

Sternal, Philipp

**Working Paper**

## Interpreting cynical beliefs about others

Working Paper, No. 465

**Provided in Cooperation with:**

Department of Economics, University of Zurich

*Suggested Citation:* Sternal, Philipp (2025) : Interpreting cynical beliefs about others, Working Paper, No. 465, University of Zurich, Department of Economics, Zurich, <https://doi.org/10.5167/uzh-276326>

This Version is available at:

<https://hdl.handle.net/10419/316346>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



**University of  
Zurich** <sup>UZH</sup>

University of Zurich  
Department of Economics

Working Paper Series  
ISSN 1664-7041 (print)  
ISSN 1664-705X (online)

---

Working Paper No. 465

# **Interpreting Cynical Beliefs About Others**

Philipp Sternal

March 2025

---

# Interpreting Cynical Beliefs About Others

Philipp Sternal\*

March 14, 2025

*A growing number of studies suggest that individuals are cynical about others' behavior. But these findings often rely on self-reported rather than actual behavior as benchmark. A well-documented limitation of self-reports is their tendency to overstate good behavior. I introduce a simple, portable test to assess the extent to which inattention to others' potential misreporting drives apparently cynical beliefs about stated behavior. Drawing people's attention to the possibility of misreporting in self-reports increases beliefs about others' stated desirable climate and health behaviors by an average of 0.33 standard deviations, substantially reducing apparent cynicism. (JEL C90, D83, D91)*

In recent years, there has been growing interest in understanding whether the beliefs we hold about others are well calibrated (for a review, see Bursztyn and Yang, 2022). A regular observation is that people hold cynical beliefs about others, i.e., they underestimate others' good behavior, for example in the domains of climate action (Andre et al., 2024b), gender equality (Bursztyn, González and Yanagizawa-Drott, 2020), health-promoting behavior (e.g., Liu and Niederdeppe, 2020), and pro-sociality (e.g., Epley et al., 2022). This observation suggests that correcting misperceptions about others could be a valuable policy tool to promote good behavior by leveraging individuals' tendency for conditional cooperation (e.g., Bursztyn, González and Yanagizawa-Drott, 2020; Andre et al., 2024b).

In this line of research, *evaluators'* misperceptions of *targets'* behavior are often measured for stated rather than actual behavior. Specifically, evaluators' beliefs about targets' self-reported behavior are often compared to targets' self-reported behavior.<sup>1</sup> A potential concern with this approach is that the targets' self-reported behavior can be prone to socially desirable responding and other response biases<sup>2</sup> (Bertrand and Mullainathan, 2001), which could in principle contribute to misperceptions (Bursztyn and Yang, 2022). With self-reported behavior as the benchmark, it is not clear whether evaluators misperceive (i) targets' actual behavior/preferences (*real misperception*) or (ii) targets' response bias (*confounding factor*). Isolating the real misperception would therefore require access to targets' actual behavior, which can often be costly or even impossible to observe.<sup>3</sup>

\*University of Zurich (email: philipp.sternal@econ.uzh.ch). I want to thank my advisors, Roberto Weber and Michel Maréchal, for their insightful feedback. I also want to thank Sandro Ambühl, Björn Bartling, Ernst Fehr, Krishna Srinivasan as well as seminar participants at the University of Zurich and ETH Zurich for their helpful comments. Funding from the ZGSE Director's Grant is gratefully acknowledged. This study has been pre-registered on AsPredicted (#182030; <https://aspredicted.org/xkmq-q4hh.pdf>) and has received IRB approval from the Human Subjects Committee of the Faculty of Economics, Business Administration and Information Technology at the University of Zurich (#2024-061).

<sup>1</sup>There are two types of exceptions that come to mind. First, donation tasks with financial stakes allow for the computation of misperceptions about actual behavior (e.g., Frey and Meier, 2004; Drouvelis and Marx, 2022; Andre et al., 2024b). Second, beliefs about fellow players' actual contributions are sometimes measured in public-good games (e.g., Croson, 2007; Neugebauer et al., 2009; Fischbacher and Gächter, 2010; Gächter and Renner, 2010). These studies find both cynical and optimistic beliefs about others. For example, using students from the University of Zurich, Frey and Meier (2004) document cynical beliefs about fellow students' contributions to a student fund while Fischbacher and Gächter (2010) find that, for a public-good game, "*contributions are lower than beliefs in almost all instances*" (p. 545).

<sup>2</sup>For the purposes of this paper, the distinction between the different motives for misreporting will not be a primary concern.

<sup>3</sup>In some cases, targets' behavior can be observed on the aggregate level. However, this data often provides only limited insights into the share of targets engaging in a given behavior (extensive margin).

To illustrate this interpretational challenge, consider the following example: Suppose that 40% of survey respondents (targets) state that they support some policy X, which is perceived as desirable. However, other respondents (evaluators) believe the level of stated support is only 30%. Based on this comparison for stated support (30% vs. 40%), one might conclude that evaluators are cynical about targets' actual support for X. However, an alternative interpretation is that evaluators accurately perceive the actual support for X at 30% but rather fail to account for the 10% of targets merely claiming to support X (e.g., due to social desirability bias or self-deception). Consequently, the interpretation of cynical beliefs about the stated support for X will be inherently ambiguous.

This paper addresses an important question arising from this ambiguity in interpretation: can we take cynical beliefs about targets' self-reported behaviors at face value as evidence for real misperception, without having access to the actual behavior? To tackle this question, I propose a simple and portable method to test whether estimates of targets' behavior already take into account the possibility of response bias, and this test does not require any knowledge of the targets' actual behavior. Specifically, the treatment provided by my method guides evaluators' attention to their own considerations of targets' potential misreporting. The idea behind this approach is straightforward. If evaluators' beliefs about targets' behavior remain unchanged, we can conclude that they already fully considered the possibility of targets misreporting and the treatment thus has no effect. While this observation does not rule out that evaluators underestimate response bias, it makes response bias as a major confounding factor less plausible: evaluators would have to fail to account for a sizeable response bias despite being explicitly prompted to consider its potential existence. Conversely, if evaluators' beliefs about targets change due to this shift in attention, and particularly in the direction of likely response bias, it suggests that evaluators initially failed to fully attend to their considerations of self-report truthfulness and that the prompt leads them to engage in a correction for this bias. While this observation does not guarantee that evaluators underestimate response bias, the initial failure to fully consider the possibility of response bias makes response bias as an important confounding factor more plausible.

I implement this test for survey questions on climate action and health behavior, using an online experiment with a broadly representative U.S. sample. In my experiment, subjects occupy the role of evaluators and are asked to make financially-incentivized guesses about the self-reported behavior of targets (i.e., other subjects). Behaviors are binary (Yes/No) and differ in perceived desirability. My primary experimental variation randomly assigns some evaluators to an additional task prior to belief elicitation. The purpose of this additional task is to draw evaluators' attention to their own considerations of targets' potential misreporting, henceforth the *attention manipulation*. The attention manipulation consists of two parts. First, evaluators must state for each behavior whether they believe that the number of targets' "Yes" responses is lower/the same/higher than the number of targets actually doing this activity. Second, evaluators must write a short essay on why they think targets might or might not tell the truth in online surveys. A simple between-subject comparison of beliefs about targets' self-reported behavior then allows me to causally identify the effect, if any, of this attention manipulation.

At the outset, I consider two opposing hypotheses. Under the null, beliefs about targets will remain unchanged by the attention manipulation. This hypothesis is consistent with the general idea that financial incentives for accuracy already ensure that evaluators use all available information in attempting to provide accurate guesses of targets' self-reported behavior, including thorough consideration of any potential response bias (e.g., Andre et al., 2024b). Under the alternative hypothesis, the attention ma-

nipulation will shift beliefs about targets in the direction of perceived response bias. This hypothesis is consistent with the idea that evaluators do not fully adjust for targets' potential misreporting, which could be driven by a variety of potential mechanisms. While I am agnostic about the precise mechanism(s) at work, candidates may include: (i) an adaptive truth-default, which means that evaluators expect targets to report honestly unless this assumption is overturned by a sufficiently strong trigger (e.g., Levine, 2014), (ii) dual process theory, à la Kahneman (2012), in which the fast System 1 ignores social image concerns faced by targets while the slow System 2 takes image concerns into account (for a discussion, see also Braghieri, 2024), or (iii) limited strategic sophistication more generally.

In the main analysis of this paper, I find that the attention manipulation indeed shifts beliefs about targets' self-reported behavior in the direction of perceived response bias. On average, beliefs about targets' desirable behavior increase by one third of a standard deviation, while beliefs about the one undesirable behavior included in the study decrease at least directionally, though not significantly. Importantly, I do not observe any effect of the attention manipulation in a placebo test with neutral behavior involving no desirable responses. Taken together, this suggests that beliefs about targets do not fully reflect perceived desirable misreporting by targets by default and that guiding evaluators' attention to think about potential misreporting can adjust for this bias in beliefs.

As a robustness check, I compare the effect of the attention manipulation to the effect of the evaluators themselves taking the same survey as targets prior to belief elicitation. This comparison is relevant because jointly eliciting self-reports and beliefs is commonly done in economics (e.g., Bursztyn, González and Yanagizawa-Drott, 2020; Bursztyn et al., 2024; Andre et al., 2024*b,a*) and could generate at least some introspection about targets' misreporting (Brownback, Burke and Gagnon-Bartsch, 2024). Therefore, I randomly assign some evaluators to a third condition in which they have to answer the survey questions themselves prior to belief elicitation. While including a preceding survey indeed shifts beliefs about targets in a manner consistent with increased introspection and awareness of response bias, the effect of the attention manipulation is still about one fifth of a standard deviation larger for desirable behavior. This suggests that the joint elicitation of self-reports and beliefs can mitigate, but not fully, resolve the issue of incomplete attention to targets' potential misreporting.

In exploratory analysis, I examine whether the measured magnitude of cynical beliefs is sensitive to the experimental condition used to elicit evaluators' beliefs. I find that the attention manipulation typically reduces cynical beliefs for desirable behavior by at least half compared to when no task precedes belief elicitation. A preceding survey often achieves a similar, albeit smaller, reduction in cynical beliefs compared to the attention manipulation. However, the difference between these two preceding tasks is largely inconsequential for measuring cynical beliefs. A notable exception occurs in the climate domain: here, the effect of the attention manipulation compared to the preceding survey-taking is not only largest, but may also be pivotal in the sense that there is no longer evidence for cynical beliefs.

Finally, I consider whether my results could be explained by an experimenter demand effect or another systematic change induced by my attention manipulation that operates through a channel other than attention to perceived misreporting. However, the precautionary measures featured by my experiment (i.e., financial incentives for belief accuracy, neutrally-worded instructions), along with a comparison to the typical size of demand effects from De Quidt, Haushofer and Roth (2018) and a placebo test with neutral behavior, speak against such alternative explanations as the sole driver of my results.

