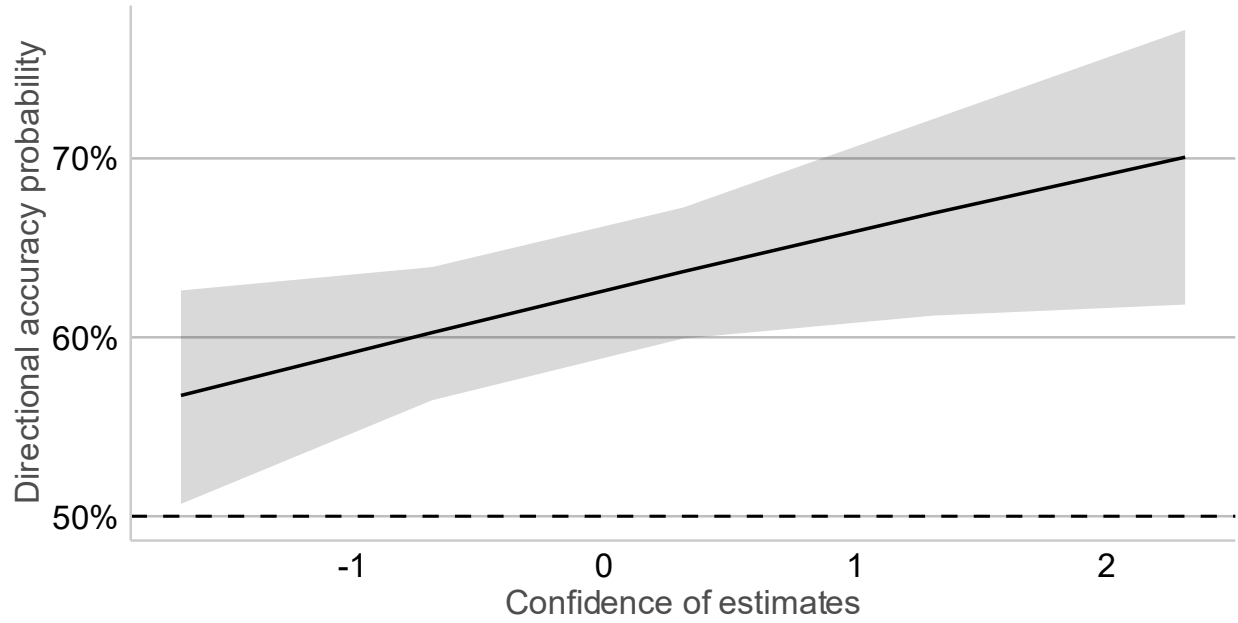
*Analysis of deviation table (Type 3 Wald  $\chi^2$ ) of main effects of variables of interests and interactions with domain on estimations of societal change*

Predictor	Model 1 (Key predictors)			Model 2 (+ Demographics)		
	$\chi^2$	<i>df</i>	<i>p</i>	$\chi^2$	<i>df</i>	<i>p</i>
Intercept	<b>56.27</b>	<b>1</b>	<b>&lt;.001</b>	<b>32.10</b>	<b>1</b>	<b>&lt;.001</b>
Period (vs 6 months)	<b>546.26</b>	<b>1</b>	<b>&lt;.001</b>	<b>540.41</b>	<b>1</b>	<b>&lt;.001</b>
Response condition	<b>4.37</b>	<b>1</b>	<b>.037</b>	<b>4.16</b>	<b>1</b>	<b>.041</b>
Confidence in estimates	<b>5.79</b>	<b>1</b>	<b>.016</b>	<b>4.44</b>	<b>1</b>	<b>.035</b>
Deliberation on task	<b>13.11</b>	<b>1</b>	<b>&lt;.001</b>	<b>13.48</b>	<b>1</b>	<b>&lt;.001</b>
General knowledge	2.60	1	.107	1.44	1	.230
Knowledge calibration	<b>4.92</b>	<b>1</b>	<b>.027</b>	3.82	1	.051
Knowledge confidence	0.15	1	.701	0.07	1	.791
Metacognition	<b>9.74</b>	<b>1</b>	<b>.002</b>	<b>8.76</b>	<b>1</b>	<b>.003</b>
Domain ×	<b>1922.89</b>	<b>11</b>	<b>&lt;.001</b>	<b>1908.19</b>	<b>11</b>	<b>&lt;.001</b>
Confidence in estimates	8.21	11	.694	6.88	11	.809
Deliberation on task	<b>67.45</b>	<b>11</b>	<b>&lt;.001</b>	<b>70.17</b>	<b>11</b>	<b>&lt;.001</b>
General knowledge	13.25	11	.277	14.65	11	.199
Knowledge calibration	<b>49.41</b>	<b>11</b>	<b>&lt;.001</b>	<b>42.81</b>	<b>11</b>	<b>&lt;.001</b>
Knowledge confidence	13.88	11	.240	14.71	11	.196
Metacognition	17.10	11	.105	17.27	11	.100
Age				<b>6.57</b>	<b>1</b>	<b>.010</b>
Gender				0.76	1	.383
Income				0.98	1	.321
Political orientation				2.71	1	.100
<b>Random effects</b>						
$\sigma^2$	3.29			3.29		
$\tau_{00}$ ResponseId	0.00			0.00		
<i>ICC</i>	0.00			0.00		
Observations	12991			12815		
<i>N</i>	591			583		
AIC	14719			14500		
Marginal $R^2$ / Condition $R^2$	0.330 / —			0.332 / —		

*Note:* Bold indicates significant effect.

