

Supplementary Appendix for: Who's to Blame for Survey Instability: Respondents with Nonexistent Preferences or Researchers with Flawed Measures?*

Libby Jenke[†] Gary King[‡]

June 16, 2026

Contents

A1 Variants of our Experimental Platform	1
A2 Self-reported memory of Q1a does not affect recall accuracy	1
A3 Priming	2
A4 Response Option Divergence and Approximate Indifference	3
A5 Failed Attempts to Experimentally Control Mind Wandering	4
A6 Preoccupation Predicts Instability on Non-Conjoint Questions	7
A7 Eye Tracking Data Collections	8
A8 Proposed Norms for Ethical Survey Research	11
A9 Data collections	12

*This appendix, the paper, associated software, and data are available at GaryKing.org/instability.

[†]Assistant Professor, Department of Political Science, University of Houston; LibbyJenke.com, LJenke@uh.edu.

[‡]Albert J. Weatherhead III University Professor, Institute for Quantitative Social Science, Harvard University; GaryKing.org, King@Harvard.edu.

A1 Variants of our Experimental Platform

Figure A1 presents three alternative experimental setups. In Survey 2, we tested whether increasing the number of distractor questions between Q1a and Q1b affects response instability. The results indicate that the number of distractor questions had little to no impact, which is likely because conjoints are sufficiently complicated that keeping all the attributes for both choices in your head for even a single question is too difficult, and so many distractors may not be necessary. In Survey 1, we investigated whether introducing burn-in questions prior to Q1a would reduce instability. This intervention also showed no meaningful effect.

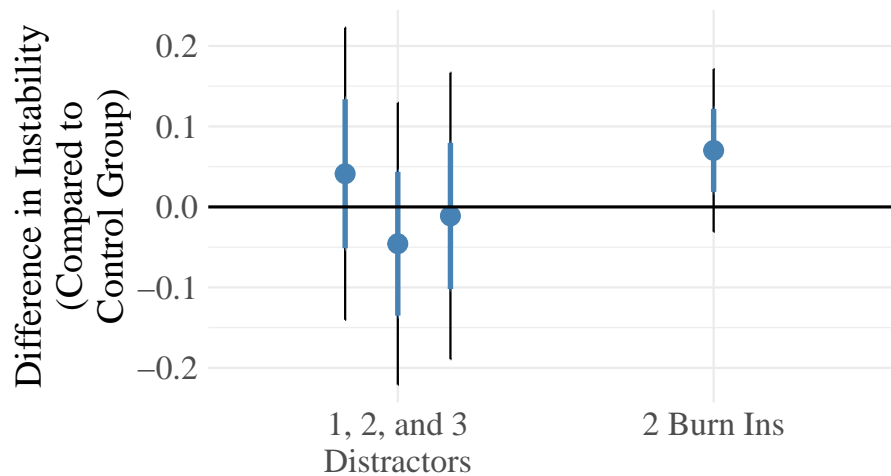


Figure A1: Experimental Platform Validation: No difference in instability with different numbers of distractor questions ($n = 208$; control group received no distractor questions) or burn-in questions ($n = 298$; control group received no burn-in questions).

A2 Self-reported memory of Q1a does not affect recall accuracy

Do respondents remember encountering Q1a and the answer they gave when they later receive Q1b? Self-reported memory may be an unreliable indicator, as respondents often report remembering prior questions even when they do not. Figure A2 compares the accuracy of recalling responses to Q1a among three groups: those who reported full memory, partial memory, or no memory of having seen Q1a (Survey 10). The results reveal no

statistically or substantively significant differences in recall accuracy across these groups. Even respondents who claimed to remember seeing the candidates were no more likely to accurately recall their earlier preference than those who reported no memory at all.

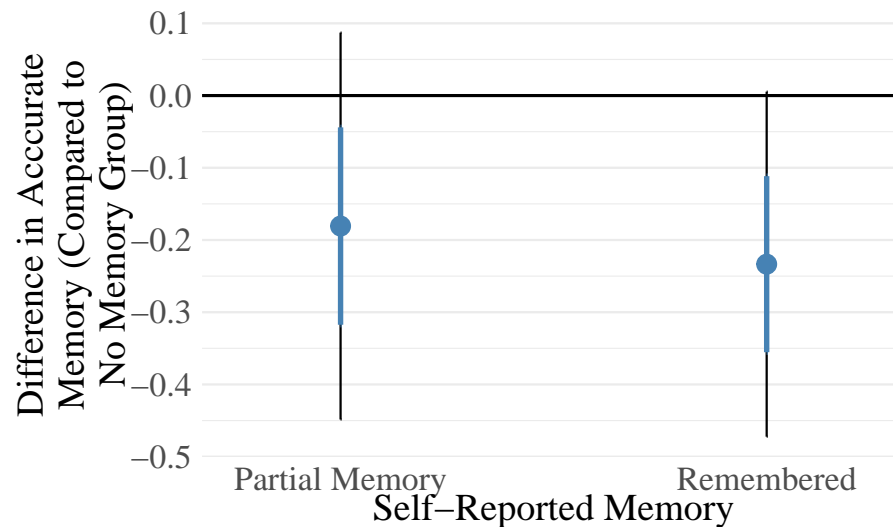


Figure A2: Experimental Platform Validation: No memory condition

A3 Priming

Figure A3 displays the difference in the proportion of respondents choosing Candidate A between the control group and each treatment group, separately for Q1a and Q1b. The priming treatments shifted candidate choice on individual questions, though in different directions across Q1a and Q1b. On Q1a—the first conjoint question, asked immediately after the prime—the age prime produced the largest shift, with the control group choosing Candidate A 11.3 percentage points more than the age treatment group ($p = 0.019$). The personal qualities prime showed a similar though marginally significant effect (9.1 pp, $p = 0.065$). The race and gender primes had no detectable effect on Q1a.

On Q1b—the repeated conjoint question, asked after a second exposure to the prime—the pattern reversed. The gender prime produced the largest shift, with the control group choosing Candidate A 10.8 percentage points more than the gender treatment group ($p = 0.028$). The age and personal qualities effects, which were present on Q1a, attenuated to near zero on Q1b. The race prime showed a moderate but nonsignificant difference (6.9

pp, $p = 0.159$).

These results suggest that the primes did introduce bias into individual candidate choices relative to the control group, but the bias was question-specific—different primes affected Q1a and Q1b differentially. Importantly, despite these shifts in the level of candidate preference, none of the primes produced a significant change in instability, consistent with the interpretation that priming shifts the direction of responses but does not reduce the stochastic component of decision-making.

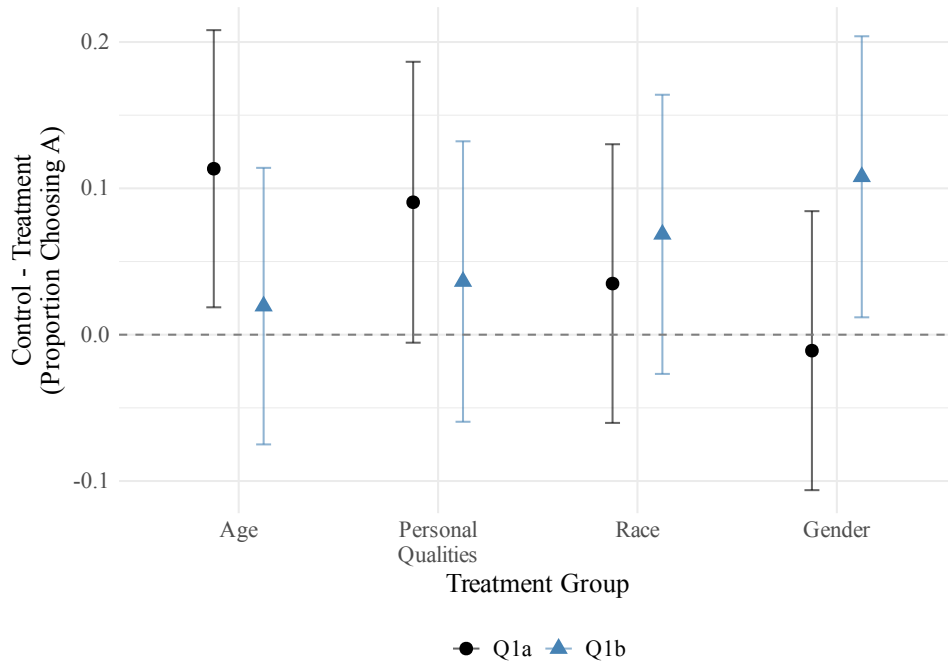


Figure A3: Effect of priming on candidate choice (Control – Treatment), separately for Q1a and Q1b. Primes shift the level of candidate preference on individual questions, but different primes affect Q1a and Q1b differentially (Survey 58; $n = 1,037$).