## I. Related Literature and Contribution

While there is a long-standing, extensive literature on the mitigation of social desirability bias (for a review, see Nederhof, 1985), this paper is only tangentially related to this strand of literature. Rather than trying to reduce bias in self-reports, my study contributes by focusing on how self-reports are used to compute misperception and how evaluators' anticipation of social desirability bias influences estimates of misperceptions.

My paper is thus more closely related to existing discussions about the use of self-reported benchmarks in misperception studies. While the absence of actual behavior is discussed as a potential challenge in studies on misperception, these studies, by and large, argue that the use of self-reports is not a relevant confounder (for a review, see Bursztyn and Yang, 2022). The two most common ways in which this concern is addressed can be summarized as follows. First, it is argued that self-reports are adequate proxies of actual behavior because there is no evidence for a strong response bias in the specific setting under investigation. This claim is often backed up by showing a correlation between self-reports and some other financially-incentivized behavior in the experiment (e.g., Andre et al., 2024b) or by employing techniques to mitigate social desirability bias in self-reports, such as list experiments (e.g., Cantoni et al., 2019; Bursztyn, González and Yanagizawa-Drott, 2020).<sup>4</sup> Second, it is argued that financial incentives should ensure that evaluators account for any remaining social desirability bias (e.g., Andre et al., 2024b).

In this paper, I focus on a central yet underappreciated angle in the existing debate: without observing actual behavior, one cannot test whether response bias is absent or at least anticipated by evaluators in a specific setting. For example, techniques to foster truthful answers might reduce response bias, but they do not necessarily eliminate it altogether (see, for example, Rosenfeld, Imai and Shapiro, 2016). Similarly, anticipation of targets' response bias may be limited despite financial incentives (Brownback, Burke and Gagnon-Bartsch, 2024; Aycinena, Bogliacino and Kimbrough, 2024; Braghieri, 2024; Ho and Huang, 2024).

I address this gap in the existing misperception literature by providing a simple belief-based test to identify the magnitude of potential concerns associated with the use of self-reported benchmarks in virtually any setting. The key difference with the existing work is that my test treats the absence of actual behavior as an empirical constraint and instead focuses on whether beliefs about targets are complete with respect to evaluators' own considerations of potential response bias. The attention manipulation I use is straightforward to use in any survey-based setting and provides an easily interpretable measure of the degree to which limited attention about potential response bias produces beliefs that may be mistakenly attributed to real misperceptions.

This paper is thus also closely related to recent experimental work on the anticipation of social desirability bias (Brownback, Burke and Gagnon-Bartsch, 2024; Aycinena, Bogliacino and Kimbrough, 2024; Braghieri, 2024; Ho and Huang, 2024). Reassuringly, my results are consistent with the general insight emerging from this literature: evaluators' anticipation of targets' response bias is limited by default, but improves as targets' image concerns become more salient to evaluators. For example, Brownback, Burke and Gagnon-Bartsch (2024) find that evaluators insufficiently correct for social desirability bias in targets' hypothetical donation decisions when using them to predict targets' real donation

<sup>4</sup>Recently, Bursztyn et al. (2024) proposed to measure social desirability bias through beliefs. But their approach assumes that evaluators are able to account for this bias after taking the survey themselves.

decisions. However, there is suggestive evidence that previous participation in the hypothetical donation task can improve inference, presumably because of heightened introspection about image concerns. Similarly, Braghieri (2024) finds that evaluators underestimate the effect of image concerns when predicting targets’ private and public support for political statements if those image concerns are not made salient. In his study, image concerns are made more salient by revealing to evaluators that there are two conditions in which self-reports are elicited.<sup>5</sup> In addition, Ho and Huang (2024) show that evaluators insufficiently account for targets’ silence in a public debate when inferring private support for controversial statements. But inference improves if the number of targets who choose to remain silent is made more salient to evaluators.

However, these studies primarily examine whether evaluators’ beliefs about targets’ self-reports track in a one-to-one fashion how these self-reports react to experimental variations (e.g., exogenous changes in the level of social desirability bias). Their focus is thus distinct from the setting of large-scale (online) surveys discussed in this paper. More specifically, evaluators’ limited ability to anticipate how targets change their self-report as a result of variations in the experimental environment does not allow us to understand whether evaluators are fully attentive to targets’ potential response bias at baseline (especially because evaluators may be less familiar with experimental variations than with conventional surveys).<sup>6</sup> This study thus contributes to this evolving literature by providing a simple and portable test to explore the practical consequences of evaluators’ limited anticipation of targets’ response bias for virtually any survey setting.

## II. Experimental Design

In this experiment, I measure the beliefs that subjects (*evaluators*) hold about the self-reported behavior of other subjects (*targets*). Specifically, evaluators make financially-incentivized guesses about the share of targets stating that they engage in certain behaviors that vary in perceived desirability. In the experiment’s baseline condition, ONLYBELIEF, evaluators only face this belief-elicitation task. To assess the extent to which evaluators already pay attention to targets’ potential misreporting at baseline, I exogenously vary which task, if any, precedes the belief elicitation common to all experimental conditions in two pre-registered treatments. My first and main treatment, ATTENTION+BELIEF, randomly assigns some evaluators to an additional task aimed at shifting their attention towards their considerations about targets’ potential misreporting prior to belief elicitation. My second treatment, SURVEY+BELIEF, randomly assigns some evaluators to a third condition in which they answer the survey questions themselves prior to belief elicitation to foster introspection about potential misreporting. Survey answers of subjects in this condition also serve as the benchmark for targets’ self-reported behavior. Next, I will describe the different tasks and the data collection in more detail. An illustration of the experimental design and the full instructions are available in Supplemental Appendix D.

<sup>5</sup>This is not the focus of his paper. Accordingly, there is only one statement (illegal immigration) for which this pattern is documented. Braghieri (2024) interprets this finding as consistent with “fast” and “slow” thinking à la Kahneman (2012).

<sup>6</sup>The approach of having two different benchmarks may also not be well-suited for large-scale surveys as it tends to be very costly (adding another benchmark also requires the elicitation of additional beliefs) and manipulations of social desirability are difficult to implement in a convincing and cost-effective fashion outside the laboratory.

Table 1: Behaviors of Interest

Topic	Question	Desirability	Adapted from
Climate	Would you be willing to contribute 1% of your household income every month to fight global warming? This would mean that you would contribute \$1 for every \$100 of this income.	Positive	Andre et al. (2024a)
Exercise	Do you typically participate in 150 minutes (2.5 hours) or more of aerobic physical activity (like walking, running, or bicycling) per week?	Positive	Liu and Niederdeppe (2020)
Vegetables	Do you typically consume vegetables at least once per day?	Positive	Liu and Niederdeppe (2020)
Coffee	Do you typically drink coffee every day?	Neutral	—
DST	Would you support making Daylight Saving Time (DST) permanent in the US to avoid changing clocks twice a year?	Neutral	Ho and Huang (2024)
Drinks	One drink is equivalent to a 12-ounce beer, a 5-ounce glass of wine, or a drink with one shot of liquor. In the past 30 days, have you consumed five or more drinks on any single occasion if you are male, or four or more drinks if you are female?	Negative	Liu and Niederdeppe (2020)

*Notes:* This table presents the behaviors of interest used throughout this study. The desirability reported in the third column refers to answering a given question with “Yes”. For each topic, the question and desirability classification were pre-registered.

### A. Tasks

This experiment features three different tasks, to be discussed in turn: belief elicitation, attention manipulation, and survey. Importantly, all three tasks focus on the same binary questions concerning different behaviors of interest presented in Table 1. Out of these six questions, three are about desirable behavior (*Climate*, *Exercise*, *Vegetables*), two are about neutral behavior (*Coffee*, *DST*), and one is about undesirable behavior (*Drinks*).<sup>7</sup> The question for *Climate* is taken from Andre et al. (2024a); items similar to *Exercise*, *Vegetables* and *Drinks* have been studied by Liu and Niederdeppe (2020).<sup>8</sup> Crucially, both studies find evidence for cynical beliefs. The question for *DST* is inspired by Ho and Huang (2024). To minimize order effects, behaviors of interest are displayed in random order for each subject and task.

*Belief Elicitation.*—In the belief-elicitation task, evaluators must guess the share of targets’ “Yes” responses for each survey question.<sup>9</sup> Accuracy is financially incentivized by a binarized scoring rule (Hossain and Okui, 2013), implemented as in Wilson and Vespa (2016). Explanations of the incentive mechanism are adapted, with some minor modifications, from Danz, Vesterlund and Wilson (2022). Evaluators must also answer two comprehension questions, similar to the ones used in Brownback, Burke and Gagnon-Bartsch (2024), to demonstrate their understanding of the task and the incentives. The aim of this belief elicitation is to measure beliefs about targets’ self-reported behavior, which is the (pre-registered) main outcome.

*Attention Manipulation.*—As part of the attention manipulation, evaluators must indicate for each of the six survey questions whether they think that the number of targets’ “Yes” responses is lower/the same/higher than the number of targets actually doing this activity. The order of the three answer options

<sup>7</sup>This classification of questions was pre-registered, but it can also be validated in-sample (see Section III).

<sup>8</sup>Note that Liu and Niederdeppe (2020) do not report the precise wording of their questions in their paper, but their items generally relate to the CDC’s Behavioral Risk Factor Surveillance System (BRFSS). I thus try to construct similarly-worded binary questions based on their article and instructions from the BRFSS.

<sup>9</sup>As in Brownback, Burke and Gagnon-Bartsch (2024), I frame the share as the number of participants out of 100.

is randomized on the subject level, with “the same” always being displayed as the second option. Finally, evaluators must write a short essay for the following prompt:

*Remember that, when people are asked about their behavior in surveys, they might or might not tell the truth. The extent to which they tell the truth might vary depending on the nature of the topic.*

*Based on the 6 questions you just saw, can you think of reasons why fellow U.S. participants might or might not tell the truth in a broadly representative and anonymous online survey? Please briefly describe such potential reasons (1-3 sentences).*

Importantly, the experimental instructions do not specify that there is misreporting by targets, let alone in which direction, so as to minimize demand effects. The main aim of this attention manipulation is to draw evaluators’ attention to targets’ potential misreporting. Evaluators’ answers can also be used to validate the pre-registered classification of behaviors in-sample.

*Survey.*—The survey consists of the six Yes-No questions from Table 1. The aim of this survey is two-fold. First, survey-taking conceivably fosters evaluators’ introspection about potential misreporting. Second, the answers collected on the survey can serve as benchmarks for targets’ self-reported behavior.