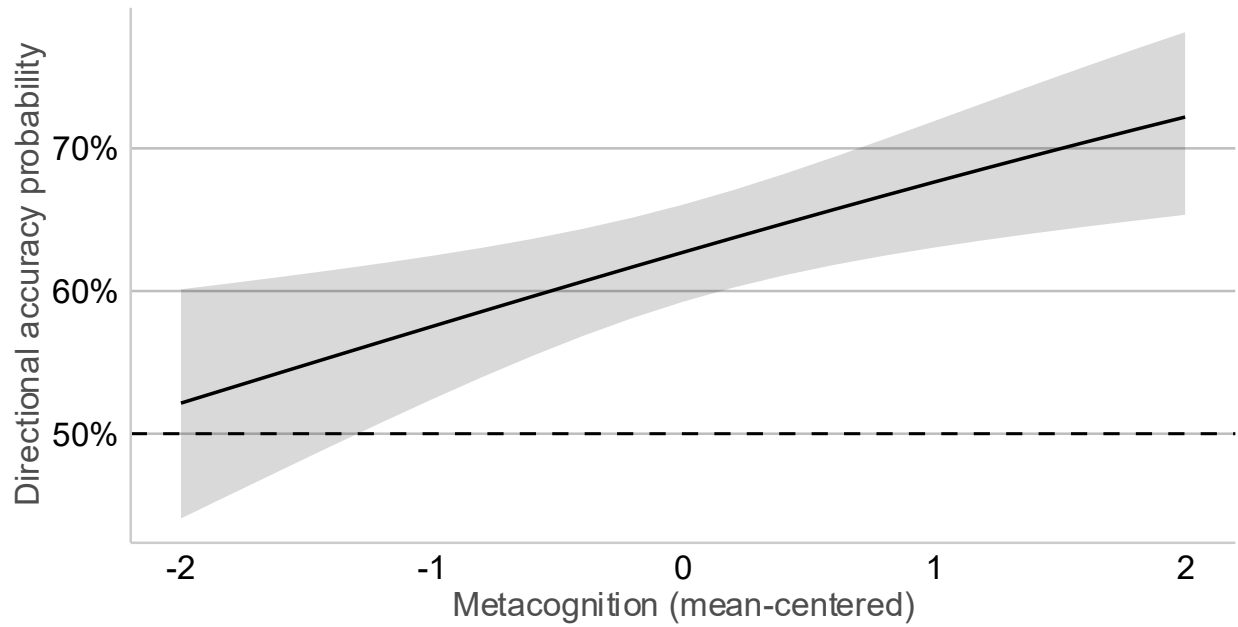
**Figure 3**

*Predicted probabilities of directional accuracy by confidence*



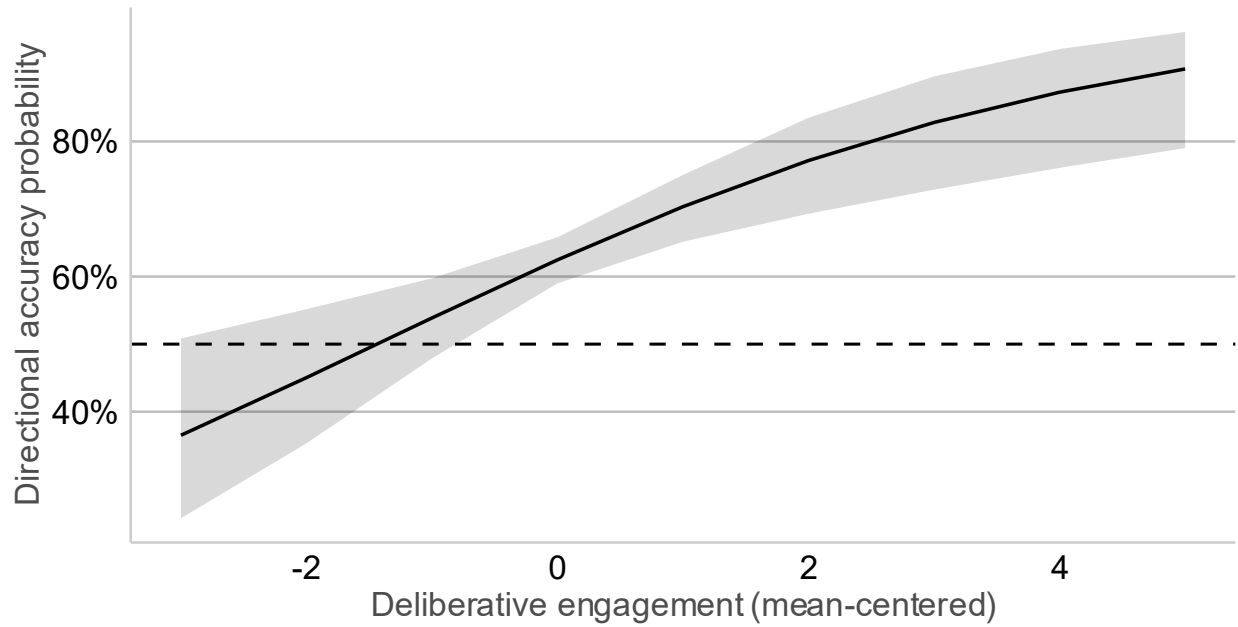
**Figure 4**

*Predicted probabilities of directional accuracy by metacognitive engagement*



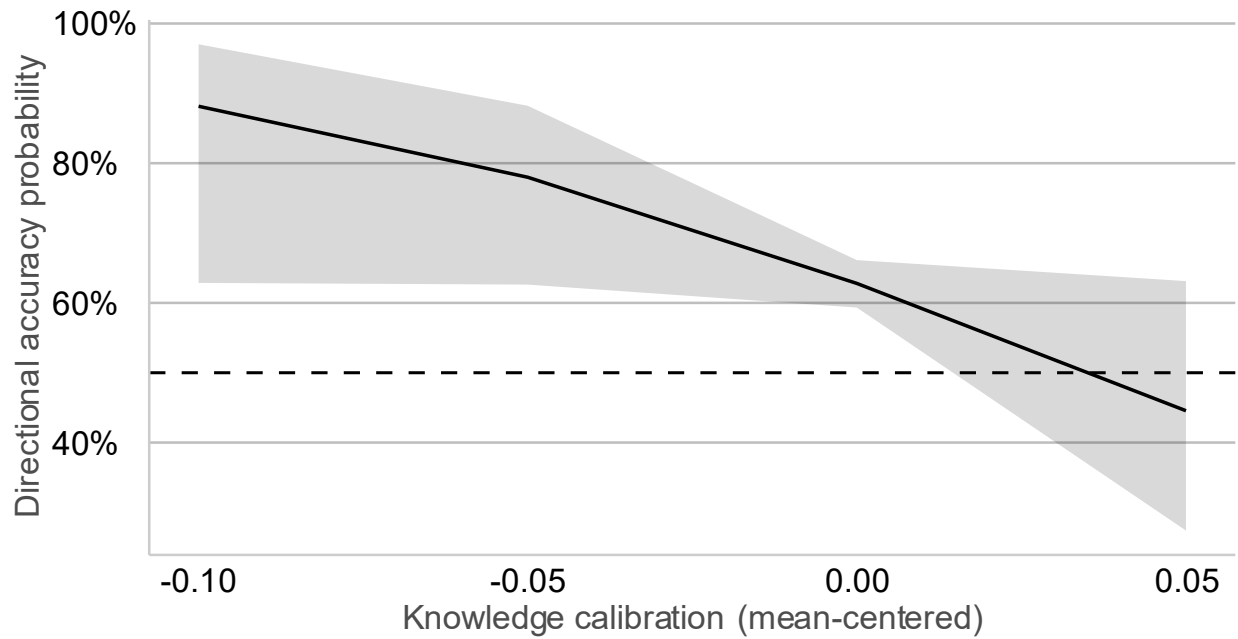
**Figure 5**

*Predicted probabilities of directional accuracy by deliberative task-engagement*