A4 Response Option Divergence and Approximate Indifference

Figure A4 plots response instability (vertically) by the number of shared attributes between the two conjoint candidates, where we use more shared attributes as a measure of less divergence between the answer options and, likely greater respondent indifference between them. Consistent with random utility models (Section 5.1), instability increases as the response options (Candidates) become more similar: respondents who received

candidates differing on all attributes exhibited approximately 20% instability, while those with nearly identical candidates were more unstable. This positive relationship confirms that indifference may be part of the story—when the deterministic components of utility for the two options are closer together, the stochastic component is more likely to tip the decision in different directions across repeated presentations. However, the starting point of instability (with maximum divergence) means that indifference alone fails to explain a large portion of the observed instability. That is, even among respondents who were asked about candidates that differed on every attribute—party, race, gender, age, education, and personal qualities—one in five respondents still changed their answer. Random utility models predict that instability should approach zero as the choice becomes increasingly lopsided, yet the data show a large, persistent baseline of instability that cannot be attributed to similarity between the options. Some additional mechanism must be generating response instability independent of the strength of the respondent’s preference. See also Clayton et al. (2025).

A5 Failed Attempts to Experimentally Control Mind Wandering

We attempted to manipulate mind wandering in several ways across four surveys.

In Survey 5 ($n = 154$), two treatment groups received prompts intended to increase cognitive distraction—and thus, increase mind wandering—during the survey. The first group was asked:

“Please write down the four most important things that you need to do in the next week.”

While this prompt may have briefly diverted attention, it may have allowed respondents to subsequently refocus on the task. To sustain distraction over a longer period, a second treatment group received a modified prompt:

“At the end of the survey, you will have 30 seconds to write down the four most important things that you need to do in the next week. It is important to our research that you get these four right. You will have the length of the survey (about 5 minutes) to think

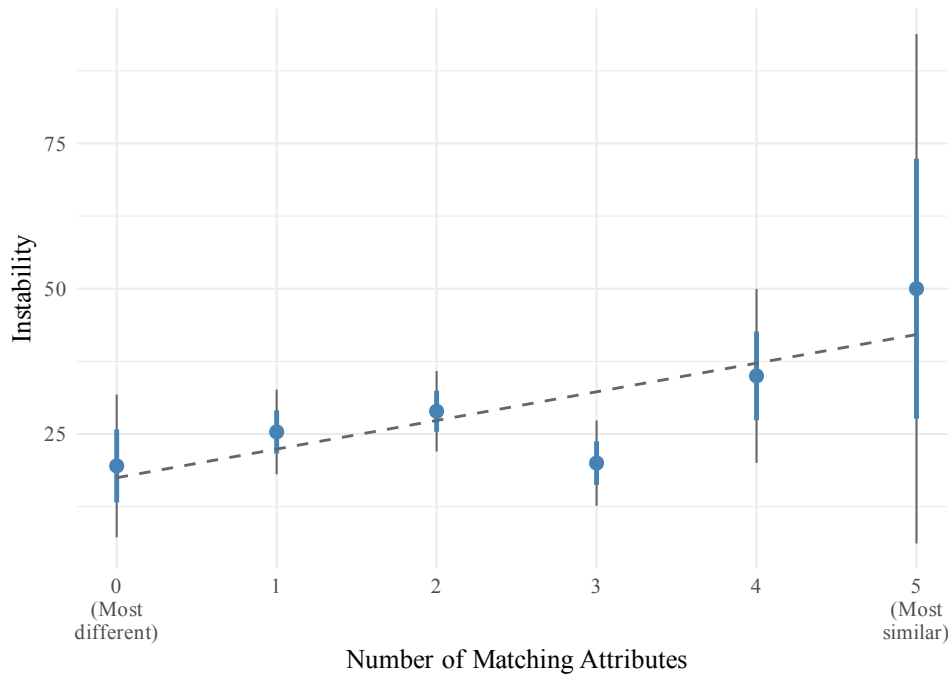


Figure A4: Instability by candidate similarity. The x-axis counts the number of shared attributes (out of six) between the two conjoint candidates; higher values indicate more similar — and thus harder to distinguish — candidates. Instability increases with similarity, as would be expected, but remains substantial ($\approx 20\%$) even when candidates differ on every attribute (Survey 57; $n = 506$).

of what to include on your list.”

This version was designed to keep respondents cognitively engaged with the task throughout the survey, thereby increasing the potential for mind wandering during the intervening questions. However, as shown in Figure A5, neither treatment group showed a statistically significant difference in reported mind wandering relative to the control group, which received no prompt.

In Survey 17 ($n = 201$), the treatment group was given the following prompt:

“Everyone comes to a new task (like this survey) with things on their minds, such as their partner, job, school, health, children, friends, money, or other concerns. What is the one thing that most preoccupies you at the moment? Please tell us about it — take your time and elaborate!”

We hypothesized that asking respondents to reflect on a current preoccupation would cause continued cognitive distraction during subsequent questions. The control group did not receive this prompt. Mind wandering was measured after Q1a. As shown in Figure

A5, there was no significant difference in mind wandering between the treatment and control groups.

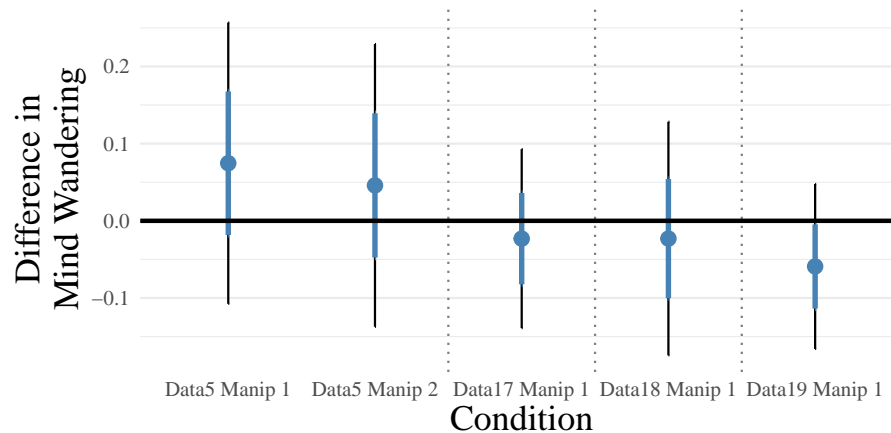


Figure A5: Null Effects of Attempted Mind Wandering Manipulations

Survey 18 tested a different approach, aimed at reducing mind wandering by encouraging respondents to acknowledge their distractions and consciously set them aside. The prompt read:

“Everyone comes to a new task (like this survey) with things on their minds, such as their partner, job, school, health, children, friends, money, or other concerns. What is the one thing that most preoccupies you at the moment? Please tell us about it — take your time and elaborate! It will help us, and maybe it will help you set it aside for a while.”

This intervention also failed to produce a statistically significant difference in mind wandering rates compared to the control group.

We further refined this intervention in Survey 19 by adding more structured prompts to encourage thoughtful engagement. The revised prompt was:

“Everyone comes to a new task (like this survey) with things on their minds, such as their partner, job, school, health, children, friends, money, or other concerns.

What is the one thing that most preoccupies you at the moment? Please tell us about it — take your time and elaborate! It will help us, and maybe it will help you set it aside for a while.

I am preoccupied with . . .

Because . . .”