### B. Implementation

Given the ubiquity of online surveys to gauge public opinion, I conducted my experiment online using Qualtrics. For the experiment, I recruited a sample of U.S. residents on Prolific in July 2024 using the platform’s representative sampling option in terms of age, gender, and political affiliation. Each subject had to pass an attention check before being allowed to take part in the study.<sup>10</sup> After the attention check(s), subjects were randomly allocated to one of the three conditions. The final sample contains 747 subjects<sup>11</sup>: 247 in ONLYBELIEF, 249 in ATTENTION+BELIEF, and 251 in SURVEY+BELIEF. The mean duration of the experiment was 7.84 minutes in ONLYBELIEF, 11.99 minutes in ATTENTION+BELIEF, and 9.27 minutes in SURVEY+BELIEF.<sup>12</sup> Subjects received a fixed payment of \$2.40 for participating. Additionally, subjects could win a \$5 bonus based on a randomly selected guess they made in the belief-elicitation task. At the conclusion of the experiment, one in five subjects was randomly selected to have their bonus payments implemented. Out of the 150 eligible subjects, 106 subjects won this bonus.

## III. Data and Empirical Strategy

*Sample.*—Subjects are balanced on observable co-variates across conditions, and the overall sample is broadly representative of the U.S. population (see Supplemental Appendix A for details).<sup>13</sup>

<sup>10</sup>Prolific requires that subjects fail at least two attention checks to be excluded from studies of five minutes or longer. Subjects thus saw up to two attention checks at the beginning of the study and were allowed to proceed as soon as they had passed one of them. Only seven subjects failed the first, but subsequently passed the second attention check. I employed a commonly used attention check with some minor rephrasing as in Andre et al. (2024b).

<sup>11</sup>I pre-registered a total sample of 750 U.S. residents. This minor difference in sample size is due to some subjects who did not return their submission so that their places could not be filled in a timely manner.

<sup>12</sup>To prevent subjects from speeding through the experiment, I implemented minimum durations for some screens. More details can be found in Supplemental Appendix D.2.

<sup>13</sup>I use IPUMS data (Ruggles et al., 2024) from the American Community Survey 2023 to compute characteristics for the U.S. population. Political affiliation comes from Gallup (<https://news.gallup.com/poll/15370/party-affiliation.aspx>) as of 1 July 2024. Note that, as can be expected, the Prolific sample has fewer elderly people, fewer top earners, and a higher level of education. I will use raking weights to compute misperceptions as part of the exploratory analysis in Section IV.C, and this calibration will not change the overall picture (Supplemental Appendix C).

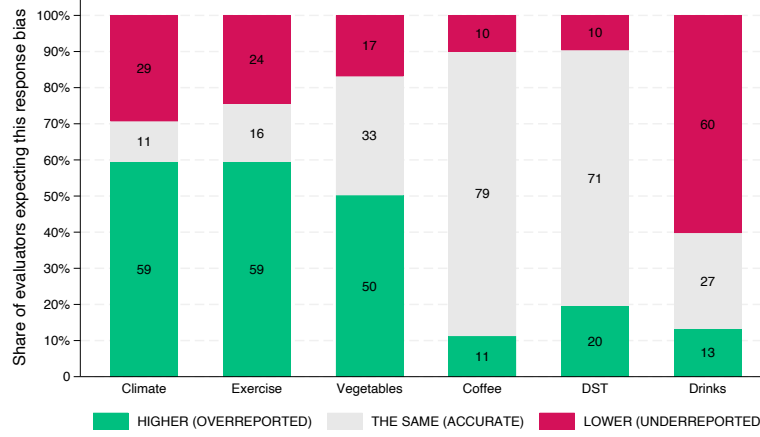


Figure 1: Perceived Response Bias

*Notes:* This figure presents the share of evaluators in the attention-manipulation task who think that the number of targets’ “Yes” responses is higher, the same, or lower than the number of targets actually doing this activity. This corresponds to perceived overreporting, accurate reporting, or underreporting for a given behavior, respectively.

*Perceived Response Bias.*—Figure 1 shows evaluators’ perceptions of misreporting by targets, i.e., the expected discrepancy between targets’ stated and actual behavior. Crucially, for each question, a majority of evaluators answers in line with the question’s pre-registered desirability classification.

*Empirical Strategy.*—As pre-registered, the main regression equation writes:

$$b_{i,j} = \sum_{j=1}^6 \gamma_j + \delta^+ \cdot T_i \cdot 1_{j \in \text{desirable}} + \delta^0 \cdot T_i \cdot 1_{j \in \text{neutral}} + \delta^- \cdot T_i \cdot 1_{j \in \text{undesirable}} + \varepsilon_{i,j} \quad (1)$$

where  $b_{i,j}$  is the belief that evaluator  $i$  holds about the targets’ stated prevalence for behavior  $j$  (0-100);  $\gamma_j$  are fixed effects for each behavior  $j$ ;  $T_i$  is equal to 1 if evaluator  $i$  is treated and equal to 0 otherwise. Errors are clustered on the individual level. For the main analysis, treatment will correspond to being assigned to ATTENTION+BELIEF, with the control condition corresponding to being assigned to ONLYBELIEF. With misperception results often focusing on desirable behavior, the main coefficient of interest is  $\delta^+$ . Absent any general shift in beliefs induced by the treatment, my simple belief-based test writes as  $H_0 : \delta^+ = 0$  vs  $H_1 : \delta^+ > 0$ .<sup>14</sup>

## IV. Results

I first test whether the attention manipulation moves evaluators’ beliefs about targets’ self-reported behavior in the direction of perceived response bias relative to the stand-alone belief elicitation. Next, as a robustness check, I compare the effect of the attention manipulation to the effect of the evaluators themselves taking the survey prior to belief elicitation. In exploratory analysis, I then examine whether the measured magnitude of cynical beliefs is sensitive to the experimental condition used to elicit evaluators’ beliefs. Finally, I consider demand effects and other systematic changes as a potential driver of my results.

<sup>14</sup>In what follows, I will report the more conservative two-sided p-values. All regression tables are available in Supplemental Appendix B.

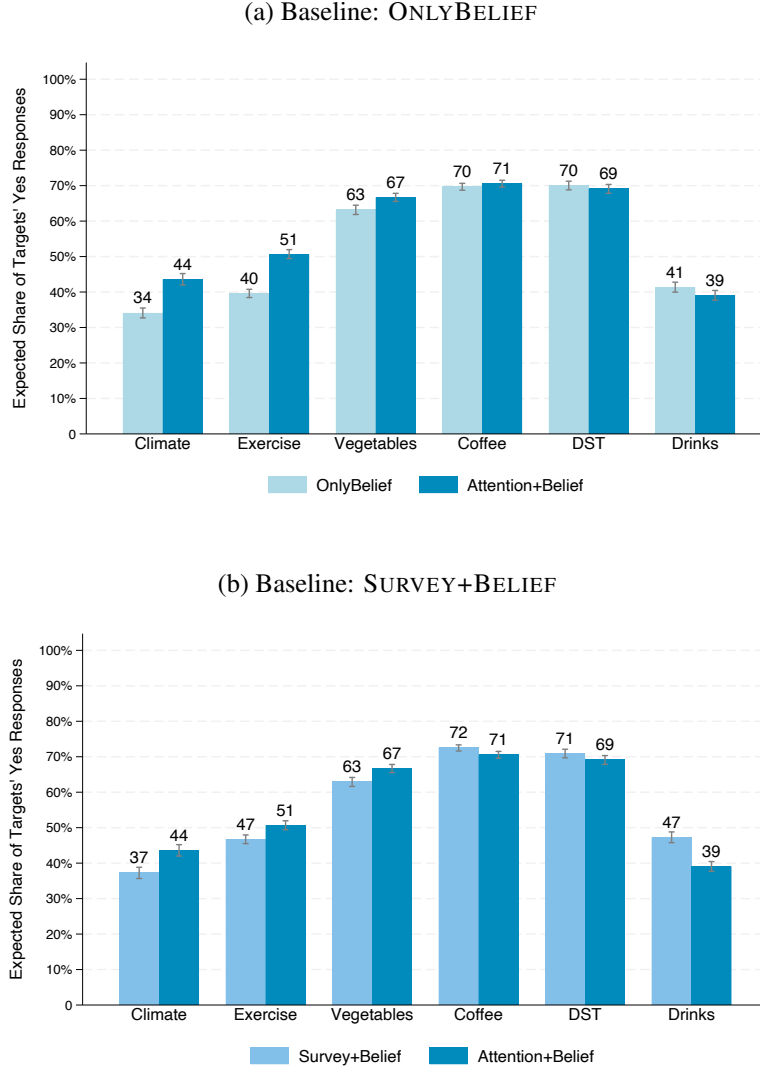


Figure 2: Effect of Attention Manipulation on Evaluators' Beliefs

*Notes:* This figure presents evaluators' mean beliefs about the share of targets stating they engage in a given behavior. The error bars represent standard errors of the mean. Evaluators with the attention manipulation as preceding task (ATTENTION+BELIEF) are compared to evaluators without any preceding task (ONLYBELIEF) in Panel (a) and to evaluators with the survey as preceding task (SURVEY+BELIEF) in Panel (b).

#### A. Stand-Alone Belief Elicitation

My primary manipulation consists of drawing evaluators' attention to targets' potential misreporting. If evaluators' beliefs about targets remain unchanged, we can conclude that evaluators already fully considered the possibility of targets misreporting and that the treatment thus has no influence. Conversely, if evaluators' beliefs about targets change due to this shift in attention, and particularly in the direction of likely response bias, it suggests that evaluators initially failed to fully attend to their considerations of self-report truthfulness. In my experiment, I can implement this simple test by comparing beliefs between ONLYBELIEF and BELIEF+ATTENTION, with differences between the two conditions being due to the attention manipulation.

Figure 2a illustrates the relevant comparisons. Overall, the data strongly support the alternative hypothesis. Specifically, beliefs about targets' desirable behavior significantly increase on average by 8.04

percentage points ( $p = 0.000$ ) in the ATTENTION+BELIEF condition, suggesting that perceived overreporting of good behavior is not fully reflected by beliefs in ONLYBELIEF despite meaningful financial incentives for belief accuracy. With a Cohen’s  $d$  of about 0.33, this change in beliefs is non-negligible. Importantly, I do not observe a similar increase in beliefs for the other topic categories. In a placebo test with neutral behavior, the average effect of the attention manipulation is not significantly different from 0 (point estimate:  $-0.03, p = 0.982$ ). Moreover, the belief about targets’ undesirable behavior at least moves in the direction of the perceived response bias (i.e., decreases), but not significantly so (point estimate:  $-2.30, p = 0.242$ ). Taken together, these results suggest that including the attention manipulation does not change beliefs per se, but changes in beliefs rather occur in the direction of perceived response bias by targets. This interpretation is further supported by the fact that beliefs increase for every desirable behavior (Climate:  $p = 0.000$ ; Exercise:  $p = 0.000$ ; Vegetables:  $p = 0.043$ ) but do not change for any neutral behavior (Coffee:  $p = 0.515$ ; DST:  $p = 0.591$ ).