**Figure 6**

*Predicted probabilities of directional accuracy by knowledge calibration*

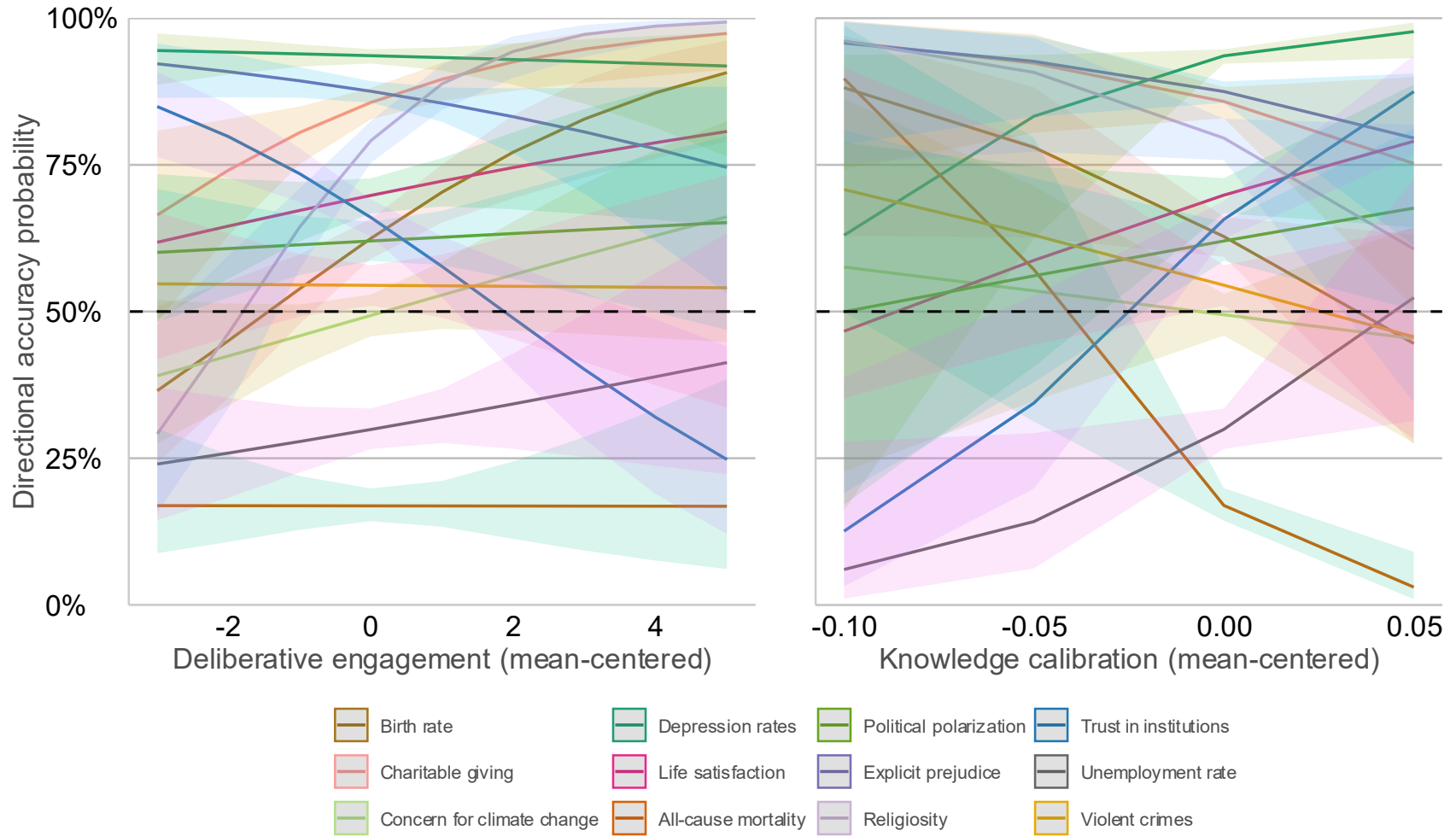


However, both deliberation and general knowledge calibration were also qualified by interactions with domain,  $\chi^2(11, N = 583) = 70.17$  and  $\chi^2(11, N = 583) = 42.81$ ,  $ps < .001$ , respectively (see Figure 7 for both interactions). Specifically, greater deliberative engagement was associated with a 35.4% increase in directional accuracy for birth rate estimates,  $SE = 9.6\%$ ,  $z = 3.67$ ,  $p < .001$ , and a 36.7% increase for estimates of charitable giving,  $SE = 12.7\%$ ,  $z = 2.90$ ,  $p = .004$ . Accuracy also highly improved for assessments of religiosity, showing a 74% increase,  $SE = 14.0\%$ ,  $z = 5.31$ ,  $p < .001$ . In contrast, deliberate engagement was also associated with a 35.5% decrease in accuracy for estimates of trust in institutions,  $SE = 8.9\%$ ,  $z = -3.97$ ,  $p < .001$ .

General knowledge calibration was associated with a 14.8% decrease in directional accuracy for birth rate estimates (though marginal),  $SE = 7.6\%$ ,  $z = -1.95$ ,  $p = .050$ , and a 37.5% decrease for mortality estimates,  $SE = 10.0\%$ ,  $z = -3.75$ ,  $p < .001$ . However, there was strong evidence suggesting that directional accuracy increased for estimates for trust in institutions by 25.9%,  $SE = 7.8\%$ ,  $z = 3.33$ ,  $p < .001$ , and marginal evidence for an increase of directional accuracy for unemployment estimates by 18.9%,  $SE = 9.7\%$ ,  $z = 1.95$ ,  $p = .051$ .

**Figure 7**

*Predicted probabilities of directional accuracy for domain by deliberative engagement and knowledge calibration*



## Discussion

The first year of the COVID-19 pandemic was a period of turbulence and uncertainty that profoundly affected societies worldwide. For many within the public, it was perceived as a clear demarcation between the world as it was and the world to come. Yet between integrating individuals' personal and local experiences and the broader collective narratives circulating at the time, how did the public actually perceive this societal shift? Was it seen as a harbinger of decline and crisis, or rather as an opening for social change and renewal of community? The present research suggests the former for the majority of domains. Many individuals expressed pessimism consistently across estimates of depression rates, all-cause mortality, political polarization, and unemployment despite these outcomes being less severe or even moving in the opposite direction. Only perceptions of explicit prejudice and trust in institutions suggested optimism consistently across estimates.

Perhaps related to this pessimism, the present research found that the public was largely unable to accurately estimate the direction of societal change during the first year of the COVID-19 pandemic, accuracy being at chance levels (49.5%). However, accuracy in the direction of societal change also varied as a function of domain. Specifically, estimates of depression rates, explicit prejudice, life satisfaction, religiosity, and charitable giving were often above chance. Estimates of depression rates and explicit prejudice were especially accurate, often near or above 80% in likelihood in estimating direction. On the other hand, participants were below chance on estimates of all-cause mortality, violent crimes, unemployment rates, and concern for climate change, and at chance for political polarization, trust in institutions, and birth rates.