Respondents in the treatment group were provided two response boxes to complete the thought. Despite this added structure, the manipulation did not produce a significant change in mind wandering relative to the control group.

Across all four studies, none of our experimental manipulations produced a reliable effect on self-reported mind wandering. Perhaps others can develop reliable methods of inducing sufficiently high levels of mind wandering to be useful, but we have not discovered a way to do this.

A6 Preoccupation Predicts Instability on Non-Conjoint Questions

In Survey 30, we examined whether respondents' reported preoccupation was associated with their answers to a set of non-conjoint items. After the preoccupation question, respondents completed a series of abortion attitude questions drawn from the General Social Survey. Each item asked: "Do you think it should be possible for a pregnant woman to obtain a legal abortion if . . . ," followed by one of the following conditions: (1) there is a strong chance of serious defect in the baby; (2) she is married and does not want any more children; (3) the family has a very low income and cannot afford any more children; (4) she is not married and does not want to marry the man; (5) the woman wants it for any reason; (6) the woman's own health is seriously endangered by the pregnancy; or (7) she became pregnant as a result of rape.

The first abortion item served as a "burn-in" question. Respondents then answered four abortion items in random order, completed the mind-wandering question, answered the remaining two abortion items, and finally answered the same four-item set again in the same random order.

As Figure A6 shows, the effect of preoccupation on instability was positive across all four questions.

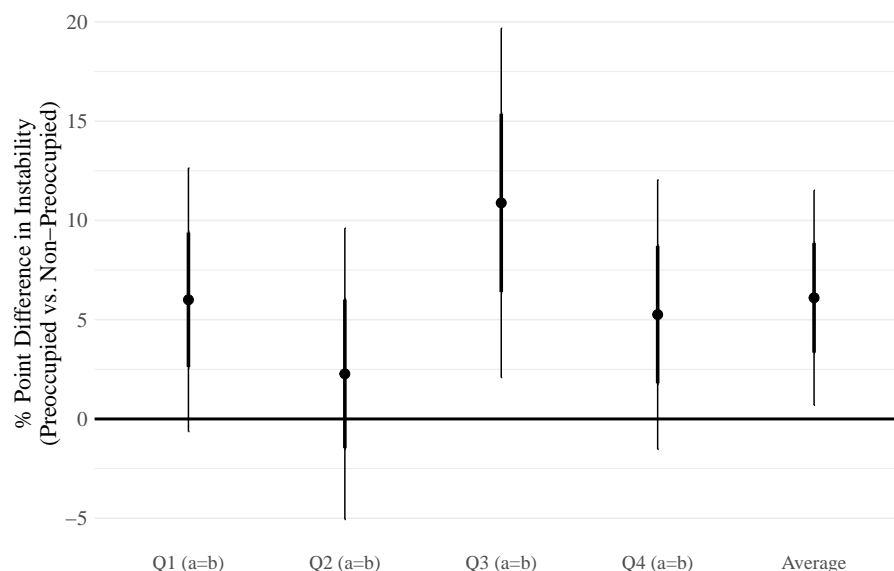


Figure A6: Preoccupation Predicts Instability on Non-Conjoint Questions

A7 Eye Tracking Data Collections

We ran three surveys using webcam eye tracking to improve our understanding of response instability by capturing participants' gaze paths as they choose between hypothetical political candidates.

We carried out the experiments using the webcam-based eye-tracking model WebGazer (Papoutsaki, 2015). WebGazer predicts participants' gaze locations using machine learning and operates successfully through commonly used commercial-grade webcams, eliminating the need for external hardware. WebGazer predicts gaze locations by identifying the participant's ocular regions through a calibration process in which participants look at dots on the screen. A pixel grid of the participant's eyes in greyscale is generated, and a ridge regression model is used to map this detailed eye representation onto the point coordinates. In order to preserve the model's generalizability, a term is added to the equation to penalize overfitting. During the task itself, WebGazer can sample at a maximum rate of about 60Hz (about every 16ms), but the actual frequency of collected gaze points depends on the participant's webcam frame rate.

We originally sought to use WebGazer through Gorilla, an online experiment host that has WebGazer functionality built-in (Survey 54). However, we discovered that Go-

rilla’s normalized coordinate data appeared to be systematically compressed, a finding confirmed but—to our knowledge—still unaddressed by Gorilla.¹ The biases in Gorilla’s algorithm were large enough to invalidate our use of its platform. Additionally, Gorilla’s built-in calibration procedure allows only for gaze-based, not click-based, calibration. Gaze-based calibration is accurate only if participants actually look at the dots when instructed. While the same is true for click-based calibration, prior research confirms that gaze position and cursor position are highly correlated (Chen, Anderson, and Sohn, 2001; Huang, White, and Buscher, 2012). We thus judge click-base calibration to be the better calibration modality for webcam eye tracking.

Therefore, in Surveys 55 and 56, we ran WebGazer directly through jsPsych 7.0 rather than through Gorilla. jsPsych is a JavaScript library that is commonly used for behavioral experiments and allows for direct control over WebGazer’s calibration and validation procedures. Participants were recruited through Prolific, who were required to be fluent English speakers, living in the United States, and using a webcam-equipped computer of sufficient size to display the conjoint task (at least 13 inches). Respondents were directed to the experiment via a link to Cognition.run, an environment optimized for running jsPsych experiments online. After finishing the survey, participants were directed back to Prolific for compensation.

WebGazer’s calibration and validation procedures are highly customizable. We implemented a click-based calibration phase before the first conjoint screen, with seven click points positioned according to the coordinates of salient elements of the conjoint screen: the Candidate A and B column headers, the third row in the central (attribute identification) column, the fifth rows in Candidate A and B’s columns, and the two candidate choice buttons. Participants were instructed to click on each point five times while looking at them. Next, participants saw a validation screen, which allows for judgment of the accuracy of the calibration results. Dots appeared consecutively in the same positions

¹An email from Gorilla’s support team said, “The development team has continued to investigate the issue you raised and believe we may have identified a minor discrepancy in how the normalised coordinates are being calculated. We still need additional time to test this thoroughly; however, if our understanding is correct, the plotted coordinates may be appearing slightly more centred [sic] on the x-axis than expected” (Gorilla Support, personal communications, 10 Dec 2025).

as the calibration points, and participants were instructed to look at each point but, this time, not to click on them. Each dot flashed for two seconds, with a one-second saccade window to allow participants to shift their gaze.

To ensure that we fairly compensated participants for their time even if the calibration/validation phases did not work for them, we allowed all participants to finish the survey regardless of their validation score. But in our analyses, we excluded participants with validation scores below 25 percent (i.e., where fewer than 25 percent of recorded gaze points during the validation phase fell within 200 pixels of the target point presented on the screen). This 200-pixel radius is consistent with prior WebGazer studies (Yang and Krajbich, 2021; Vos, Minor, and Ramchand, 2022) and arises from Yang and Krajbich (2021)'s finding that the average distance between true and measured gaze locations when using WebGazer is about 200 pixels.

The conjoint task screen was designed slightly differently from our prior surveys to improve salient gaze inference from the eye tracking data. Each candidate's information was presented on one side of a central attribute identification column rather than both being to the right of it, as was typical in our other surveys. Given the typical fixation offset that occurs using webcam eye tracking, this placement allowed us to more clearly distinguish which candidate a participant was looking at. Candidate A and Candidate B choice buttons were located under their respective columns.

The conjoint task was presented in a fixed 1200x800-pixel frame centered in participants' browsers. Because participants' displays vary in size and resolution, the table was not rendered at same absolute size across participants, nor did it appear at identical absolute screen coordinates across participants. WebGazer records the pixel coordinates of each stimulus element on each participant's unique screen [i.e., the top attribute for Candidate A might fall at (700, 110) for one participant but at (600, 90) for another], while maintaining the same 1200x800 frame on that participant's display. Webgazer's (x,y) gaze coordinates for each participant are likewise expressed in the participant's screen coordinate system. As a result, gaze and stimulus coordinates share the same reference frame and are directly aligned in the data (but differ across participants).