### B. *Joint Elicitation of Self-Reports and Beliefs*

In light of the attention manipulation’s sizeable effect on beliefs, one might object that studies in economics frequently ask evaluators to answer the survey themselves prior to belief elicitation, making SURVEY+BELIEF the more relevant baseline for comparison in practice. Moreover, the introspection gained by answering the survey could, in principle, substitute for the attention manipulation, making the latter less effective or even redundant.<sup>15</sup> Next, I therefore compare beliefs about targets between SURVEY+BELIEF and ATTENTION+BELIEF. If the preceding survey already induces evaluators to fully account for their considerations about targets’ misreporting, then beliefs for a given behavior should not systematically differ between SURVEY+BELIEF and ATTENTION+BELIEF. Under the alternative hypothesis of incomplete adjustment, beliefs about targets in ATTENTION+BELIEF should still differ from the ones in SURVEY+BELIEF in line with the direction of perceived response bias.

Figure 2b illustrates the relevant comparisons. Overall, the data still clearly support the alternative hypothesis. Specifically, beliefs about targets’ desirable behavior significantly increase on average by 4.70 percentage points in the ATTENTION+BELIEF condition ( $p = 0.001$ ), suggesting that perceived overreporting of good behavior is still not fully reflected by evaluators’ beliefs in SURVEY+BELIEF despite meaningful financial incentives for belief accuracy and, conceivably, increased introspection from taking the survey themselves prior to belief elicitation. With a Cohen’s  $d$  of about 0.20, the change in beliefs is smaller than with ONLYBELIEF as the baseline, but still non-negligible. As in the previous section, I do not observe a comparable increase in beliefs for the other topic categories. In a placebo test with neutral behavior, the average treatment effect of the attention manipulation is not significantly different from 0 (point estimate:  $-1.86, p = 0.105$ ). Moreover, the belief about targets’ undesirable behavior moves in the direction of the perceived response bias (i.e., decreases), this time even significantly so (point estimate:  $-8.22, p = 0.000$ ). Taken together, these results suggest that the attention manipulation leads evaluators to account for perceived misreporting by targets’ beyond the introspection offered by taking the survey themselves. This interpretation is further supported by the fact that beliefs increase for every desirable behavior (Climate:  $p = 0.005$ ; Exercise:  $p = 0.025$ ; Vegetables:  $p = 0.027$ ) but do not change for any neutral behavior (Coffee:  $p = 0.139$ ; DST:  $p = 0.295$ ).

<sup>15</sup>Brownback, Burke and Gagnon-Bartsch (2024) find suggestive evidence for increased introspection from survey-taking in a subsequent inference task. For a comparison of ONLYBELIEF and SURVEY+BELIEF, see Supplemental Appendix B.3.

### C. Exploratory Analysis: Potential Implications for Cynical Beliefs

While the focus of this study is on the malleability of evaluators' beliefs about targets' self-reported behavior, a related question to ask is whether the measured magnitude of misperceptions is sensitive to the experimental condition used to elicit evaluators' beliefs. Next, I will thus explore how misperceptions for each sensitive behavior vary across the study's three conditions. Misperceptions are computed as the difference between (i) the mean belief of evaluators and (ii) the share of targets in the SURVEY+BELIEF condition stating that they engage in a given behavior, with both quantities calibrated to be more representative of the U.S. population.<sup>16</sup> Importantly, this exploratory analysis is only meant to be illustrative as the estimation of misperceptions with high precision would require a significantly larger sample size, especially due to the binary nature of the behavioral benchmarks. Moreover, misperceptions may, in principle, also be influenced by the fact that the sample in this study is only broadly representative of the U.S. population, with calibration only mitigating this issue to some extent.

Table 2: Misperceptions by Behavior and Condition (Calibrated)

	<b>Experimental Condition</b>		
	ONLYBELIEF	SURVEY+BELIEF	ATTENTION+BELIEF
<b>Behavior of Interest</b>			
Climate	-8.0	-5.0	2.7
Exercise	-24.5	-16.1	-11.5
Vegetables	-21.0	-14.9	-12.6
Drinks	20.1	21.8	17.1

*Notes:* This table reports point estimates for the average misperceptions, in percentage points, about sensitive behavior (rows), separately for each experimental condition assigned to evaluators (columns). Misperceptions are computed as the difference between the mean belief of evaluators and the share of targets in the SURVEY+BELIEF condition stating that they engage in a given behavior, with both quantities calibrated to be more representative of the U.S. population. Calibration is done using the raking procedure based on the population characteristics from Table A2. Positive (negative) values indicate that evaluators overestimate (underestimate) the share of targets stating they engage in a given behavior. Cynical beliefs correspond to negative values for *Climate*, *Exercise*, and *Vegetables* and positive values for *Drinks*.

Table 2 reports the average misperception for each sensitive behavior as a function of the condition assigned to evaluators, with positive (negative) values indicating that evaluators overestimate (underestimate) the share of targets stating that they engage in a given behavior. While this table is only illustrative for the reasons outlined earlier, it, by and large, replicates the cynical beliefs about targets' self-reported behavior documented by previous studies, albeit not necessarily with the same magnitude.<sup>17</sup> Specifically, there is a tendency to underestimate (overestimate) the prevalence of targets' self-reported desirable (undesirable) behavior. Importantly, the magnitude of misperception greatly varies across conditions. Cynical beliefs tend to be most pronounced in ONLYBELIEF and least pronounced in ATTENTION+BELIEF, with SURVEY+BELIEF representing the intermediate case. This pattern is consistent with increasing attention to perceived response bias as one progresses from ONLYBELIEF to

<sup>16</sup>Specifically, I use the *ipfraking* package in Stata (Kolenikov, 2014) to implement the raking procedure based on the population characteristics from Table A2. For technical details, see also Valliant and Dever (2018). As in Capozza and Srinivasan (2024), I derive the raking weights separately for each condition.

<sup>17</sup>It is worth pointing out that some benchmarks seem inherently volatile, possibly due to concurrent events at the time of data collection. I herein rely on the data from the main experiment to compute benchmark values as they were collected at the same time as beliefs. In addition, I ex-post collected about 500 additional survey responses a few days after the main experiment to increase estimation precision for the benchmark values. However, the resulting point estimates for the benchmark values obtained from the additional data can differ greatly from the ones in the main experiment. Importantly, this does not affect the general insights on the link between cynical beliefs and elicitation method put forward in this section. I report an overview of all benchmark values in Table C1 for completeness.

SURVEY+BELIEF and ultimately to ATTENTION+BELIEF. The use of the attention manipulation often greatly reduces cynical beliefs by about half or more relative to ONLYBELIEF, whereas the misperceptions in SURVEY+BELIEF and ATTENTION+BELIEF typically differ only by a few percentage points. However, in the climate domain, the effect of the attention manipulation compared to the preceding survey-taking is not only largest, but may also be pivotal in the sense that the former condition no longer generates evidence for cynical beliefs.

## V. Discussion

While this study is not a test of any specific theory, many of the observations in this study are at least consistent with attention to targets’ potential misreporting as the main channel. For example, observing smaller changes in beliefs for SURVEY+BELIEF than for ATTENTION+BELIEF is consistent with the idea that the attention manipulation is more powerful at correcting inattention to targets’ misreporting than is the introspection stemming from preceding survey-taking. Perhaps counterintuitively, I do not observe a change in beliefs for *Drinks* even though this behavior has the clearest social-desirability rating among all topics used in this study. But this finding is not necessarily at odds with attention to targets’ misreporting as the main channel. Importantly, if social image concerns are already very salient even without any additional intervention (for example, if evaluators are already attentive to the possibility that targets might lie about how much they drink), then there might also be very little inattention to correct. This would suggest that the effect of the attention manipulation may not necessarily be monotone in the strength of the desirability ratings, but more research would be needed to investigate the link between desirability ratings and the effect size of the attention manipulation.

An alternative interpretation of my results is that the attention manipulation induces some form of demand effect or another systematic change operating through a channel other than attention to perceived misreporting. I consider this to be unlikely for at least four reasons. First, the instructions are worded in a way that they do not take a stance on whether there is misreporting by targets in the first place, let alone in which direction.<sup>18</sup> Second, evaluators face meaningful financial incentives for belief accuracy, which is generally considered a precautionary measure against large demand effects (see, for example, Haaland, Roth and Wohlfart, 2023, for an influential review on the topic). Third, De Quidt, Haushofer and Roth (2018) find that demand effects are generally moderate in size, with bounds (i.e., difference of two opposite demand-effect treatments) averaging only 0.13 standard deviations. It is thus hard to imagine how a non-trivial demand effect could selectively cause changes in the magnitude of 0.20 to 0.33 standard deviations in one direction alone. Fourth, beliefs barely move in a placebo test with neutral behavior while they move in the direction of perceived response bias for sensitive behavior. This suggests that my attention manipulation does not generate any systematic changes in beliefs operating outside of perceived misreporting.

<sup>18</sup>One could go on to argue that asking evaluators about targets’ reporting tendencies causes evaluators to believe that targets are generally dishonest (or it could discourage evaluators from stating that targets are honest to not sound naive). But this seems unlikely as a supermajority of evaluators identifies *Coffee* and *DST* as neutral. In practical terms, saying that self-reports might or might not correspond to real behavior is only a very weak form of opposition compared to the common political discourse.

## **VI. Conclusion**

This study asks how to interpret cynical beliefs in the absence of actual behavior. To assess first-order concerns associated with the use of self-reported behavior, I propose a portable attention manipulation as a simple test. Comparing beliefs about targets' self-reported behavior with and without this attention manipulation then yields an easily interpretable measure of the degree to which limited attention about potential response bias produces beliefs that may be mistakenly attributed to real misperceptions. The results of my study have important practical implications for interpreting existing studies and designing future ones. In general, caution is warranted if misperceptions are computed from stand-alone belief elicitation. For some topics, even adding a preceding survey may be insufficient to ensure full adjustment for perceived misreporting. For future studies, researchers thus might wish to include additional precautionary measures to alleviate concerns about incomplete belief adjustment. The attention manipulation proposed by this study represents one easily implementable option to address this challenge. Alternatively, jointly eliciting self-reports and beliefs could represent a convenient and effective second-best option. However, further research is needed to provide a more systematic identification of settings at risk of incomplete belief adjustment and to explore its practical implications for measures of misperception.