Accuracy was also moderated by the timeframe of estimates, with longer estimation windows associated with reduced accuracy. This decline may stem from multiple factors. One possible explanation is the increased cognitive burden involved in recalling broader temporal trends, which may exceed the general public's typical capacity, especially during periods of uncertainty. However, another important consideration (not mutually exclusive with the prior) is that many individuals' lives during COVID-19 isolation were highly repetitive, where days and even months were indistinguishable from one another. Consequently, memories of any certain event may have been fuzzier given the lack of distinction of events and associated cues. Despite the overall difficulty of the task, one might still be expected to retain a general sense of how society had changed before forming expectations about the future. Yet even social scientists who routinely monitor societal trends often perform no better than chance when forecasting change (Hutcherson et al., 2023). In contrast, experts who directly engage with relevant historical data demonstrate significantly greater predictive accuracy (Grossmann et al., 2024b), highlighting the importance of a grounded and systematic understanding of the past into predicting the future. These findings suggest that the accuracy of societal change estimates may depend less on general expertise and more on which information is accessible, salient, or actively reinforced at the time of judgment. This raises a critical question: what made specific judgments of past societal change more accurate compared to others?

One possible general explanation is that the public's estimates are shaped by the salience and accessibility of information related to the domain in question. When a domain is

cognitively salient, either frequently encountered in personal experience or highly visible in media coverage, it becomes easier to recall and evaluate. Conversely, when domains are less prominent or overshadowed by competing information (such as the extreme negative media coverage during this time), individuals may rely on distorted or emotionally charged heuristics often biased towards a specific perception (similar to primal world beliefs; Clifton, 2020).

Crucially, domain-specific accuracy may not primarily reflect access to knowledge, but instead prominence of the domain at the time of estimation. In highly salient domains in everyday life, individuals may have had more opportunities to validate their impressions through lived experience or media exposure. However, when salience is externally amplified (e.g., by sensationalist coverage) or when personal and media experiences become misaligned, estimations may be skewed.

Such distortions may emerge from imbalances between local perceptions (e.g., one's direct experiences or those of close others) and collective narratives (e.g., public discourse, news media, social media). When one source dominates, individuals may (i) place excessive trust in that source, (ii) fail to question the reliability of their own or others' knowledge, or (iii) miss opportunities to moderate their views through comparison with alternate perspectives. In short, estimation errors may arise not from ignorance, but from asymmetries in how information is accessed, weighted, and integrated. This speculative framework may provide some explanation on why directional accuracy varied so much across domains. Participants were considerably more accurate when estimating changes in depression rates,

explicit prejudice, charitable giving, religiosity, and life satisfaction. Most notably, directional accuracy for depression and explicit prejudice consistently exceeded 80% across overall, six-month, and one-year estimates. These domains were not only substantively important during the pandemic, but also highly salient in media, discourse, and individuals' personal lives during the first pandemic year, providing multiple opportunities for comparative validation between information sources.

For instance, the rise in depression, was likely self-evident amid prolonged social isolation and frequently reinforced by widespread media coverage of mental health decline (e.g., Bernstein, 2022; Kwai & Peltier, 2021; Rogers, 2021). Similarly, increased attention to conspiracy theories such as the “lab-leak theory” may have shaped perceptions of rising prejudice, particularly against Asian communities (e.g., *BBC*, 2021; De Luce, 2025; Qiu, 2025). Other racially-motivated crimes were of great concern during this period as well, such as the murder of George Floyd and the consequent Black Lives Matter movement (Hill et al., 2020; Silverstein, 2021). Regardless of individuals' beliefs on the source of COVID-19, the public was likely to recognize the impact of spreading misinformation in their local and online communities.

Notably, despite widespread assumptions and substantial evidence of the pandemic's psychological toll, objective estimates indicated that life satisfaction actually increased during the April 2020–2021 period. This observation may seem counterintuitive given the concurrent rise in depression; however, these two constructs likely operate on different psychological timescales. Depression, requiring time to make a professional diagnosis, may

require reflection on longer-term patterns, while life satisfaction may be more sensitive to recent or situational experiences. Participants' underestimation of life satisfaction could also reflect distorted perceptions shaped by media narratives, such as those emphasizing the hardships of isolation policies or perceived inequities in financial support (Gramlich, 2021). But more importantly, individuals may also have misremembered or even forgotten their experiences during the beginning of COVID-19 lockdowns when perceptions of the future may have been the most bleak or uncertain. Their estimates may have been purely driven by their current experiences, rather than an accurate retrospection of the past.

Objective increases in life satisfaction may have been moderated by coping adaptations and resiliency as the pandemic went on. For example, participants' accurate estimates of increased religiosity may be related to widespread perceptions of religious practices mitigating psychological distress (Dutra & Rocha, 2021; Knight et al., 2021; Pirutinsky et al., 2020). Similarly, accurate estimates of increased charitable giving may reflect the existential urgency of the pandemic and publicized community support efforts (Fridman et al., 2022; Litofcenko et al., 2023).

By contrast to prior domains, participants performed especially poorly when estimating changes in all-cause mortality and unemployment. In these domains, accurate rates hovered near or below 20%. These domains, while objectively important, may have lacked the right kind of salience to accurately assess knowledge. In these cases, estimation errors appeared to follow from distorted impressions driven by high emotional intensity or misaligned cues.

For instance, participants frequently assumed that overall mortality increased, likely due to the frequency of COVID-19 death coverage, even though when accounting for all causes, mortality decreased. Many participants referenced COVID-related deaths in their explanations (unreported in this thesis), even though instructions explicitly asked them to consider all-cause mortality. This may also reflect base rate neglect, not recognizing how unprepared societies were during the start of COVID-19 and the adaptations created afterwards. Similarly, estimates for unemployment appeared to on focus on early-pandemic job losses, overlooking subsequent recovery through remote work and policy support (Redmond & Pabilonia, 2024). In this instance, base rates may instead have been too narrowly focused on, assuming rates have stayed the same since the beginning of the pandemic and ignoring adjustments made to adapt to it.

Yet compared to prior domains which were more consistently accurate or inaccurate, why might have estimates of political polarization, trust in institutions, and birth rates been closer to chance? Perceptions of political polarization and trust in institutions are likely to have been intertwined, both shaped by divisive narratives such as the “fake news” discourse (Caceres et al., 2022) in addition to lab-leak conspiracy theories and the political tensions surrounding social justice movements. Birth rates, by contrast, may have been harder to gauge because demographic shifts often unfold over much longer timeframes than six months or a year, yet perceived demarcation may have questioned the validity of continuing to rely on such trends in such a volatile context. Moreover, public expectations may have been split: some saw pandemic conditions (i.e., greater remote work and isolation) as conducive to

starting a family (Bailey et al., 2022), while others viewed it as too risky or burdensome due to health concerns and other subjective uncertainties (Rocca et al., 2022). These conflicting interpretations, compounded by the challenge of reconciling personal experiences with broader societal patterns, likely contributed to the mixed accuracy in these domains.