A8 Proposed Norms for Ethical Survey Research

Our research design was declared “exempt” by our university Institutional Review Boards and satisfies the requirements adopted by the American Political Science Association’s “Principles and Guidance for Human Subjects Research.” For example, we obtained voluntary and informed consent from participants, did not use deception, ensured that respondents understood they were participating in a research study, offered them ways to contact us if they had any objections (no one objected), collected no personally identifiable information, among others. We also adopted several other ethical rules beyond those required, which we now describe.

Survey research is a massive, multi-sector, multi-billion dollar enterprise with large and increasing numbers of surveys being conducted by researchers, universities, other nonprofits, and commercial entities. This activity is creating many new opportunities for individual scholars, but may have led to a classic tragedy of the commons. Most obviously, oversampling has led to dramatically declining response rates (Stedman et al., 2019). In addition, the way we pay and incent respondents has changed the survey situation for some, from the Gricean maxims of ordinary trustworthy conversations, to required attention checks and attempts to trick respondents into revealing their lack of attentiveness, inadvertently turning some questions from the attempts at unobtrusive measurement into unintentional causal treatments that induce bias in the quantities of interest (Hauser and Schwarz, 2015).

These issues are particularly relevant to results of our research which provide far more targeted ways of identifying and filtering respondents. Targeting can be used for good, but it can also create problems for other researchers, and so might be used to further trample the commons if we are not careful. Thus, to preserve the ability of future researchers to also learn from survey respondents, we have imposed specific rules on our own survey work. We hope others will consider following these rules, or to at least will think about and explain why different rules might be more appropriate for their work. Our four rules, designed to generate sincere survey responses and to not affect these respondents in future surveys by others, are as follows:

1. We allow every person who signs up for a survey to finish it, even if our plan is to filter out the remaining answers from that respondent.
2. We pay every participant the full amount, regardless of their responses to any survey question, including questions used for filters.
3. We use the smallest number of respondents necessary to estimate our quantities of interest with needed uncertainty levels. (This can be done via sequential selection, power calculations, or confidence interval size estimates based on simulated data.)
4. We try to make all questions trustworthy so that the respondent does not feel like they are being tricked, or that participation or earning money for participation needs to be gamed such as via certain types of attention check questions.

A9 Data collections

We describe here all the data we collected for this project, including surveys that did not make it into our paper, as well as where evidence we collected convinced us that our early hypotheses were wrong. From the outset, we sought to understand what drives survey response instability. While most existing work focused on instability across waves of panel studies (risking bias due to attrition), few had looked within a single survey or tried to manipulate its causes experimentally. Mind wandering and preoccupation had not (to our knowledge) been linked to survey instability. We suspected mind wandering might play a role and introduced a measure for it in Survey 2. We do not consider preoccupation as a potential factor until Survey 22. And because we weren't sure how best to measure or experimentally manipulate these constructs, many of the studies we ultimately didn't use reflect early, unsuccessful attempts to do just that. (**Bold** is used for treatment and control groups for expository clarity.)

This is how the surveys are described:

- S = Socioeconomic descriptors
- B = Burn-in conjoint question
- P = Preoccupation question
- Q1a = First conjoint question, to test instability
- D = Distractor conjoint question
- Q1b = Repeated conjoint question, to test instability (same as Q1a, usually with the order of the candidates reversed)

- M = Mind wandering probe
- T = Other questions
- We also define “[Follow-up]” in Survey 3 and use it subsequently

1. Survey 1 (n = 298)

- **Treatment Group 1:** S Q1a D1 D2 D3 D4 Q1b
- **Treatment Group 2:** S B1 B2 Q1a D1 D2 D3 D4 Q1b
- **Treatment Group 3:** S T Q1a D1 D2 D3 D4 Q1b
- **Treatment Group 4:** S T B1 B2 Q1a D1 D2 D3 D4 Q1b
- *Use: Validation of survey platform (number of burn-in questions); investigate influence of attention checks*

2. Survey 2 (n = 208)

- **Treatment Group 1:** S B1 B2 B3 Q1a Q1b M1
- **Treatment Group 2:** S B1 B2 Q1a D1 Q1b M1
- **Treatment Group 3:** S B1 Q1a D1 D2 Q1b M1
- **Treatment Group 4:** S Q1a D1 D2 D3 Q1b M1
- *Use: Validation of survey platform (number of distractor questions); mind wandering analyses*

3. Survey 3 (n = 776)

- All respondents receive: S Q1a D1 D2 D3 Q1b. Then, if they were unstable:
- **Treatment Group 1:** T1 (Candidate A and Candidate B’s info in T1 shown in the same order as Q1a)
- **Treatment Group 2:** D4 D5 T1 (Candidate A and Candidate B’s info in T1 shown in the same order as Q1a)
- **Treatment Group 3:** T2 (Candidate A and Candidate B’s info in T2 shown in the same order as Q1b)
- **Treatment Group 4:** D4 D5 T2 (Candidate A and Candidate B’s info in T2 shown in the same order as Q1b),
where T1 and T2 = “We noticed that your answers to two of the questions were different, even though the candidates that you saw were equivalent between the two questions.

This is not a problem!

We want to understand better what leads people to answer the same question differently. Below, the two candidates in these questions are described again.

After thinking about it some more, which candidate do you prefer?

Why do you think you answered the two questions differently? Please explain below. ”

These questions are referred to from here on out as [Follow-up].

- *Use: Examine whether unstable respondents are sticking with their Q1b answer when asked a Q1c.*

4. Survey 4 (n = 597)

- **Treatment Group 1:** S Q1a D1 D2 Q1b
- **Treatment Group 2:** S Q1a M1 D1 D2 Q1b
- **Treatment Group 3:** S Q1a D1 D2 Q1b M1
- **Treatment Group 4:** S Q1a M1 D1 D2 Q1b M2
- *Use: Mind wandering analyses*

5. Survey 5 (n = 154)

- **Control Group:** S Q1a M1 D1 D2 Q1b M2
- **Treatment Group 1:** S T1 Q1a M1 D1 D2 Q1b M2
- **Treatment Group 2:** S T2 Q1a M1 D1 D2 Q1b M2,
where T1 = “Please write down the four most important things that you need to do in the next week” and
T2 = “At the end of the survey, you will have 30 seconds to write down the four most important things that you need to do in the next week. It is important to our research that you get these four right. You will have the length of the survey (about 5 minutes) to think of what to include on your list.”
- *Use: Unsuccessful attempt to manipulate mind wandering*

6. Survey 6 (n=173)

- **Control Group:** S Q1a M1 D1 D2 D3 D4 Q1b M2 D5 D6 [Follow-up] M3
- **Treatment Group 1:** S T1 Q1a M1 D1 D2 D3 D4 Q1b M2 D5 D6 [Follow-up] M3
- **Treatment Group 2:** S T2 Q1a M1 D1 D2 D3 D4 Q1b M2 D5 D6 [Follow-up] M3 T3, where
- T1 = “In this survey, you will answer some questions, each asking you to choose between a pair of candidates. These questions usually take about 5 minutes in total. First, please write down the 4 most important tasks you need to do in the next week (including personal, professional, or anything else, as long as they are important to you). Getting these 4 tasks right is crucial for our research, and so please think hard about which tasks belong on the list.”
T2 = “In this survey, you will answer some questions, each asking you to choose between a pair of candidates. These questions usually take about 5 minutes in total. Finally, we will give you a maximum of 30 seconds to write down the 4 most important tasks you need to do in the next week (including personal, professional, or anything else, as long as they are important to you). Getting these 4 tasks right is crucial for our research, and so you might think about what belongs on the list while you’re answering the previous questions.”
T3 = “Please write down the four most important things that you need to do in the next week.”