## References

- Andre, Peter, Teodora Boneva, Felix Chopra, and Armin Falk.** 2024a. “Globally representative evidence on the actual and perceived support for climate action.” *Nature Climate Change*, 14(3): 253–259.
- Andre, Peter, Teodora Boneva, Felix Chopra, and Armin Falk.** 2024b. “Misperceived social norms and willingness to act against climate change.” *Review of Economics and Statistics*, 1–46.
- Aycinena, Diego, Francesco Bogliacino, and Erik O Kimbrough.** 2024. “Measuring norms: Assessing the threat of social desirability bias to the Bicchieri and Xiao elicitation method.” *Journal of Economic Behavior & Organization*, 222: 225–239.
- Bertrand, Marianne, and Sendhil Mullainathan.** 2001. “Do people mean what they say? Implications for subjective survey data.” *American Economic Review*, 91(2): 67–72.
- Braghieri, Luca.** 2024. “Political correctness, social image, and information transmission.” *Conditionally Accepted at the American Economic Review*.
- Brownback, Andy, Nathaniel Burke, and Tristan Gagnon-Bartsch.** 2024. “Inference from Biased Polls.” *Games and Economic Behavior, Forthcoming*.
- Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott.** 2020. “Misperceived social norms: Women working outside the home in Saudi Arabia.” *American Economic Review*, 110(10): 2997–3029.
- Bursztyn, Leonardo, Alexander W Cappelen, Bertil Tungodden, Alessandra Voena, and David H Yanagizawa-Drott.** 2024. “How are gender norms perceived?” Working Paper.
- Bursztyn, Leonardo, and David Y Yang.** 2022. “Misperceptions about others.” *Annual Review of Economics*, 14: 425–452.
- Cantoni, Davide, David Y Yang, Noam Yuchtman, and Y Jane Zhang.** 2019. “Protests as strategic games: experimental evidence from Hong Kong’s antiauthoritarian movement.” *The Quarterly Journal of Economics*, 134(2): 1021–1077.
- Capozza, Francesco, and Krishna Srinivasan.** 2024. “Who Should Get Money? Estimating Welfare Weights in the US.” *Working Paper*.
- Croson, Rachel TA.** 2007. “Theories of commitment, altruism and reciprocity: Evidence from linear public goods games.” *Economic Inquiry*, 45(2): 199–216.
- Danz, David, Lise Vesterlund, and Alistair J Wilson.** 2022. “Belief elicitation and behavioral incentive compatibility.” *American Economic Review*, 112(9): 2851–83.
- De Quidt, Jonathan, Johannes Haushofer, and Christopher Roth.** 2018. “Measuring and bounding experimenter demand.” *American Economic Review*, 108(11): 3266–3302.
- Drouvelis, Michalis, and Benjamin M Marx.** 2022. “Can charitable appeals identify and exploit belief heterogeneity?” *Journal of Economic Behavior & Organization*, 198: 631–649.
- Epley, Nicholas, Michael Kardas, Xuan Zhao, Stav Atir, and Juliana Schroeder.** 2022. “Undersociality: Miscalibrated social cognition can inhibit social connection.” *Trends in Cognitive Sciences*, 26(5): 406–418.
- Fischbacher, Urs, and Simon Gächter.** 2010. “Social preferences, beliefs, and the dynamics of free riding in public goods experiments.” *American Economic Review*, 100(1): 541–556.

- Frey, Bruno S, and Stephan Meier.** 2004. "Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment." *American economic review*, 94(5): 1717–1722.
- Gächter, Simon, and Elke Renner.** 2010. "The effects of (incentivized) belief elicitation in public goods experiments." *Experimental Economics*, 13: 364–377.
- Haaland, Ingar, Christopher Roth, and Johannes Wohlfart.** 2023. "Designing information provision experiments." *Journal of Economic Literature*, 61(1): 3–40.
- Hossain, Tanjim, and Ryo Okui.** 2013. "The binarized scoring rule." *Review of Economic Studies*, 80(3): 984–1001.
- Ho, Yuen, and Yihong Huang.** 2024. "Breaking the Spiral of Silence." *Working Paper*.
- Kahneman, Daniel.** 2012. "Thinking fast and slow."
- Kolenikov, Stanislav.** 2014. "Calibrating survey data using iterative proportional fitting (raking)." *The Stata Journal*, 14(1): 22–59.
- Levine, Timothy R.** 2014. "Truth-default theory (TDT) a theory of human deception and deception detection." *Journal of Language and Social Psychology*, 33(4): 378–392.
- Liu, Jiawei, and Jeff Niederdeppe.** 2020. "Misperceptions of the prevalence of health conditions and behaviors." *Journal of Health Communication*, 25(11): 903–916.
- Nederhof, Anton J.** 1985. "Methods of coping with social desirability bias: A review." *European Journal of Social Psychology*, 15(3): 263–280.
- Neugebauer, Tibor, Javier Perote, Ulrich Schmidt, and Malte Loos.** 2009. "Selfish-biased conditional cooperation: On the decline of contributions in repeated public goods experiments." *Journal of Economic Psychology*, 30(1): 52–60.
- Rosenfeld, Bryn, Kosuke Imai, and Jacob N Shapiro.** 2016. "An empirical validation study of popular survey methodologies for sensitive questions." *American Journal of Political Science*, 60(3): 783–802.
- Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rodgers, and Megan Schouweiler.** 2024. "IPUMS USA: Version 15.0 [dataset]."
- Valliant, Richard, and Jill A Dever.** 2018. *Survey weights: a step-by-step guide to calculation*. Vol. 1, Stata Press College Station, TX.
- Wilson, Alistair, and Emanuel Vespa.** 2016. "Paired-uniform scoring: Implementing a binarized scoring rule with non-mathematical language."

# Supplemental Appendices

## A. Preliminary Analysis

Table A1: Randomisation Check

Variable	Sample Shares				Balance Tests			
	OB	S+B	A+B	Overall	OB vs S+B	OB vs A+B	S+B vs A+B	Overall
<b>Gender</b>					0.93	0.86	0.79	0.96
Male	0.49	0.49	0.50	0.50				
Female	0.51	0.51	0.50	0.50				
<b>Age</b>					0.30	0.66	0.60	0.54
18–24	0.10	0.14	0.11	0.12				
25–34	0.21	0.14	0.17	0.17				
35–44	0.17	0.18	0.18	0.17				
45–54	0.16	0.18	0.14	0.16				
55–64	0.25	0.26	0.26	0.26				
65+	0.11	0.11	0.14	0.12				
<b>Income</b>					0.95	0.87	0.88	0.98
less than 25,000	0.12	0.10	0.12	0.11				
25,000–49,999	0.22	0.20	0.24	0.22				
50,000–74,999	0.18	0.21	0.21	0.20				
75,000–99,999	0.19	0.18	0.15	0.17				
100,000–149,999	0.18	0.19	0.16	0.18				
150,000 or more	0.11	0.12	0.12	0.12				
<b>Education</b>					0.33	0.81	0.78	0.75
Some high school or less	0.00	0.01	0.01	0.01				
High school diploma or GED	0.09	0.11	0.10	0.10				
Some college, but no degree	0.21	0.23	0.22	0.22				
Associates or technical degree	0.11	0.10	0.12	0.11				
Bachelor's degree	0.38	0.41	0.37	0.38				
Graduate or professional degree (MA, MS, MBA, PhD, JD, MD, DDS etc.)	0.21	0.14	0.18	0.17				
<b>Region</b>					0.11	0.24	0.16	0.13
Northeast	0.18	0.15	0.20	0.18				
Midwest	0.21	0.21	0.15	0.19				
South	0.45	0.39	0.44	0.43				
West	0.16	0.25	0.21	0.21				
<b>Politics</b>					0.73	0.06	0.18	0.16
Democrat	0.29	0.31	0.39	0.33				
Republican	0.28	0.30	0.27	0.28				
Independent	0.43	0.39	0.35	0.39				

Notes: p-values are based on chi-square tests of independence. OB: OnlyBelief; S+B: Survey+Belief; A+B: Attention+Belief

Table A2: Sample Representativeness

Variable	Prolific Sample	U.S. Population
<b>Gender</b>		
Male	0.50	0.49
Female	0.50	0.51
<b>Age</b>		
18–24	0.12	0.12
25–34	0.17	0.17
35–44	0.17	0.17
45–54	0.16	0.15
55–64	0.26	0.16
65+	0.12	0.23
<b>Income</b>		
less than 25,000	0.11	0.11
25,000–49,999	0.22	0.14
50,000–74,999	0.20	0.15
75,000–99,999	0.17	0.13
100,000–149,999	0.18	0.20
150,000 or more	0.12	0.26
<b>Education</b>		
Some high school or less	0.01	0.10
High school diploma or GED	0.10	0.27
Some college, but no degree	0.22	0.20
Associates or technical degree	0.11	0.08
Bachelor's degree	0.38	0.21
Graduate or professional degree (MA, MS, MBA, PhD, JD, MD, DDS etc.)	0.17	0.13
<b>Region</b>		
Northeast	0.18	0.17
Midwest	0.19	0.20
South	0.43	0.39
West	0.21	0.24
<b>Politics</b>		
Democrat	0.33	0.28
Republican	0.28	0.30
Independent	0.39	0.41

Notes: I use IPUMS data (Ruggles et al., 2024) from the American Community Survey 2023 to derive U.S. population shares. Political affiliation comes from Gallup (<https://news.gallup.com/poll/15370/party-affiliation.aspx>) as of 1 July 2024.

Table A3: Descriptive Statistics (Beliefs About Targets)

	OnlyBelief (N=247)		Survey+Belief (N=251)		Attention+Belief (N=249)	
	mean	sd	mean	sd	mean	sd
Climate	34.10	21.90	37.25	25.14	43.61	25.29
Exercise	39.61	18.17	46.73	19.28	50.68	19.93
Vegetables	63.17	20.57	62.90	20.10	66.69	17.98
Coffee	69.68	15.17	72.47	13.81	70.56	15.00
DST	70.03	19.18	70.91	19.14	69.09	19.60
Drinks	41.37	22.05	47.29	23.98	39.07	21.72

## B. Detailed Empirical Results

### B.1. ONLYBELIEF vs ATTENTION+BELIEF

Table B1: Beliefs About Targets  
(OnlyBelief vs Attention+Belief)

	Pooled
Attention+Belief $\times$ Desirable	8.04 (0.000)
Attention+Belief $\times$ Neutral	-0.03 (0.982)
Attention+Belief $\times$ Undesirable	-2.30 (0.242)
Climate	34.84 (0.000)
Exercise	41.13 (0.000)
Vegetables	60.91 (0.000)
Coffee	70.14 (0.000)
DST	69.57 (0.000)
Drinks	41.37 (0.000)
H0: Attention+Belief $\times$ Desirable = Attention+Belief $\times$ Neutral (p-value)	0.000
H0: Attention+Belief $\times$ Undesirable = Attention+Belief $\times$ Neutral (p-value)	0.275
Observations	2976

*Notes:* Errors are clustered on the subject level. Two-sided p-values are reported in parentheses.