Overall, these findings suggest that accuracy in estimating societal change depends not only on knowledge or intelligence, but also on how people engage with available information: what is accessible, salient, and trusted at the time of judgment. Domains that were both prominent in public discourse and reinforced by personal experience may have been easier to assess accurately by critically engaging and integrating both sources of information. In contrast, domains where judgments relied too heavily on a single salient source, or where alternative perspectives were inaccessible (possibly due to personal biases), may be more vulnerable to error due to a lack of critical source comparison. Chance-level accuracy may be a product of deeper structural biases, such as political divisions splitting perspectives prior to aggregation (e.g., trust in institutions, political polarization) or lack of coherence between collective narratives and local experience to estimate a clear direction (e.g., birth rates). This then raises the question of who, if anyone, was able to overcome such informational distortions and asymmetries shaping public perception.

Results on individual differences suggest that metacognition and deliberative task-engagement may play key roles in estimating societal change. Specifically, participants who demonstrated higher levels of metacognitive and deliberate engagement were more likely to produce accurate directional estimates overall. These relationships persisted even when other

individual factors well-established in the forecasting literature, such as cognitive abilities, did not (e.g., Mellers, Stone, Atanasov, et al., 2015b). However, calibration of confidence and accuracy in participants' general knowledge, another form of metacognitive engagement, was found to be inversely related to directional accuracy of estimates. Given these mixed findings, it is possible that this may reflect a lack of similarity of metacognition between these two different domains. Indeed, although metacognition in the social context was focused on societal issues related to COVID-19 pandemic, it still lacked direct assessment of actual performance. On the other hand, metacognition in assessing general knowledge actively assessed participants' calibration but was unrelated to topics of societal change. This divergence suggests that metacognitive engagement may be context-specific, and that accuracy in societal estimation may depend more on reflective thinking about social and informational dynamics than on confidence regulation of general intelligence.

Greater deliberative engagement was linked to improved directional accuracy for several domains, including birth rates, charitable giving, and religiosity. However, it was associated with decreased accuracy in estimating trust in institutions. Similarly, general knowledge calibration showed mixed effects depending on the domain: it was related to reduced accuracy for birth rates and mortality, but increased accuracy for trust in institutions and unemployment estimates. These findings highlight that both deliberative engagement and knowledge calibration influence judgmental accuracy may not uniformly influence judgment accuracy but vary across different domains and societal indicators.

These nuanced findings suggest that more cognitive effort and reflection is not inherently beneficial across all domains of societal change. In some cases, greater deliberation or metacognition may amplify attention to noise, volatility, or competing narratives. For instance, domains like climate change or prejudice may have evoked conflicting cues, balancing long-term trends and shifting public attention. Under such conditions, deliberation may not clarify trends but instead increase confusion or susceptibility to recent, unrepresentative signals when simpler models may be more sufficient (Gigerenzer & Gaissmaier, 2011; Green & Armstrong, 2015; Grossmann et al., 2023; Hertwig & Herzog, 2009).

Confidence in estimates was also consistently associated with greater directional accuracy, similar to prior findings in predictions of societal change (Hutcherson et al., 2023). It is unclear how this relationship might occur, given that confidence may act as both signal and bias of accuracy in estimates. Moreover, the circular nature between confidence and accuracy of estimates further creates difficulties in identifying the role of confidence (i.e., “I am confident because I am accurate, and I am accurate because I was confident in my answer”). Future research would benefit in identifying the specific pathways that explain this relationship.

To further explore the speculative framework of the role of cognitive salience and active metacognitive processing, future research could analyze the explanations participants gave for their estimates, across and within domain. Specifically, topic modelling approaches, such as with Latent Dirichlet Allocation (Blei et al., 2003) or more recent neural network

methods like BERTopic (Grootendorst, 2022), could be employed to identify key themes (such as content of explanations, strategies employed, or evidence of metacognition) related to salience and metacognitive engagement. In terms of salience, future analyses could identify the extent to which participants' reasoning reflected (i) direct references to COVID-19 as an influence on estimates, (ii) temporal focus on current versus past events, and (iii) reliance on local experience versus broader trends when making estimates. One could then identify usage of each dimension as a proportion within the public, within each domain, and the extent each dimension might correlate or predict directional accuracy of estimates overall and within domain.

Similarly, topic modeling may also be used to identify evidence of active metacognitive processing in participants' responses. For example, analyses could examine whether participants: (iv) compared their beliefs to those of others or to objective data; (v) reflected on the limits of their own or others' knowledge; or (vi) adopted an outside perspective in their reasoning. These strategies align closely with key dimensions of metacognition outlined in the SWiS framework (Brienza et al., 2018). A self-assessment of participants' metacognitive strategies in estimating change could also complement these analyses by offering a more direct measure of metacognitive engagement. Such an assessment would help capture not just what participants concluded, but how they arrived at their estimates. Combining multiple lines of analysis from retrospective strategy reports, confidence-accuracy calibration of specific estimates, and qualitative coding of explanations may yield a more comprehensive and coherent picture of metacognition. This triangulated

approach would help clarify when and how metacognition contributes to more accurate societal judgments.

### **Limitations**

Before moving forward with such analyses, it is important to acknowledge the limitations of the present research prior to building upon its findings. First, this study was exploratory in nature and cannot support strong causal claims or definitive conclusions. As noted earlier, the domains examined were primarily selected based on data availability, which means they may not fully capture the entire range of societal areas most affected or significant during the first year of the COVID-19 pandemic. However, these domains were also chosen with reference to existing social science theories to provide a meaningful framework. In addition, the domains chosen aimed to cover a large range of potential concerns during this period, rather than be limited to those most apparent to the pandemic.

Additionally, since this research relied on existing data rather than collecting new measures, some variables serve as proxies rather than ideal measures. For example, general knowledge accuracy was used as a proxy for general intelligence, though it likely does not capture the construct as comprehensively as other established assessments like Raven's Matrices. However, it was selected for its accessibility and suitability for use with general population samples, which may not be familiar with psychological assessments. Similarly, deliberation-based task engagement was measured by time spent on estimates, which may not fully reflect deeper qualitative aspects of engagement. Unfortunately, given the focus on

the quantitative data from this study, more nuanced measures of deliberation could not be incorporated to existing analyses.

Metacognition was also not directly assessed during participants' estimation tasks. While the study focused on societal change, directly capturing participants' metacognitive reflections about their estimation process could provide richer insights. Furthermore, accuracy was assessed based on direction rather than magnitude of error. Although precision in estimates is linked to expert forecasting performance (Mellers, Stone, Atanasov, et al., 2015a), previous research (Hutcherson et al., 2023) cautions that magnitude-based accuracy may be disproportionately affected by overestimation, potentially inflating error measures. Nevertheless, given the importance of precision, future research could benefit from including both directional and magnitude-based metrics to offer a more comprehensive evaluation of accuracy.

## **Conclusion**

In conclusion, these findings highlight that accurate judgments of societal change depend not solely on preference for deliberation, access to knowledge or general intelligence, but on how individuals engage with and evaluate information in socially complex, uncertain environments. Domains that were more cognitively and socially salient tended to elicit more accurate estimates, but this accuracy also varied meaningfully based on individual differences in metacognitive engagement. Metacognition, in particular, emerged as a critical factor suggesting that reflection on the limits of one's knowledge and consideration of others' perspectives on it may help mitigate informational distortions. However,

metacognitive engagement is not universally beneficial as under certain conditions, it may amplify noise or bias when taken to the extreme. Together, these insights underscore the importance of developing more nuanced, context-sensitive models of how people form judgments about large-scale social trends — models that account not only for what people know, but how they think about what they know — as a foundation for preparing for the future.