- Reason not used: No difference in mind wandering (M1 or M2) between control group and either treatment group

7. Survey 7 (n = 205)

- **Treatment Group 1:** S Q1a D1 D2 D3 D4 Q1b (all questions from Q1a on allocated a maximum of 10 seconds) D5 D6 [Follow up]
- **Treatment Group 2:** S Q1a D1 D2 D3 D4 Q1b (all questions Q1a on allocated a maximum of 20 seconds) D5 D6 [Follow up]
- **Treatment Group 3:** S Q1a D1 D2 D3 D4 Q1b (all questions Q1a on allocated a maximum of 30 seconds) D5 D6 [Follow up]
- **Treatment Group 4:** S Q1a D1 D2 D3 D4 Q1b (all questions Q1a on allocated a maximum of 40 seconds) D5 D6 [Follow up]
- Reason not used: There is a notable decrease in instability as the time respondents were forced to spend on questions increased. That said, the times are likely too long to be spent by anyone thinking about the question.

8. Survey 8 (n = 202)

- **Treatment Group 1:** S B1 Q1a D1 D2 D3 Q1b D4 D5 [follow-up], where Q1a and Q1b were limited to 10 seconds
- **Treatment Group 2:** S B1 Q1a D1 D2 D3 Q1b D4 D5 [follow-up], where Q1a and Q1b were limited to 20 seconds
- **Treatment Group 3:** S B1 Q1a D1 D2 D3 Q1b D4 D5 [follow-up], where Q1a and Q1b were limited to 30 seconds
- **Treatment Group 4:** S B1 Q1a D1 D2 D3 Q1b D4 D5 [follow-up], where Q1a and Q1b were limited to 40 seconds
- Reason not used: When participants are forced to spend the allotted amounts of time on only Q1a and Q1b, instability decreases but not significantly; we cannot force participants to focus on the question for a certain amount of time.

9. Survey 9 (n=110)

- **Control Group:** S Q1a D1 D2 D3 Q1b D4 D5 [Follow-up]
- **Treatment Group 1:** S Q1a T1a T2a D1 T1b T2b D2 T1c T2c D3 T1d T2d Q1b T1e T2e D4 T1f T2f D5 T1g T2g [Follow-up], where T1 = “Why do you prefer this candidate? (Please write at least one or two full sentences explaining your choice so that we can understand)” and T2 = “Some people who have taken this survey chose the other candidate. Say that you were to try to persuade them that your favored candidate was the better choice. What would your argument be?”
- Reason not used: The manipulation failed to move instability.

10. Survey 10 (n = 106)

- **S Q1a D1 D2 T1a D3 T1b T2a Q1b T2b T3a T3b [Follow-up]**
 where T1 = “You may or may not have seen the following candidates in a previous question. Do you think you have?” (Answer options: “I have seen these candidates in a previous question,” “I am unsure if I have seen these candidates in a previous question,” and “I have not seen these candidates in a previous question”) and
 T2 = “Actually, you did not see these candidates in a previous question,” “You did not see these candidates in a previous question,” and “That’s correct. You did not see these candidates in a previous question.” T2 options corresponded to answers to T1, respectively.
 T3a = “Do you remember which candidate you preferred?” (Answer options: “I remember,” “I do not remember but could probably guess,” and “I do not remember”);
 T3b = “Try to remember: which candidate did you prefer in the previous question?” (Answer options: Candidate A, Candidate B)
- *Use: Validation of survey platform (self-reported memory of Q1a does not make a difference in memory’s accuracy)*

11. Survey 11 (n = 322)

- **Control Group:** S Q1a D1 D2 D3 Q1b D4 D5 [Follow-up]
- **Treatment Group 1:** S T1a Q1a D1 D2 D3 T1b Q1b D4 D5, where T1 is a video on **mindfulness**.
- **Treatment Group 2:** S T2a Q1a D1 D2 D3 T2b Q1b D4 D5 [Follow-up], where T2 is a **Tom & Jerry video**.
- Reason not used: Failed to manipulate instability.

12. Survey 12 (n = 553)

- **Control Group:** S B1 Q1a D1 D2 D3 Q1b D4 D5 [Follow-up]
- **Treatment Group 1:** S B1 Q1a D1 D2 D3 Q1b D4 D5 [Follow-up] (Q1b allowed a maximum of two seconds)
- **Treatment Group 2:** S B1 Q1a D1 D2 D3 Q1b D4 D5 [Follow-up] (Q1b allowed a maximum of four seconds)
- **Treatment Group 3:** S B1 Q1a D1 D2 D3 Q1b D4 D5 [Follow-up] (Q1b allowed a maximum of six seconds)
- **Treatment Group 4:** S B1 Q1a D1 D2 D3 Q1b D4 D5 [Follow-up] (Q1b allowed a maximum of eight seconds)
- *Use: Experimental manipulation of time on task, to demonstrate its relationship with instability*

13. Survey 13 (n = 210)

- **S B1 Q1a M1a (if M1a = 1, T1a) D1 D2 D3 Q1b M1b (if M1b = 1, T1b) D4 D5 [Follow-up]** where T1 = “Most people get distracted at some point when

taking surveys by other important things that they need to do. How long would you say you spent thinking about other things, out of the time you spent on that page?”

- *Use: Evidence that time on task effects instability; mind wandering analyses*

14. Survey 14 (n = 118 in control group)

- **Control Group:** S B1 Q1a M1a D1 D2 D3 Q1b M1b
- **Treatment Group:** S B1 Q1a M1a D1 D2 D3 Q1b M1b (four seconds on all questions)
- *Use: Mind wandering analyses (control group only)*

15. Survey 15

- Reason not used: mistake in coding up experiment

16. Survey 16 (n = 641)

- **S B1 P Q1a D1 D2 D3 Q1b M1**
- *Use: Evidence that time on task causes instability; preoccupation analyses; mind wandering analyses; individual characteristics analyses*

17. Survey 17 (n = 201)

- **Control Group:** S Q1a M1 D1 D2 D3 Q1b M2
- **Treatment Group:** S T1 Q1a M1 D1 D2 D3 Q1b M2
where T1 = “Everyone comes to a new task (like this survey) with things on their minds, such as their partner, job, school, health, children, friends, money, or other concerns. What is the one thing that most preoccupies you at the moment? Please tell us about it – take your time and elaborate!”
- *Use: evidence that time on task causes instability; additionally, an unsuccessful attempt to manipulate mind wandering; mind wandering analyses*

18. Survey 18 (n = 103)

- **Control Group:** S B1 Q1a D1 D2 D3 Q1b M1
- **Treatment Group:** S B1 T1 Q1a D1 D2 D3 Q1b M1
where T1 = “Everyone comes to a new task (like this survey) with things on their minds, such as their partner, job, school, health, children, friends, money, or other concerns. What is the one thing that most preoccupies you at the moment? Please tell us about it – take your time and elaborate! It will help us, and maybe it will help you set it aside for awhile.”
- *Use: evidence that time on task causes instability; an unsuccessful attempt to manipulate mind wandering; mind wandering analyses*

19. Survey 19 (n = 254)

- **Control Group:** S Q1a M1 D1 D2 D3 Q1b M2

- **Treatment Group:** S T1 Q1a M1 D1 D2 D3 Q1b M2
 where T1 = “Everyone comes to a new task (like this survey) with things on their minds, such as their partner, job, school, health, children, friends, money, or other concerns. What is the one thing that most preoccupies you at the moment? Please tell us about it – take your time and elaborate! It will help us, and maybe it will help you set it aside for awhile.
 I am preoccupied with . . .
 Because . . .”
- *Use: evidence that time on task causes instability; additionally, an unsuccessful attempt to manipulate mind wandering*

20. Survey 20 (n = 1050)

- **S B1 B2 B3**, followed by:
- IF $\min(\text{time on B1, B2, B3}) < 5 \text{ sec}$, then half of respondents see
 - **Treatment Group 1:** T1 Q1a D1 D2 D3 Q1b,
 where T1 = “Everyone comes to a new task (like this survey) with things on their minds, such as their partner, job, school, health, children, friends, money, or other concerns. What is the one thing that most preoccupies you at the moment? Please tell us about it – take your time and elaborate! It will help us, and maybe it will help you set it aside for awhile.”
 The other half **Control Group 1** sees Q1a D1 D2 D3 Q1b.
- ELSE half of the remaining respondents see one of the following:
 - **Treatment Group 2:** Q1a D1 D2 D3 T2 Q1b T3,
 where T2 = “Please do your best to memorize the following six digit number. Survey it, or repeat it to yourself, but do not write it down. Next, we will ask you one more candidate question. After that, we will ask you to remind us what the number was. The number is: 527364.”
 T3 = “Please write the number that you saw before as well as you can remember it.”
 - **Control Group 2:** Q1a D1 D2 D3 Q1b
- Reason not used: No effect on instability