Table B2: Beliefs About Targets  
(OnlyBelief vs Attention+Belief, Topics)

	Climate	Exercise	Vegetables	Coffee	DST	Drinks
Attention+Belief	9.51 (0.000)	11.08 (0.000)	3.52 (0.043)	0.88 (0.515)	-0.94 (0.591)	-2.30 (0.242)
Constant	34.10 (0.000)	39.61 (0.000)	63.17 (0.000)	69.68 (0.000)	70.03 (0.000)	41.37 (0.000)
Observations	496	496	496	496	496	496

*Notes:* Errors are clustered on the subject level. Two-sided p-values are reported in parentheses.

Table B3: Beliefs About Targets  
(OnlyBelief vs Attention+Belief; Robustness Check)

	Dem.	Rep.	Ind.
Attention+Belief $\times$ Desirable	5.40 (0.017)	6.22 (0.036)	11.63 (0.000)
Attention+Belief $\times$ Neutral	1.15 (0.573)	-1.07 (0.631)	-0.21 (0.914)
Attention+Belief $\times$ Undesirable	-4.34 (0.179)	-2.21 (0.600)	0.48 (0.875)
Climate	35.29 (0.000)	38.89 (0.000)	31.90 (0.000)
Exercise	42.18 (0.000)	45.00 (0.000)	37.81 (0.000)
Vegetables	61.46 (0.000)	64.95 (0.000)	57.89 (0.000)
Coffee	68.27 (0.000)	72.99 (0.000)	69.60 (0.000)
DST	69.28 (0.000)	69.58 (0.000)	69.67 (0.000)
Drinks	38.92 (0.000)	46.04 (0.000)	39.93 (0.000)
H0: Attention+Belief $\times$ Desirable = Attention+Belief $\times$ Neutral (p-value)	0.097	0.044	0.000
H0: Attention+Belief $\times$ Undesirable = Attention+Belief $\times$ Neutral (p-value)	0.116	0.799	0.833
Politics Control	SPLIT	SPLIT	SPLIT
Observations	1002	816	1158

Notes: Errors are clustered on the subject level. Two-sided p-values are reported in parentheses.

Table B4: Beliefs About Targets  
(OnlyBelief vs Attention+Belief, Topics; Robustness Check)

	Climate	Exercise	Vegetables	Coffee	DST	Drinks
Attention+Belief	9.59 (0.000)	11.08 (0.000)	3.51 (0.045)	1.00 (0.457)	-0.99 (0.573)	-1.87 (0.344)
Democrat	32.88 (0.000)	38.92 (0.000)	62.55 (0.000)	68.35 (0.000)	70.52 (0.000)	37.49 (0.000)
Republican	37.26 (0.000)	42.64 (0.000)	66.27 (0.000)	71.98 (0.000)	69.54 (0.000)	45.88 (0.000)
Independent	32.82 (0.000)	38.06 (0.000)	61.55 (0.000)	69.05 (0.000)	70.02 (0.000)	40.99 (0.000)
Politics Control	FE	FE	FE	FE	FE	FE
Observations	496	496	496	496	496	496

Notes: Errors are clustered on the subject level. Two-sided p-values are reported in parentheses.

## B.2. SURVEY+BELIEF vs ATTENTION+BELIEF

Table B5: Beliefs About Targets  
(Survey+Belief vs Attention+Belief)

	Pooled
Attention+Belief $\times$ Desirable	4.70 (0.001)
Attention+Belief $\times$ Neutral	-1.86 (0.105)
Attention+Belief $\times$ Undesirable	-8.22 (0.000)
Climate	38.08 (0.000)
Exercise	46.36 (0.000)
Vegetables	62.45 (0.000)
Coffee	72.45 (0.000)
DST	70.93 (0.000)
Drinks	47.29 (0.000)
H0: Attention+Belief $\times$ Desirable = Attention+Belief $\times$ Neutral (p-value)	0.000
H0: Attention+Belief $\times$ Undesirable = Attention+Belief $\times$ Neutral (p-value)	0.003
Observations	3000

Notes: Errors are clustered on the subject level. Two-sided p-values are reported in parentheses.

Table B6: Beliefs About Targets  
(Survey+Belief vs Attention+Belief, Topics)

	Climate	Exercise	Vegetables	Coffee	DST	Drinks
Attention+Belief	6.36 (0.005)	3.95 (0.025)	3.79 (0.027)	-1.91 (0.139)	-1.82 (0.295)	-8.22 (0.000)
Constant	37.25 (0.000)	46.73 (0.000)	62.90 (0.000)	72.47 (0.000)	70.91 (0.000)	47.29 (0.000)
Observations	500	500	500	500	500	500

Notes: Errors are clustered on the subject level. Two-sided p-values are reported in parentheses.

### B.3. Direct Comparison: ONLYBELIEF vs SURVEY+BELIEF

A natural question is whether one of the two standard approaches, ONLYBELIEF or SURVEY+BELIEF, is preferable to the other. Therefore, I compare beliefs about targets in these two conditions. Figure B1 illustrates the relevant comparisons, while Tables B7 and B8 report the full analysis. Overall, the picture is mixed. On the one hand, beliefs in SURVEY+BELIEF go up by an average 3.34 percentage points for desirable behavior ( $p = 0.015$ ), suggesting at least partial adjustment for evaluators' considerations about targets' self-report truthfulness. On the other hand, this adjustment does not occur for all desirable behaviors, and I even observe increases for neutral (point estimate: 1.84,  $p = 0.096$ )<sup>19</sup> and undesirable behavior (point estimate: 5.92,  $p = 0.004$ ). Taken together, this suggests that preceding survey-taking by evaluators prior to belief elicitation works as a mitigation strategy only in some cases and might even introduce unintended considerations.

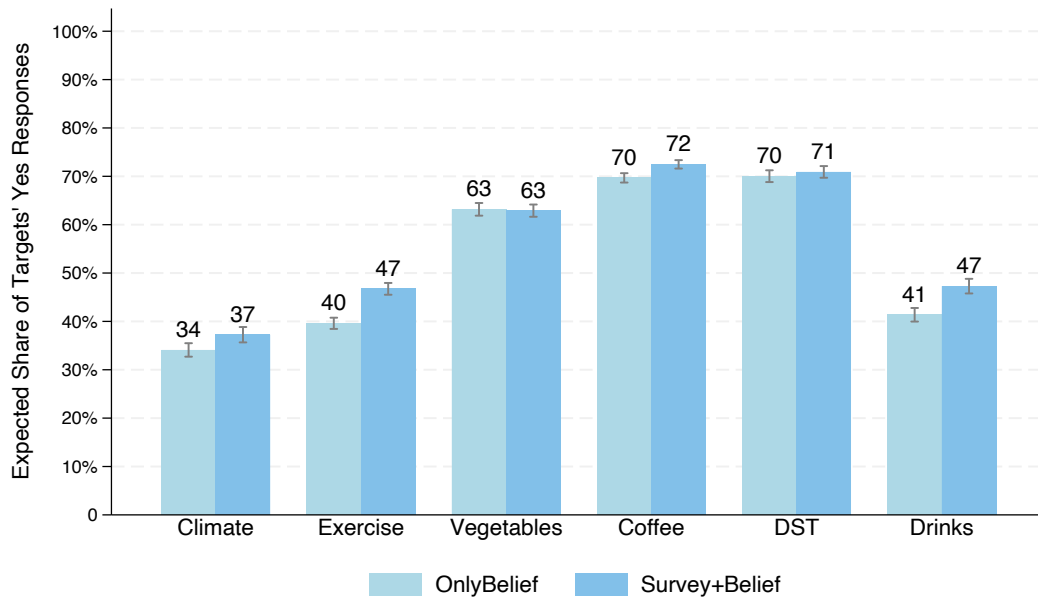


Figure B1: Effect of Preceding Survey

<sup>19</sup>The effect of the preceding survey does not significantly differ for desirable and neutral behavior ( $p = 0.361$ ).

Table B7: Beliefs About Targets  
(OnlyBelief vs Survey+Belief)

	Pooled
Survey+Belief $\times$ Desirable	3.34 (0.015)
Survey+Belief $\times$ Neutral	1.84 (0.096)
Survey+Belief $\times$ Undesirable	5.92 (0.004)
Climate	34.01 (0.000)
Exercise	41.52 (0.000)
Vegetables	61.36 (0.000)
Coffee	70.16 (0.000)
DST	69.55 (0.000)
Drinks	41.37 (0.000)
H0: Survey+Belief $\times$ Desirable = Survey+Belief $\times$ Neutral (p-value)	0.361
H0: Survey+Belief $\times$ Undesirable = Survey+Belief $\times$ Neutral (p-value)	0.064
Observations	2988

Notes: Errors are clustered on the subject level. Two-sided p-values are reported in parentheses.

Table B8: Beliefs About Targets  
(OnlyBelief vs Survey+Belief, Topics)

	Climate	Exercise	Vegetables	Coffee	DST	Drinks
Survey+Belief	3.15 (0.136)	7.13 (0.000)	-0.27 (0.881)	2.79 (0.032)	0.88 (0.609)	5.92 (0.004)
Constant	34.10 (0.000)	39.61 (0.000)	63.17 (0.000)	69.68 (0.000)	70.03 (0.000)	41.37 (0.000)
Observations	498	498	498	498	498	498

Notes: Errors are clustered on the subject level. Two-sided p-values are reported in parentheses.