## References

- Ashby, N. J. S., & Rakow, T. (2014). Forgetting the past: Individual differences in recency in subjective valuations from experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1153–1162.  
<https://doi.org/10.1037/a0036352>
- Bailey, M. J., Currie, J., & Schwandt, H. (2022). *The Covid-19 Baby Bump: The Unexpected Increase in U.S. Fertility Rates in Response to the Pandemic* (Working Paper No. 30569). National Bureau of Economic Research. <https://doi.org/10.3386/w30569>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48.  
<https://doi.org/10.18637/jss.v067.i01>
- BBC. (2021, May 27). *Covid origin: Why the Wuhan lab-leak theory is so disputed*.  
<https://www.bbc.com/news/world-asia-china-57268111>
- Berg, S. A., & Moss, J. H. (2022). Anchoring and Judgment Bias: Disregarding Under Uncertainty. *Psychological Reports*, *125*(5), 2688–2708.  
<https://doi.org/10.1177/003329412111016750>
- Bernstein, L. (2022, September 30). Suicides increased in 2021, especially among younger people. *The Washington Post*.  
<https://www.washingtonpost.com/health/2022/09/29/suicide-increase-2021/>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, *3*(null), 993–1022.

- Brienza, J. P., Kung, F. Y. H., Santos, H. C., Bobocel, D. R., & Grossmann, I. (2018). Wisdom, bias, and balance: Toward a process-sensitive measurement of wisdom-related cognition. *Journal of Personality and Social Psychology, 115*(6), 1093–1126. <https://doi.org/10.1037/pspp0000171>
- Caceres, M. M. F., Sosa, J. P., Lawrence, J. A., Sestacovschi, C., Tidd-Johnson, A., Rasool, M. H. U., Gadamidi, V. K., Ozair, S., Pandav, K., Cuevas-Lou, C., Parrish, M., Rodriguez, I., & Fernandez, J. P. (2022). The impact of misinformation on the COVID-19 pandemic. *AIMS Public Health, 9*(2), 262–277. <https://doi.org/10.3934/publichealth.2022018>
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1–22. <https://doi.org/10.1037/h0046743>
- Chang, W., Chen, E., Mellers, B., & Tetlock, P. (2016). Developing expert political judgment: The impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments. *Judgment and Decision Making, 11*(5), 509–526. <https://doi.org/10.1017/S1930297500004599>
- Clifton, J. D. W. (2020). Testing If Primal World Beliefs Reflect Experiences—Or at Least Some Experiences Identified ad hoc. *Frontiers in Psychology, 11*. <https://doi.org/10.3389/fpsyg.2020.01145>
- De Luce, D. (2025, January 25). *CIA shifts assessment on Covid origins, saying lab leak likely caused outbreak*. NBC News. <https://www.nbcnews.com/politics/politics->

- news/cia-shifts-assessment-covid-origins-saying-lab-leak-likely-caused-outbreak-rcna189284
- De Neys, W. (2025). Defining deliberation for dual-process models of reasoning. *Nature Reviews Psychology*, 4(8), 544–552. <https://doi.org/10.1038/s44159-025-00466-6>
- Dijkstra, K. A., van der Pligt, J., & van Kleef, G. A. (2013). Deliberation Versus Intuition: Decomposing the Role of Expertise in Judgment and Decision Making. *Journal of Behavioral Decision Making*, 26(3), 285–294. <https://doi.org/10.1002/bdm.1759>
- Dupuis, M., Studer, J., Wicki, M., Marmet, S., & Gmel, G. (2023). Was retrospective change measurement conducted with Covid-19 containment inconsistent? Comparing prospective and retrospective change measures using data from a national survey on substance use and addictive behaviors. *PLOS ONE*, 18(6), e0286597. <https://doi.org/10.1371/journal.pone.0286597>
- Dutra, C. C. D., & Rocha, H. S. (2021). Religious Support as a Contribution to Face the Effects of Social Isolation in Mental Health During the Pandemic of COVID-19. *Journal of Religion and Health*, 60(1), 99–111. <https://doi.org/10.1007/s10943-020-01140-2>
- Fiske, S. T. (Tufts), & Taylor, S. E. (2020). *Social Cognition: From brains to culture*. 1–672.
- Flora, D. B. (2020). Thinking about effect sizes: From the replication crisis to a cumulative psychological science. *Canadian Psychology / Psychologie Canadienne*, 61(4), 318–330. <https://doi.org/10.1037/cap0000218>

- Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology: General*, *141*(1), 124–133.  
<https://doi.org/10.1037/a0024006>
- Fridman, A., Gershon, R., & Gneezy, A. (2022). Increased generosity under COVID-19 threat. *Scientific Reports*, *12*(1), 4886. <https://doi.org/10.1038/s41598-022-08748-2>
- Friedman, J. A., Baker, J. D., Mellers, B. A., Tetlock, P. E., & Zeckhauser, R. (2018). The Value of Precision in Probability Assessment: Evidence from a Large-Scale Geopolitical Forecasting Tournament. *International Studies Quarterly*, *62*(2), 410–422. <https://doi.org/10.1093/isq/sqx078>
- Fujita, K., Eyal, T., Chaiken, S., Trope, Y., & Liberman, N. (2008). Influencing attitudes toward near and distant objects. *Journal of Experimental Social Psychology*, *44*(3), 562–572. <https://doi.org/10.1016/j.jesp.2007.10.005>
- Funder, D. C., & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. <https://doi.org/10.1177/2515245919847202>
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, *62*(Volume 62, 2011), 451–482. <https://doi.org/10.1146/annurev-psych-120709-145346>
- Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78.  
<https://doi.org/10.1016/j.paid.2016.06.069>

- Godlonton, S., Hernandez, M. A., & Murphy, M. (2018). Anchoring Bias in Recall Data: Evidence from Central America. *American Journal of Agricultural Economics*, *100*(2), 479–501. <https://doi.org/10.1093/ajae/aax080>
- Gramlich, C. D., Kim Parker and John. (2021, March 5). A Year of U.S. Public Opinion on the Coronavirus Pandemic. *Pew Research Center*.  
<https://www.pewresearch.org/social-trends/2021/03/05/a-year-of-u-s-public-opinion-on-the-coronavirus-pandemic/>
- Green, K. C., & Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research*, *68*(8), 1678–1685.  
<https://doi.org/10.1016/j.jbusres.2015.03.026>
- Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure* (No. arXiv:2203.05794). arXiv. <https://doi.org/10.48550/arXiv.2203.05794>
- Grossmann, I., & Johnson, S. G. B. (2025). Cultivating wisdom through metacognition: A new frontier in decision-making under radical uncertainty. *Journal of Applied Research in Memory and Cognition*. <https://doi.org/10.1037/mac0000235>
- Grossmann, I., Rotella, A., Hutcherson, C. A., Sharpinskyi, K., Varnum, M. E. W., Achter, S., Dhami, M. K., Guo, X. E., Kara-Yakoubian, M., Mandel, D. R., Raes, L., Tay, L., Vie, A., Wagner, L., Adamkovic, M., Arami, A., Arriaga, P., Bandara, K., Baník, G., ... The Forecasting Collaborative. (2023). Insights into the accuracy of social scientists' forecasts of societal change. *Nature Human Behaviour*, *7*(4), 484–501.  
<https://doi.org/10.1038/s41562-022-01517-1>