21. Survey 21 (n = 1057)

- **S B1 B2 B3**, followed by:
- IF $\min(\text{time on B1, B2, B3}) < 7 \text{ sec}$, then
 - **Treatment Group 1:** T1 Q1a D1 D2 D3 Q1b,
 where T1 = “We have 5 more candidate questions for you. We know that these are complicated! And we also know that everyone comes to a new task (like this survey) with things on their mind that make it hard to focus — concerns such as their partner, job, school, health, children, friends, or money.
 But it would really help our research if you would direct your total focus to the next five questions. Can you give each of the following 5 questions

your full attention?” (Answer options: I can direct my full attention to these questions; I just can't focus that much right now)

– **Control Group 1:** Q1a D1 D2 D3 Q1b

• ELSE

– T2 Q1a T3 D1 D2 D3 T4 Q1b T5,

where T2 = “Please do your best to memorize the following six digit number. Survey it, or repeat it to yourself, but do not write it down. Next, we will ask you another candidate question. After that, we will ask you what the number was. The number is: 472935. ”

T3 = “Please write the number that you saw before as well as you can remember it.”

T4 = “One more time, please do your best to memorize the following six digit number. Survey it, or repeat it to yourself, but do not write it down. Next, we will ask you one more candidate question. After that, we will ask you what the number was. The number is: 527364. ” T5 = “Please write the number that you saw before as well as you can remember it.”

• Reason not used: No effect on instability

22. Survey 22 (n = 344)

• **Control Group:** S B1 B2 B3 P Q1a D1 D2 D3 Q1b

• **Treatment Group:** S B1 B2 B3 P Q1a D1 D2 D3 Q1b (with all questions past P limited to four seconds)

• Reason not used: Not a useful treatment for distinguishing the effect of preoccupation or mind wandering.

23. Survey 23 (n = 145)

• **Control Group:** S B1 P Q1a D1 D2 D3 Q1b

• **Treatment Group:** S B1 P Q1a D1 D2 D3 Q1b (all questions past P limited to four seconds)

• This Survey was the same as Survey 22 in order to validate the result.

• Reason not used: Not a useful treatment for distinguishing the effect of preoccupation or mind wandering on instability.

24. Survey 24 (n = 129)

• **Control Group:** S P B1 Q1a M1 D1 D2 D3 Q1b M2

• **Treatment Group:** S P B1 Q1a M1 D1 D2 D3 Q1b M2 (all Qs and Ds limited to seven seconds)

• Reason not used: Not a useful treatment for distinguishing the effects of preoccupation or mind wandering on instability

25. Survey 25 (n = 305)

- **Treatment Group 1:** S P B1 Q1a M1 D1 D2 D3 Q1b M2 (10 seconds on all after B)
- **Treatment Group 2:** S P B1 Q1a M1 D1 D2 D3 Q1b M2 (25 seconds on all after B)
- **Treatment Group 3:** S P B1 Q1a M1 D1 D2 Q1b M2 (40 seconds on all after B)
- Reason not used: Manipulation failed to affect instability

26. Survey 26 (n = 227)

- **P1 B1 Q1a P2 D1 D2 Q1b M1 T1 T2,**
 where T1 = “Below are the two candidates that you just saw. On the last question, you chose Candidate [A/B].
 Please have another look at these candidates and decide again which you prefer. You may keep the same choice or change it, perhaps because you chose the wrong candidate the first time, changed your reasoning, decided to weigh the features differently, or maybe just like them about the same.
 Whatever your reasoning, which candidate do you now prefer?” and
 T2 = “Why did you decide to choose the other candidate this time?” (Answer options: I accidentally pressed the wrong button the first time; I changed my mind after thinking more about it; I liked the candidates about the same; Other)
- *Use: evidence that time on task causes instability; preoccupation analyses; mind wandering analyses; individual characteristics analyses*

27. Survey 27 (n = 374)

- **Control Group:** P S B1 Q1a D1 D2 Q1b Q1c,
 where Q1c = “Below are the two candidates that you just saw. On that question, you chose Candidate [A/B].
 Please have another look at these candidates and decide again which you prefer. You may keep the same choice or change it, perhaps because you chose the wrong candidate the first time, changed your reasoning, decided to weigh the features differently, or maybe just like them about the same.
 Whatever your reasoning, which candidate do you now prefer?”
- **Treatment Group 1:** P S B1 Q1a T1 Q2 Q3 Q1b Q1c,
 where T1 = “How sure are you about your previous answer?” (Answer options: Very sure; fairly sure; slightly sure; not at all sure)
- **Treatment Group 2:** P S B1 Q1a T Q2 Q3 Q1b Q1c,
 where answer options to conjoint questions = Candidate A, Candidate B, and I can’t choose.
- Reason not used: First treatment did not impact instability. Second treatment did, but cannot differentiate if this was caused by having three answer options or the specific “I can’t choose” answer option.

28. Survey 28 (n = 160; control group n = 79)

- **Control Group:** P S B1 Q1a D1 D2 Q1b T1, where T1 = “How sure are you about your previous answer?” (Answer options: Very sure, fairly sure, slightly sure, not at all sure)
- **Treatment Group:** P S B1 Q1a D1 D2 Q1b T1 (with three answer options, “I can’t choose” being the third)
- *Use: Preoccupation analyses (control group only); individual characteristics analyses*

29. Survey 29

- NA; accidentally mislabeled Survey 29 as Survey 30

30. Survey 30 (n = 237)

- **S P B1 T1a T2a T3a T4a (order of Ts randomized) M1 T5 T6 T1b T2b T3b T4b b(order of T1b - T4b randomized),** where B1 = “Do you think it should be possible for a pregnant woman to obtain a legal abortion if there is a strong chance of serious defect in the baby?”; T1 = “Do you think it should be possible for a pregnant woman to obtain a legal abortion if she is married and does not want any more children?”; T2 = “Do you think it should be possible for a pregnant woman to obtain a legal abortion if the family has a very low income and cannot afford any more children?”; T3 = “Do you think it should be possible for a pregnant woman to obtain a legal abortion if she is not married and does not want to marry the man?”; T4 = “Do you think it should be possible for a pregnant woman to obtain a legal abortion if the woman wants it for any reason?; T5 = “Do you think it should be possible for a pregnant woman to obtain a legal abortion if the woman’s own health is seriously endangered by the pregnancy?”; T6 = “Do you think it should be possible for a pregnant woman to obtain a legal abortion if she became pregnant as a result of rape?”
- *Use: investigates whether preoccupation has the hypothesized effect on instability when using non-conjoint questions*

31. Survey 31 (n = 215)

- **S P T1 T2 T3 T4 M1,** where T = “You will now cast votes in an election for four offices. Two candidates, A and B, are running for each office. For each office, there is a three in four chance that Candidate A is better for you and the country, and a one in four chance that Candidate B is better. (Unfortunately, you don’t know anything else about the candidates or the offices.)” Then, they were asked to cast their vote for each office.
- *Use: Examine whether consistency on choices exists when preferences are held constant.*

32. Survey 32 (n = 220)

- **P T1 T2 T3 T4 M1**, where
T = You will now cast votes in an election for four offices. Two candidates, A and B, are running for each office. For each office, there is a one in four chance that Candidate A is better for you and the country, and a three in four chance that Candidate B is better. (Unfortunately, you don't know anything else about the candidates or the offices.)” Then, they were asked to cast their vote for each office.
- *Use: Examine whether consistency in choices exists when preferences are held constant, switching the preferred candidate from Candidate A to Candidate B*