### C. Detailed Exploratory Analysis

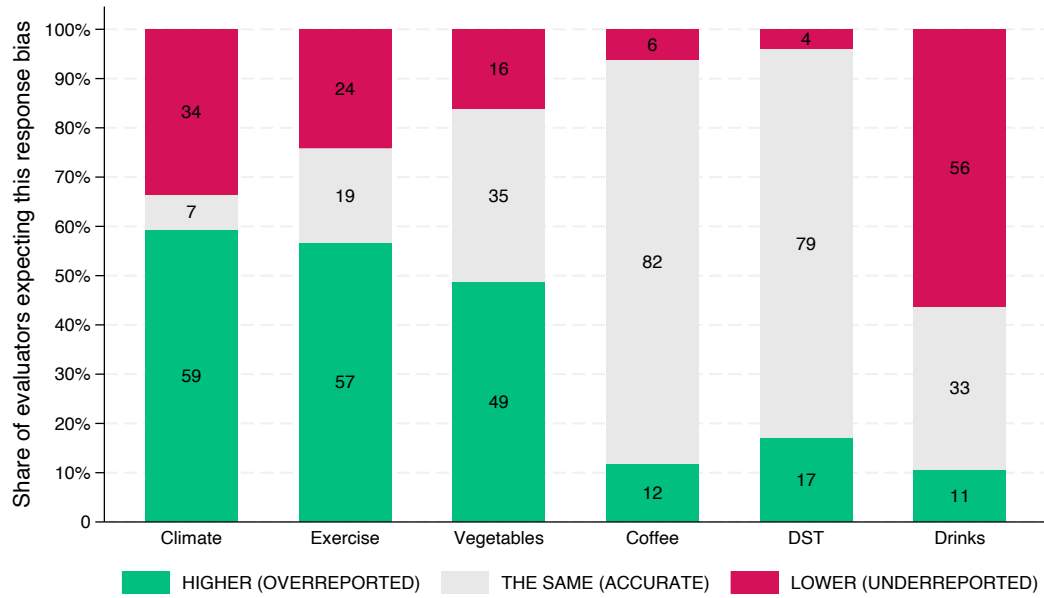


Figure C1: Perceived Misreporting (Calibrated)

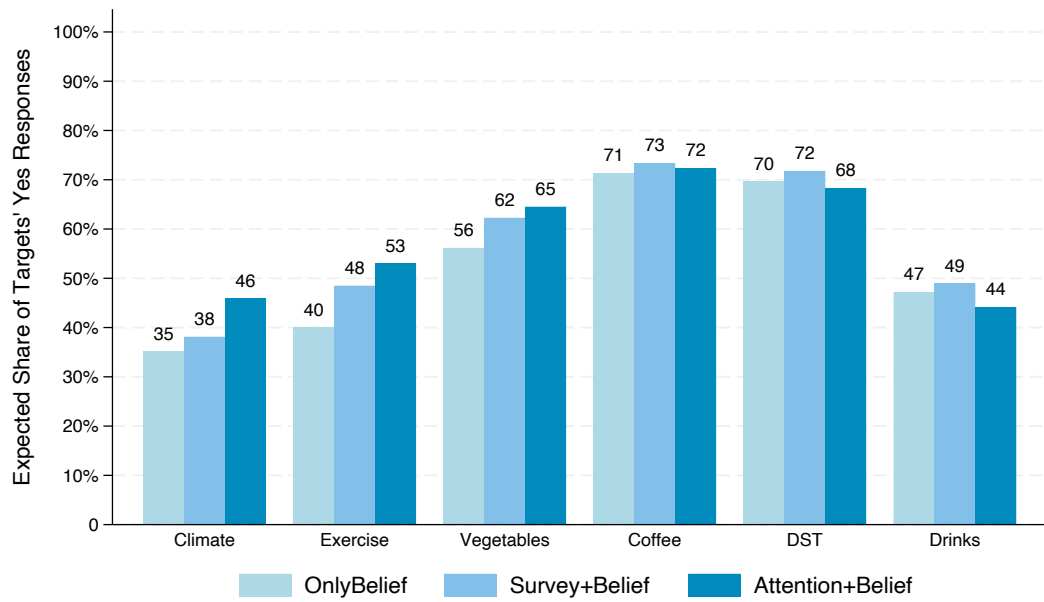


Figure C2: Beliefs About Self-Reported Behavior (Calibrated)

Table C1: Overview of Benchmarks

	Main Experiment (N=251)		Additional Data (N=496)		Combined (N=747)	
	R	C	R	C	R	C
Climate	0.38	0.43	0.32	0.31	0.34	0.35
Exercise	0.63	0.65	0.55	0.59	0.58	0.59
Vegetables	0.81	0.77	0.79	0.70	0.80	0.73
Coffee	0.57	0.58	0.52	0.46	0.54	0.53
DST	0.82	0.85	0.81	0.84	0.81	0.84
Drinks	0.25	0.27	0.26	0.21	0.26	0.25

*Notes:* R=raw; C=calibrated. Raking weights are based on the population characteristics from Table A2.

## D. Experimental Procedures

### D.1. *Design Overview*

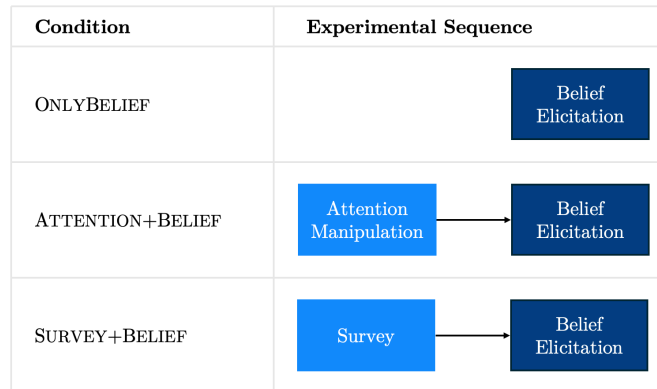


Figure D3: Design Overview

## D.2. Full Instructions

### D.2.1. Introduction (ALL CONDITIONS)

#### INFORMED CONSENT

This study is carried out as part of a research project at the University of Zurich, Switzerland. **For our research, it is important that you read all instructions and questions carefully before answering.**

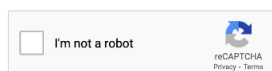
#### STUDY DETAILS:

- On average, we expect this study to take **about 12 minutes** to complete.
- You will receive a **fixed payment of \$2.40** for taking part in this study.
- You have the chance to win an **additional bonus payment of \$5** depending on your decisions during this study.
- The data collected about you will be **anonymized** and will never be used to identify you as an individual. Anonymized datasets may be transferred to academic journals or other researchers for scientific purposes.
- Participation in this study is **voluntary**. You may withdraw from it at any point during the study. Only complete submissions will be eligible for payment.

If you have any questions regarding this study, please contact us via Prolific.

☐ I have read the above and give my consent to participate

☐ I do not give my consent and wish to exit this study



---

[PAGE BREAK]

---

#### REGISTRATION

**What is your Prolific ID?**

*Please note that this response should auto-fill with the correct ID.*



---

[PAGE BREAK]

---

The next question is about the following problem. In questionnaires like ours, sometimes there are participants who do not carefully read the questions and just quickly click through the survey. This compromises the results of research studies. To show that you are reading the survey carefully, please choose both "Very strongly interested" and "Not at all interested" as your answer to the next question.

**Based on the text you read above, which answers have you been asked to enter?**

Very strongly interested

Very interested

A little bit interested

Not very interested

Not at all interested



---

[PAGE BREAK]

---

*Only if previous attention check was failed:*

The next question is about the following problem. In questionnaires like ours, sometimes there are participants who do not carefully read the questions and just quickly click through the survey. This compromises the results of research studies. To show that you are reading the survey carefully, please choose both "Very interested" and "Not very interested" as your answer to the next question.

**Based on the text you read above, which answers have you been asked to enter?**

Very strongly interested

Very interested

A little bit interested

Not very interested

Not at all interested



---

[PAGE BREAK]

---

Thank you for reading our questions carefully! You will now proceed to Part 1 of this study.



---

[PAGE BREAK]

---

### D.2.2. Attention Manipulation (ATTENTION+BELIEF ONLY)

#### PART 1: INSTRUCTIONS

When people are asked about their behavior in surveys, they might or might not tell the truth. The extent to which they tell the truth might vary depending on the nature of the topic.

In this part of the study, you will see 6 short questions.

For each question, your task is to indicate to what extent you believe the responses from fellow U.S. participants, collected anonymously in a broadly representative online survey, truly reflect their behavior.

Please read each screen carefully and answer as best you can.

*The continue button will only appear after 15 seconds.*

---

[PAGE BREAK]

---

Remember that your task is to indicate to **what extent you believe the responses** from fellow U.S. participants, collected anonymously in a broadly representative online survey, **truly reflect their behavior**.

Now consider the question displayed in the [blue box](#).

Question shown to fellow U.S. participants:

Would you be willing to contribute 1% of your household income every month to fight global warming? This would mean that you would contribute \$1 for every \$100 of this income.

Yes

No

Do you think the number of participants who answer 'Yes' to this question is lower, the same, or higher than the number of participants who would actually do this activity?

Number of 'Yes' responses is lower than the number who would actually do this activity

Number of 'Yes' responses is the same as the number who would actually do this activity

Number of 'Yes' responses is higher than the number who would actually do this activity

*The continue button will only appear after 10 seconds.*

---

[PAGE BREAK]

---

Remember that your task is to indicate to **what extent you believe the responses** from fellow U.S. participants, collected anonymously in a broadly representative online survey, **truly reflect their behavior**.

Now consider the question displayed in the [blue box](#).

Question shown to fellow U.S. participants:

Do you typically participate in 150 minutes (2.5 hours) or more of aerobic physical activity (like walking, running, or bicycling) per week?

Yes

No

Do you think the number of participants who answer 'Yes' to this question is lower, the same, or higher than the number of participants who actually do this activity?

Number of 'Yes' responses is lower than the number who actually do this activity

Number of 'Yes' responses is the same as the number who actually do this activity

Number of 'Yes' responses is higher than the number who actually do this activity

*The continue button will only appear after 10 seconds.*

---

[PAGE BREAK]

---

Remember that your task is to indicate **to what extent you believe the responses** from fellow U.S. participants, collected anonymously in a broadly representative online survey, **truly reflect their behavior**.

Now consider the question displayed in the [blue box](#).

Question shown to fellow U.S. participants:

Do you typically consume vegetables at least once per day?

Yes

No

Do you think the number of participants who answer '[Yes](#)' to this question is lower, the same, or higher than the number of participants who actually do this activity?

Number of 'Yes' responses is lower than the number who actually do this activity

Number of 'Yes' responses is the same as the number who actually do this activity

Number of 'Yes' responses is higher than the number who actually do this activity

*The continue button will only appear after 10 seconds.*

---

[PAGE BREAK]

---

Remember that your task is to indicate **to what extent you believe the responses** from fellow U.S. participants, collected anonymously in a broadly representative online survey, **truly reflect their behavior**.

Now consider the question displayed in the [blue box](#).

Question shown to fellow U.S. participants:

Do you typically drink coffee every day?

Yes

No

Do you think the number of participants who answer '[Yes](#)' to this question is lower, the same, or higher than the number of participants who actually do this activity?

Number of 'Yes' responses is lower than the number who actually do this activity

Number of 'Yes' responses is the same as the number who actually do this activity

Number of 'Yes' responses is higher than the number who actually do this activity

*The continue button will only appear after 10 seconds.*

---

[PAGE BREAK]

---

Remember that your task is to indicate **to what extent you believe the responses** from fellow U.S. participants, collected anonymously in a broadly representative online survey, **truly reflect their behavior**.

Now consider the question displayed in the [blue box](#).