- Grossmann, I., Rudnev, M., Lopez, W., Atari, M., Barr, K., Bencherifa, A., Buckwalter, W., Clancy, R., Dahua, G., Dahua, N., Deguchi, Y., Dorfman, A., Fabiano, E., Guennoun, B., Halamova, J., Hashimoto, T., Homan, J., Kanovsky, M., & Machery, E. (2025). *Decision-making preferences for intuition, deliberation, friends, or crowds in independent and interdependent societies*. [https://doi.org/10.31234/osf.io/aud8f\\_v2](https://doi.org/10.31234/osf.io/aud8f_v2)
- Grossmann, I., Twardus, O., Varnum, M. E. W., Jayawickreme, E., & McLevey, J. (2022). Expert predictions of societal change: Insights from the world after COVID project. *American Psychologist*, 77(2), 276–290. <https://doi.org/10.1037/amp0000903>
- Grossmann, I., Varnum, M. E. W., Hutcherson, C. A., & Mandel, D. R. (2024a). When expert predictions fail. *Trends in Cognitive Sciences*, 28(2), 113–123. <https://doi.org/10.1016/j.tics.2023.10.005>
- Grossmann, I., Varnum, M. E. W., Hutcherson, C. A., & Mandel, D. R. (2024b). When expert predictions fail. *Trends in Cognitive Sciences*, 28(2), 113–123. <https://doi.org/10.1016/j.tics.2023.10.005>
- Grossmann, I., Weststrate, N. M., Ferrari, M., & Brienza, J. P. (2020). A Common Model Is Essential for a Cumulative Science of Wisdom. *Psychological Inquiry*, 31(2), 185–194. <https://doi.org/10.1080/1047840X.2020.1750920>
- Hák, T., Janoušková, S., & Moldan, B. (2016). Sustainable Development Goals: A need for relevant indicators. *Ecological Indicators*, 60, 565–573. <https://doi.org/10.1016/j.ecolind.2015.08.003>

- Hauser, O. P., & Norton, M. I. (2017). (Mis)perceptions of inequality. *Current Opinion in Psychology*, 18, 21–25. <https://doi.org/10.1016/j.copsyc.2017.07.024>
- Henry, B., Moffitt, T. E., Caspi, A., Langley, J., & Silva, P. A. (1994). On the “remembrance of things past”: A longitudinal evaluation of the retrospective method. *Psychological Assessment*, 6(2), 92–101. <https://doi.org/10.1037/1040-3590.6.2.92>
- Hertwig, R., & Herzog, S. M. (2009). Fast and Frugal Heuristics: Tools of Social Rationality. *Social Cognition*, 27(5), 661–698. <https://doi.org/10.1521/soco.2009.27.5.661>
- Hill, E., Tiefenthäler, A., Triebert, C., Jordan, D., Willis, H., & Stein, R. (2020, June 1). How George Floyd Was Killed in Police Custody. *The New York Times*.  
<https://www.nytimes.com/2020/05/31/us/george-floyd-investigation.html>
- Himmelstein, M., Atanasov, P., & Budescu, D. V. (2021). Forecasting forecaster accuracy: Contributions of past performance and individual differences. *Judgment and Decision Making*, 16(2), 323–362. <https://doi.org/10.1017/S1930297500008597>
- Hirst, W., Phelps, E. A., Buckner, R. L., Budson, A. E., Cuc, A., Gabrieli, J. D. E., Johnson, M. K., Lyle, K. B., Lustig, C., Mather, M., Meksin, R., Mitchell, K. J., Ochsner, K. N., Schacter, D. L., Simons, J. S., & Vaidya, C. J. (2009). Long-term memory for the terrorist attack of September 11: Flashbulb memories, event memories, and the factors that influence their retention. *Journal of Experimental Psychology. General*, 138(2), 161–176. <https://doi.org/10.1037/a0015527>
- Hutcherson, C. A., Sharpinskyi, K., Varnum, M. E. W., Rotella, A., Wormley, A. S., Tay, L., & Grossmann, I. (2023). On the accuracy, media representation, and public

- perception of psychological scientists' judgments of societal change. *American Psychologist*, 78(8), 968–981. <https://doi.org/10.1037/amp0001151>
- Inbar, Y., Cone, J., & Gilovich, T. (2010). People's intuitions about intuitive insight and intuitive choice. *Journal of Personality and Social Psychology*, 99(2), 232–247. <https://doi.org/10.1037/a0020215>
- Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, 64(6), 515–526. <https://doi.org/10.1037/a0016755>
- Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. *Science*, 306(5702), 1776–1780. <https://doi.org/10.1126/science.1103572>
- Kashima, Y., Bain, P., Haslam, N., Peters, K., Laham, S., Whelan, J., Bastian, B., Loughnan, S., Kaufmann, L., & Fernando, J. (2009). Folk theory of social change. *Asian Journal of Social Psychology*, 12(4), 227–246. <https://doi.org/10.1111/j.1467-839X.2009.01288.x>
- Kline, R. B. (2023). *Principles and Practice of Structural Equation Modeling*. Guilford Publications.
- Knight, D., Dudenkov, D. V., & Cheshire, W. P. (2021). Religion in the US during the time of a Pandemic: A Medical Perspective. *Journal of Religion and Health*, 60(5), 3177–3192. <https://doi.org/10.1007/s10943-021-01366-8>

- Kwai, I., & Peltier, E. (2021, February 14). ‘What’s the Point?’ Young People’s Despair Deepens as Covid-19 Crisis Drags On. *The New York Times*.  
<https://www.nytimes.com/2021/02/14/world/europe/youth-mental-health-covid.html>
- Land, K. C., & Michalos, A. C. (2018). Fifty Years After the Social Indicators Movement: Has the Promise Been Fulfilled? *Social Indicators Research*, *135*(3), 835–868.  
<https://doi.org/10.1007/s11205-017-1571-y>
- Lenth, R. (2025). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (Version 1.11.2-00001) [Computer software]. <https://rvlenth.github.io/emmeans/>
- Litofcenko, J., Meyer, M., Neumayr, M., & Pennerstorfer, A. (2023). Charitable Giving in Times of Covid-19: Do Crises Forward the Better or the Worse in Individuals? *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *34*(6), 1322–1334. <https://doi.org/10.1007/s11266-023-00558-y>
- Mandel, D. R., & Irwin, D. (2021). Tracking accuracy of strategic intelligence forecasts: Findings from a long-term Canadian study. *FUTURES & FORESIGHT SCIENCE*, *3*(3–4), e98. <https://doi.org/10.1002/ffo2.98>
- Maniscalco, B., Peters, M. A. K., & Lau, H. (2016). Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Attention, Perception, & Psychophysics*, *78*(3), 923–937. <https://doi.org/10.3758/s13414-016-1059-x>