33. Survey 33 (n = 269)

- **Treatment Group 0:** S P T1 T2 T3 T4, where T = “You will now cast votes in an election for four offices. Two candidates, A and B, are running for each office.
You will be given some information about the candidates but, unfortunately, will not know anything about the offices.” This was followed by a conjoint table listing, in each question, the same candidates.
- **Treatment Group 1:** S P T1 T2 T3 T4,
where T = “You will now cast votes in an election for four offices. Two candidates, A and B, are running for each office.
For each office, there is a one in four chance that Candidate A is better for you and the country, and a three in four chance that Candidate B is better.
(Unfortunately, you don't know anything else about the candidates or the offices.)” This was followed by four choices between Candidates A and B for offices 1 through 4.
- *Use: Demonstrates that when preferences are fixed vs. not fixed in the same conjoint setup, instability is greater when they are fixed. This is indirect evidence that when π is greater, choices' instability increase.*

34. Survey 34 (n = 221)

- **Treatment Group 1:** S B1 Q1a D1 D2 Q1b M1 T1,
where T1 = “Now, you will vote in a new type of voting system. In this system, every person gets ten votes. You can cast as many of those votes for Candidate A as you want to and as many for Candidate B as you want to. . . . Please indicate the number of your votes that you would like to cast for Candidate A and the number of your votes that you would like to cast for Candidate B. They must sum to 10.” The candidates shown were in the same order as the candidates in Q1a.
- **Treatment Group 2:** S B1 Q1a D1 D2 Q1b M1 T1,
where T1 is the same as above but with the candidates shown in the same order as they were in Q1b.
- *Use: Examine stochasticity in decision making*

35. Survey 35 (n = 114)

- **S T1 B1 Q1a T2 D1 D2 Q1b T3**,
where T1 = First attention check; T2 = Second attention check; T3 = Third attention check
- Reason not used: No control group included; also, need a larger sample as result of very few people getting attention checks incorrect (e.g., only eight answered two of them incorrectly)

36. Survey 36 (n = 200)

- **Control Group:** S P B1 Q1a M D1 D2 D3 Q1b
- **Treatment Group 1:** S P B1 Q1a M D1 D2 D3 Q1b (all conjoint questions worded confusingly, including instructions)
- *Use: Most extreme possible example of question wording affecting instability (e.g., “Which candidate do you not dislike the most?”)*

37. Survey 37 (n = 225)

- **S P T1 T2 T3 T4 T5a T5b T5c T5d M1**,
where T1 = “We’d like to ensure that you understand the following instructions in the same way as we do and so we’ll first ask you some questions about them before proceeding. You will next cast votes in an election for four offices. Two candidates, A and B, are running for each office. For each office, there is a one in four chance that Candidate A is better for you and the country, and a three in four chance that Candidate B is better. (Unfortunately, you don’t know anything else about the candidates or the offices.)”
T2 = “To make sure you understand the instructions on the previous page: What is the chance that Candidate A is better for you and the country?” (Answer options: 1 in 4; 2 in 4; 3 in 4; 4 in 4). If respondents answered incorrectly, they saw the following: “Oh, sorry, that’s not what we meant. Here are the instructions again, put a little differently: You will now choose between two candidates, A and B, for four offices in an election. Candidate A: there is a 1 in 4 chance that this candidate is better for you and the country. Candidate B: there is a 3 in 4 chance that this candidate is better. (Unfortunately, you don’t know anything else about the candidates or the offices.)” If they answered incorrectly again, they saw yet another wording of the same information. If, after that, they still answered incorrectly, they saw: “We meant to convey that there is a 1 in 4 chance that Candidate A will be better for you and for the country.”
T3 = “What is the chance that Candidate B is better for you and the country?” (Answer options: 1 in 4; 2 in 4; 3 in 4; 4 in 4) Respondents who answered this question incorrectly saw, “We meant to convey that there is a 3 in 4 chance that Candidate B will be better for you and for the country.”
T4: “And for how many offices will you cast a vote?” (Answer options: one; two; three; four; five; six) Respondents who answered this question incorrectly saw, “We meant to convey that there are four offices in which you will now cast a vote.”
T5a: “Please cast your vote for Office 1. (Remember: You will vote in four

offices. For each office, there is a 1 in 4 chance that Candidate A is better for you and the country, and a 3 in 4 chance that Candidate B is better for you and the country.)” (Answer options: Candidate A, Candidate B). T5b, T5c, and T5d were the same, with the office number changed.

- *Use: Decision making stochasticity.*

38. Survey 38 (n = 83)

- **S P T1 T2 T3 T4a T4b T4c T4d,**

where T1 = “We’d like to ensure that you understand the following instructions in the same way as we do and so we’ll first ask you some questions about them before proceeding. You will next cast votes in an election for four offices. Two candidates, A and B, are running for each office. For each office, there is a one in four chance that Candidate A is better for you and the country, and a three in four chance that Candidate B is better. (Unfortunately, you don’t know anything else about the candidates or the offices.)

To make sure you understand these instructions: What is the chance that Candidate A is better for you and the country?” (Open-ended response)

T2 = “What is the chance that Candidate B is better for you and the country?” (Open-ended response)

T3 = “Thanks. Now to make sure that we’re clear, here’s another version of the instructions: You will now choose between two candidates, A and B, for four offices in an election. Candidate A: there is a 1 in 4 chance that this candidate is better for you and the country. Candidate B: there is a 3 in 4 chance that this candidate is better. (Unfortunately, you don’t know anything else about the candidates or the offices.)

How many times will you choose between Candidate A and Candidate B?” (Open-ended response.)

T5a-d = “Please cast your vote for Office [1-4]. (Remember: You will vote in four offices. For each office, there is a 1 in 4 chance that Candidate A is better for you and the country, and a 3 in 4 chance that Candidate B is better for you and the country.)” (Answer options: Candidate A, Candidate B)

- *Use: Decision making stochasticity*

39. Survey 39 (n = 121)

- **Control Group:** S P B1 Q1a M1 D1 D2 D3 Q1b
- **Treatment Group:** S P B1 Q1a M1 D1 D2 D3 Q1b, with confusing wording of questions
- All conjoint attributes were held at prespecified levels for both the control and treatment group
- Reason not used: Manipulation failed to affect instability; perhaps our wording of the questions in the treatment condition was not weird enough.

40. Survey 40 (n = 106)

- **Control Group:** S T1 T2a T3 T4 T5 T6 T2b, where all Ts are all ANES questions, and T2 = the traditional ANES defense spending question, “Some people believe that we should spend much less money for defense. Suppose these people are at one end of a scale, at point 1. Others feel that defense spending should be greatly increased. Suppose these people are at the other end, at point 7. And, of course, some other people have opinions somewhere in between, at points 2, 3, 4, 5 or 6. Where would you place yourself on this scale, or haven’t you thought much about this?” (Answer options: 1. Spend much less money on defense, 2, 3, 4, 5, 6, 7. Spending should be greatly increased)
- **Treatment Group:** S T1 T2a/b T3 T4 T5 T6 T2c/d, where all Ts are all improved 2008 ANES questions, and T2 = the improved 2008 ANES defense spending question T2a/c: “Do you think that the government should spend more on national defense, less on national defense, or about the same on national defense as it does now?”
T2b/d: “Do you think the government should spend a lot more (less), somewhat more (less), or slightly more (less) than it does now?” (Answer options as described in the question.)
- *Use: Determine whether improved survey questions reduce instability*

41. Survey 41 (n = 119)

- **Control Group:** S P T1 T2a T3 T4 T5 T6 T2b M, where T = Traditional ANES questions. T2= “Some people think the government should provide fewer services even in areas such as health and education in order to reduce spending. Suppose these people are at one end of a scale, at point 1. Other people feel it is important for the government to provide many more services even if it means an increase in spending. Suppose these people are at the other end, at point 7. And, of course, some other people have opinions somewhere in between, at points 2, 3, 4, 5, or 6. Where would you place yourself on this scale, or haven’t you thought much about this?” (Answer options: 1. Government should provide many fewer services, 2, 3, 4, 5, 6, 7. Government should provide many more services)
- **Treatment Group:** S P T1 T2a T3 T4 T5 T6 T2b M, where improved question from 2008 ANES. T2 = “Do you think the government should provide more services than it does now, fewer services than it does now, or about the same number of services as it does now?” “Do you think the government should provide a lot more (fewer) services, somewhat more (fewer) services, or slightly more (fewer) services than it does now?” (Answer options as described in the question.)]
- *Use: determine whether improved survey questions reduce instability*