Question shown to fellow U.S. participants:

Would you support making Daylight Saving Time (DST) permanent in the US to avoid changing clocks twice a year?

Yes

No

Do you think the number of participants who answer 'Yes' to this question is lower, the same, or higher than the number of participants who would actually do this activity?

Number of 'Yes' responses is lower than the number who would actually do this activity

Number of 'Yes' responses is the same as the number who would actually do this activity

Number of 'Yes' responses is higher than the number who would actually do this activity

The continue button will only appear after 10 seconds.

---

[PAGE BREAK]

---

Remember that your task is to indicate **to what extent you believe the responses** from fellow U.S. participants, collected anonymously in a broadly representative online survey, **truly reflect their behavior**.

Now consider the question displayed in the [blue box](#).

Question shown to fellow U.S. participants:

One drink is equivalent to a 12-ounce beer, a 5-ounce glass of wine, or a drink with one shot of liquor.  
In the past 30 days, have you consumed five or more drinks on any single occasion if you are male, or four or more drinks if you are female?

Yes

No

Do you think the number of participants who answer 'Yes' to this question is lower, the same, or higher than the number of participants who have actually done this activity?

Number of 'Yes' responses is lower than the number who have actually done this activity

Number of 'Yes' responses is the same as the number who have actually done this activity

Number of 'Yes' responses is higher than the number who have actually done this activity

The continue button will only appear after 10 seconds.

---

[PAGE BREAK]

---

Remember that, when people are asked about their behavior in surveys, they might or might not tell the truth. The extent to which they tell the truth might vary depending on the nature of the topic.

Based on the 6 questions you just saw, can you think of reasons why fellow U.S. participants might or might not tell the truth in a broadly representative and anonymous online survey? Please briefly describe such potential reasons (1-3 sentences).

The continue button will only appear after 10 seconds.

---

[PAGE BREAK]

---

D.2.3. *Survey (SURVEY+BELIEF ONLY)*

PART 1: INSTRUCTIONS

In this part of the study, you will complete a brief survey that consists of 6 questions.

For each question, there will be two answer options (Yes/No), and you have to select exactly one option.

All your answers will be recorded anonymously.

The continue button will only appear after 10 seconds.

[PAGE BREAK]

Would you be willing to contribute 1% of your household income every month to fight global warming? This would mean that you would contribute \$1 for every \$100 of this income.

Yes

No

The continue button will only appear after 10 seconds.

[PAGE BREAK]

Do you typically participate in 150 minutes (2.5 hours) or more of aerobic physical activity (like walking, running, or bicycling) per week?

Yes

No

The continue button will only appear after 10 seconds.

[PAGE BREAK]

Do you typically consume vegetables at least once per day?

Yes

No

The continue button will only appear after 10 seconds.

[PAGE BREAK]

Do you typically drink coffee every day?

Yes

No

The continue button will only appear after 10 seconds.

[PAGE BREAK]

Would you support making Daylight Saving Time (DST) permanent in the US to avoid changing clocks twice a year?

Yes

No

The continue button will only appear after 10 seconds.

[PAGE BREAK]

One drink is equivalent to a 12-ounce beer, a 5-ounce glass of wine, or a drink with one shot of liquor.  
**In the past 30 days, have you consumed five or more drinks on any single occasion if you are male, or four or more drinks if you are female?**

Yes

No

*The continue button will only appear after 10 seconds.*

---

[PAGE BREAK]

---

## D.2.4. Belief Task (ALL CONDITIONS)

### PART 2: INSTRUCTIONS

You can win an additional \$5 bonus in this part of the study.

#### Your Task

Your task will be to guess the answers from fellow U.S. participants, collected anonymously in a broadly representative online survey. These participants first had to pass an attention check before being given 6 questions. For each question, there were exactly two answer options (Yes/No), and they had to select exactly one option after spending at least 10 seconds on this question.

#### Your Guess

For each of the 6 questions, you will see the same screen that was shown to fellow U.S. participants. After that, you will have to guess how many fellow U.S. participants (out of 100) chose the answer option 'Yes'. The number you provide is called *Your Guess*.

#### Final Bonus Payment

- At the end of the study, the computer will randomly choose 1 in every 5 participants who may win an additional \$5 bonus.
- For those participants, the computer will then randomly choose 1 of the 6 questions for payment.
- For the selected question, we will use *Your Guess* and the answers from fellow U.S. participants for the selected question in order to determine whether you win \$5 based on a Payment Rule. You can find the details of the Payment Rule below if you are interested. **Importantly, the Payment Rule is designed so that you can secure the largest chance of winning the \$5 prize by reporting what you believe to be the most accurate guess.**

► For details about the Payment Rule, click here.

To make sure you understand these instructions, please answer the following questions:

#### What is your task in this part of the study?

I have to guess how many fellow U.S. participants chose the answer option 'Yes'

I have to choose between two answer options based on my own preferences

I have to state how fellow U.S. participants should have answered

#### What is the best you can do to maximize your chance of winning the \$5 prize?

I should report low numbers

I should report medium numbers

I should report high numbers

I should report the most accurate guess possible

The continue button will only appear after 15 seconds.

*NB: In SURVEY+BELIEF, the second sentence of Your Task instead reads 'These participants first had to pass an attention check before being given the same 6 questions that you just saw.'*

*If subjects click on details:*

▼ For details about the Payment Rule, click here.

The following explains how *Your Guess* is used to determine whether you win the \$5 prize:

1. The computer chooses two numbers between 1 and 100, where each number is equally likely, as if rolling two 100-sided dice. These numbers are called *Computer Number A* and *Computer Number B*.
2. The computer determines whether you win the \$5 prize depending on which answer option a randomly-selected fellow U.S. participant chose on the selected question:
  - **Answer option 'Yes' was chosen:** You will win the \$5 prize if *Your Guess* is **greater than or equal to either** of the two *Computer Numbers*.
  - **Answer option 'No' was chosen:** You will win the \$5 prize if *Your Guess* is **less than either** of the two *Computer Numbers*.

While it may seem complicated, this Payment Rule means that you have the highest chance of winning the \$5 prize if *Your Guess* is as close as possible to the true likelihood that a randomly-selected fellow U.S. participant chose the answer option 'Yes' for that question.

[PAGE BREAK]

You have correctly answered the comprehension questions! You are now ready to provide *Your Guesses* for Part 2 of this study.



---

[PAGE BREAK]

---

Remember that your task is to predict the answers from fellow U.S. participants, collected anonymously in a broadly representative online survey. The more accurate *Your Guess* is about their answers, the more likely you are to win the \$5 prize.

Now consider the question displayed in the [blue box](#).

**Question shown to fellow U.S. participants:**

Would you be willing to contribute 1% of your household income every month to fight global warming? This would mean that you would contribute \$1 for every \$100 of this income.

Yes

No

**Your Guess:**

How many fellow U.S. participants (out of 100) do you think chose the answer option 'Yes' to this question?

0 10 20 30 40 50 60 70 80 90 100



*The continue button will only appear after 10 seconds.*

---

[PAGE BREAK]

---

Remember that your task is to predict the answers from fellow U.S. participants, collected anonymously in a broadly representative online survey. The more accurate *Your Guess* is about their answers, the more likely you are to win the \$5 prize.

Now consider the question displayed in the [blue box](#).

**Question shown to fellow U.S. participants:**

Do you typically participate in 150 minutes (2.5 hours) or more of aerobic physical activity (like walking, running, or bicycling) per week?

Yes

No

**Your Guess:**

How many fellow U.S. participants (out of 100) do you think chose the answer option 'Yes' to this question?

0 10 20 30 40 50 60 70 80 90 100



*The continue button will only appear after 10 seconds.*

---

[PAGE BREAK]

---

Remember that your task is to predict the answers from fellow U.S. participants, collected anonymously in a broadly representative online survey. The more accurate *Your Guess* is about their answers, the more likely you are to win the \$5 prize.

Now consider the question displayed in the [blue box](#).

**Question shown to fellow U.S. participants:**

Do you typically consume vegetables at least once per day?

Yes

No

**Your Guess:**

How many fellow U.S. participants (out of 100) do you think chose the answer option 'Yes' to this question?

0 10 20 30 40 50 60 70 80 90 100



*The continue button will only appear after 10 seconds.*

---

[PAGE BREAK]

---

Remember that your task is to predict the answers from fellow U.S. participants, collected anonymously in a broadly representative online survey. The more accurate *Your Guess* is about their answers, the more likely you are to win the \$5 prize.

Now consider the question displayed in the [blue box](#).

**Question shown to fellow U.S. participants:**

Do you typically drink coffee every day?

Yes

No

**Your Guess:**

How many fellow U.S. participants (out of 100) do you think chose the answer option 'Yes' to this question?

0 10 20 30 40 50 60 70 80 90 100



*The continue button will only appear after 10 seconds.*

---

[PAGE BREAK]

---

Remember that your task is to predict the answers from fellow U.S. participants, collected anonymously in a broadly representative online survey. The more accurate *Your Guess* is about their answers, the more likely you are to win the \$5 prize.

Now consider the question displayed in the [blue box](#).

**Question shown to fellow U.S. participants:**

Would you support making Daylight Saving Time (DST) permanent in the US to avoid changing clocks twice a year?

Yes

No

**Your Guess:**

How many fellow U.S. participants (out of 100) do you think chose the answer option 'Yes' to this question?

0 10 20 30 40 50 60 70 80 90 100

*The continue button will only appear after 10 seconds.*

[PAGE BREAK]

Remember that your task is to predict the answers from fellow U.S. participants, collected anonymously in a broadly representative online survey. The more accurate *Your Guess* is about their answers, the more likely you are to win the \$5 prize.

Now consider the question displayed in the [blue box](#).

**Question shown to fellow U.S. participants:**

One drink is equivalent to a 12-ounce beer, a 5-ounce glass of wine, or a drink with one shot of liquor.  
In the past 30 days, have you consumed five or more drinks on any single occasion if you are male, or four or more drinks if you are female?

Yes

No

**Your Guess:**

How many fellow U.S. participants (out of 100) do you think chose the answer option 'Yes' to this question?

0 10 20 30 40 50 60 70 80 90 100

*The continue button will only appear after 10 seconds.*

[PAGE BREAK]

Please briefly explain how you made *Your Guesses* (1-3 sentences).

*The continue button will only appear after 10 seconds.*

[PAGE BREAK]

### D.3. *Additional Information (ALL CONDITIONS)*

#### PART 3

Please answer the following short background questions below:

How old are you?

What is your sex?

What is the highest level of education you have completed?

What was your total household income before taxes during the past 12 months?

In which state do you currently live?

What best describes your political affiliation?

---

[PAGE BREAK]

---

Lastly, do you have any comments concerning this study (e.g. technical difficulties, parts that were unclear to you)?



---

[END OF SURVEY]

---