- Mastroianni, A. M., & Dana, J. (2022). Widespread misperceptions of long-term attitude change. *Proceedings of the National Academy of Sciences*, *119*(11), e2107260119. <https://doi.org/10.1073/pnas.2107260119>
- Mastroianni, A. M., & Gilbert, D. T. (2023). The illusion of moral decline. *Nature*, *618*(7966), 782–789. <https://doi.org/10.1038/s41586-023-06137-x>
- Mellers, B., Stone, E., Atanasov, P., Rohrbaugh, N., Metz, S. E., Ungar, L., Bishop, M. M., Horowitz, M., Merkle, E., & Tetlock, P. (2015a). The psychology of intelligence analysis: Drivers of prediction accuracy in world politics. *Journal of Experimental Psychology: Applied*, *21*(1), 1–14. <https://doi.org/10.1037/xap0000040>
- Mellers, B., Stone, E., Atanasov, P., Rohrbaugh, N., Metz, S. E., Ungar, L., Bishop, M. M., Horowitz, M., Merkle, E., & Tetlock, P. (2015b). The psychology of intelligence analysis: Drivers of prediction accuracy in world politics. *Journal of Experimental Psychology: Applied*, *21*(1), 1–14. <https://doi.org/10.1037/xap0000040>
- Mellers, B., Stone, E., Murray, T., Minster, A., Rohrbaugh, N., Bishop, M., Chen, E., Baker, J., Hou, Y., Horowitz, M., Ungar, L., & Tetlock, P. (2015). Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions. *Perspectives on Psychological Science*, *10*(3), 267–281. <https://doi.org/10.1177/1745691615577794>
- Moore, D. A., Swift, S. A., Minster, A., Mellers, B., Ungar, L., Tetlock, P., Yang, H. H. J., & Tenney, E. R. (2017). Confidence Calibration in a Multiyear Geopolitical Forecasting Competition. *Management Science*, *63*(11), 3552–3565. <https://doi.org/10.1287/mnsc.2016.2525>

- O’Muirheartaigh, C., Gaskell, G., & Wright, D. B. (1995). Weighing Anchors: Verbal and Numeric Labels for Response Scales. *Journal of Official Statistics*, 11(3), 295–307.
- Pachur, T., & Spaar, M. (2015). Domain-specific preferences for intuition and deliberation in decision making. *Journal of Applied Research in Memory and Cognition*, 4(3), 303–311. <https://doi.org/10.1016/j.jarmac.2015.07.006>
- Pasek, J. (2018). *anesrake: ANES Raking Implementation* [Computer software].
- Pinker, S. (2018). *Enlightenment Now: The Case for Reason, Science, Humanism, and Progress*. Penguin UK.
- Pirutinsky, S., Cherniak, A. D., & Rosmarin, D. H. (2020). COVID-19, Mental Health, and Religious Coping Among American Orthodox Jews. *Journal of Religion and Health*, 59(5), 2288–2301. <https://doi.org/10.1007/s10943-020-01070-z>
- Posit Team. (2025). *RStudio: Integrated Development Environemtn for R* [Computer software]. Posit Software, PBC. <http://www.posit.co/>
- Protzko, J., & Schooler, J. W. (2019). Kids these days: Why the youth of today seem lacking. *Science Advances*, 5(10), eaav5916. <https://doi.org/10.1126/sciadv.aav5916>
- Qiu, J. (2025, June 25). The Covid ‘lab leak’ theory isn’t just a rightwing conspiracy – pretending that’s the case is bad for science. *The Guardian*. <https://www.theguardian.com/commentisfree/2025/jun/25/covid-lab-leak-theory-right-conspiracy-science>

- R Core Team. (2024). *R: A Language and Environment for Statistical Computing* (Version 4.4.0) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rahnev, D. (2025). A comprehensive assessment of current methods for measuring metacognition. *Nature Communications*, *16*(1), 701. <https://doi.org/10.1038/s41467-025-56117-0>
- Redmond, J. J., & Pabilonia, S. W. (2024). *The rise in remote work since the pandemic and its impact on productivity*. Bureau of Labor Statistics. <https://www.bls.gov/opub/btn/volume-13/remote-work-productivity.htm>
- Rocca, C. H., Parra, M., Muñoz, I., Foster, D. G., Boscardin, W. J., & Ralph, L. J. (2022). Comparison of Pregnancy Preferences Preceding vs Year 1 of the COVID-19 Pandemic. *JAMA Network Open*, *5*(7), e2220093. <https://doi.org/10.1001/jamanetworkopen.2022.20093>
- Rogers, K. (2021, January 4). *Mental health is one of the biggest pandemic issues we'll face in 2021*. CNN. <https://www.cnn.com/2021/01/04/health/mental-health-during-covid-19-2021-stress-wellness>
- Sacerdote, B., Sehgal, R., & Cook, M. (2020). *Why Is All COVID-19 News Bad News?* (Working Paper No. 28110). National Bureau of Economic Research. <https://doi.org/10.3386/w28110>

- Schwarz, N., Kahneman, D., & Xu, J. (2009). Global and Episodic Reports of Hedonic Experience. In *Calendar and Time Diary: Methods in life course research* (pp. 156–174). SAGE Publications, Inc. <https://doi.org/10.4135/9781412990295>
- Silverstein, J. (2021, June 4). *The global impact of George Floyd: How Black Lives Matter protests shaped movements around the world - CBS News*.  
<https://www.cbsnews.com/news/george-floyd-black-lives-matter-impact/>
- Tetlock, P. E. (1999). Theory-Driven Reasoning About Plausible Pasts and Probable Futures in World Politics: Are We Prisoners of Our Preconceptions? *American Journal of Political Science*, 43(2), 335–366. <https://doi.org/10.2307/2991798>
- Thomas, M., & Kyung, E. J. (2019). Slider Scale or Text Box: How Response Format Shapes Responses. *Journal of Consumer Research*, 45(6), 1274–1293.  
<https://doi.org/10.1093/jcr/ucy057>
- Trewin, D., & Hall, J. (2010, November 16). *Developing Societal Progress Indicators*. OECD. [https://www.oecd.org/en/publications/developing-societal-progress-indicators\\_5kghzxp6k7g0-en.html](https://www.oecd.org/en/publications/developing-societal-progress-indicators_5kghzxp6k7g0-en.html)
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- Wilkinson, L. (2011). ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H. *Biometrics*, 67(2), 678–679. <https://doi.org/10.1111/j.1541-0420.2011.01616.x>