42. Survey 42 (n = 108)

- **Control Group:** S P T1a T2 T3 T4 T5 T6 T7 T1b M1, where T1a = Traditional ANES discussion of politics questions

- **Treatment Group:** S P T1a T2 T3 T4 T5 T6 T7 T1b M1, where T1 = Improved discussion of politics question [from 2008 ANES]: “During a typical week, how many days do you discuss politics with your family or friends?” (Zero days to seven days.)]
- *Use: Determine whether improved survey questions reduce instability*

43. Survey 43 (n = 120)

- **Control Group:** S P T1 T2a T3 T4 T5 T6 T7 T2b M1, where T2 = Traditional ANES medical insurance question
- **Treatment group:** S P T1 T2a T3 T4 T5 T6 T7 T2b M1, where T2 = Improved medical insurance question [from 2008 ANES: Q1: “Do you favor, oppose, or neither favor nor oppose the U.S. government paying for all necessary medical care for all Americans?” Q2: “Do you favor that a great deal, moderately, or a little?” (Answer options as described in the question.)]
- *Use: Determine whether improved survey questions reduce instability*

44. Survey 44 (n = 115)

- **Control Group:** S T1a Q1a M1 D1 D2 D3 Q1b
- **Treatment Group 1:** S T1a Q1a M1 D1 D2 D3 T1b Q1b, where T1a and T1b are the same candidates (shown in reverse order across the two questions)
- *Use: Investigate whether priming explains instability*

45. Survey 45 (n = 635)

- **Control Group:** S B1 B2 Q1a D1 D2 D3 Q1b
- **Treatment Group 1:** S B1 T3a T3b T3c B2 Q1a D1 D2 D3 Q1b
- **Treatment Group 2:** S/T1 (interspersed) B1 B2 T2 Q1a D1 D2 Q1b T4,
- where T3 = Conjoint attention checks and T1, T2, and T4 = instructional manipulation checks
- *Use: Validation of survey platform (attention check questions)*

46. Survey 46 (n = 224)

- **Control Group:** S Q1a D1 D2 D3 D4 Q1b
- **Treatment Group 1:** S B1 B2 Q1a D1 D2 D3 D4 Q1b
- **Treatment Group 2:** S B3 Q1a D1 D2 D3 D4 Q1b
- *Use: Investigate potential effects of burn-ins*

47. Survey 47 (n = 301)

- **Control Group:** S P B1 Q1a M1 D1 D2 D3 Q1b

- **Treatment Group:** S P B1 T1a M2 D1 D2 D3 T1b, where T = “For this election, say that you know that there is a three in four chance that Candidate A is better for you and the country, and a one in four chance that Candidate B is better. (Unfortunately, you don’t know anything else about the candidates.) Which candidate do you prefer?” (Candidate A and B have swapped probabilities in T1b)
- *Use: Investigate instability when $\pi \neq 1$*

48. Survey 48 (n = 278)

- **Control Group:** S P B1 Q1a M1 D1 D2 D3 Q1b
- **Treatment Group:** S P B1 T1a M2 D1 D2 D3 T1b, where T = “For this election, say that you know that there is a three in five chance that Candidate A is better for you and the country, and a two in five chance that Candidate B is better. (Unfortunately, you don’t know anything else about the candidates.) Which candidate do you prefer?” (Candidate A and B have swapped probabilities in T1b)
- *Use: Decision making stochasticity*

49. Survey 49 (n = 289)

- **Control Group:** S P B1 Q1a M1 D1 D2 D3 Q1b
- **Treatment Group:** S P B1 T1a M2 D1 D2 D3 T1b, where T = “For this election, say that you know that there is a nine in ten chance that Candidate A is better for you and the country, and a one in ten chance that Candidate B is better. (Unfortunately, you don’t know anything else about the candidates.) Which candidate do you prefer?” (Candidate A and B have swapped probabilities in T1b)
- *Use: Decision making stochasticity*

50. Survey 50 (n = 857)

- **Control Group:** S(including age and education) P B1 Q1a M1 D1 D2 D3 Q1b S(only age and education)
- **Treatment Group:** T1 S(including gender and education) P B1 Q1a M1 D1 D2 D3 Q1b S(only gender and education), where T = “How would you characterize yourself on the last survey you took?” (Answer options: I finished the survey as quickly as possible; I balanced speed and giving thoughtful, accurate answers; I answered thoughtfully and accurately, regardless of the time it took)
- *Use: Investigate instability on gender and education; investigate persona*

51. Survey 51 (n = 591)

- **Control Group:** S (including age and education) P B1 Q1a M1 D1 D2 D3 Q1b S (only age and education)

- **Treatment Group:** T1 S (including age and education) P B1 Q1a M1 D1 D2 D3 Q1b S (only age and education), where T = “How would you characterize yourself on the last survey you took?” (Answer options: I finished the survey as quickly as possible; I balanced speed and giving thoughtful, accurate answers; I answered thoughtfully and accurately, regardless of the time it took)
- *Use: get more data on instability for gender and education; get more data on persona*

52. Survey 52 (n = 331)

- **Control Group:** S B1 Q1a D1 D2 D3 Q1b M Per P (with T1 and P’s order randomized)
- **Treatment Group 1:** S Per1 S B1 Q1a D1 D2 D3 Q1b M Per2 P,
- **Treatment Group 2:** S P1 S B1 Q1a D1 D2 D3 Q1b M P2 Per
- *Use: Analyze exogeneity and treatment effects of persona and preoccupation*

53. Survey 53 (n = 418)

- **Control Group:** S B1 Q1a D1 D2 D3 Q1b
- **Treatment Group 1:** S T1a B1 Q1a T3 D1 D2 T2 D3 Q1b T1b
- **Treatment Group 2:** S T2a B1 Q1a T1 D1 D2 T3 D3 Q1b T2b
- **Treatment Group 3:** S B1 T3a Q1a T2 D1 D2 T1 D3 Q1b T3b, where T1, T2, and T3 are attention check questions
- *Use: Analysis of instability in attention check questions*

54. Survey 54

- S P B Q1a D1 D2 D3 Q1b, where Q1a and Q1b display one Republican and one Democrat
- Reason not used: Data collected on Gorilla, which we discovered incorrectly compresses normalized coordinate data. Despite repeated attempts, we failed to receive details about the causes and nature of the compression from Gorilla.

55. Survey 55 (n = 404; 298 after exclusion criteria applied)

- S P B Q1a D1 D2 D3 Q1b, where Q1a and Q1b display one Republican and one Democrat
- *Use: Eyetracking data*

56. Survey 56 (n = 397; 311 after exclusion criteria applied)

- S P B Q1a D1 D2 D3 Q1b, where Q1a and Q1b display two Democrats
- *Use: Eyetracking data*

57. Survey 57 (n = 506)

- **Control Group:** S P B Q1a M D1 D2 D3 Q1b T
- **Treatment Group 1:** S P B T1 Q1a M D1 D2 T1 Q1b T
- **Treatment Group 2:** S P B T2 Q1a M D1 D2 T2 Q1b T
- **Treatment Group 3:** S P B T3 Q1a M D1 D2 T3 Q1b T
- **Treatment Group 4:** S P B T4 Q1a M D1 D2 T4 Q1b T,
where T1 = bipartisan quotes (Buttigieg, Haley) about the importance of young people in political leadership, followed by an agree/disagree question;
T2 = bipartisan quotes (Obama, Bush) about the importance of empathic, morally upstanding people in political leadership, followed by an agree/disagree question;
T3 = bipartisan quotes (Obama, Rubio) about the importance of diverse people in political leadership, followed by an agree/disagree question;
T4 = factual statistics about women's underrepresentation in political office, followed by an agree/disagree question.
The control group receives an additional distractor conjoint (D3) in place of the second priming to keep survey length comparable. Conjoint profiles are fully randomized (both attribute values and attribute order).
- *Reason not used: Priming did not work when conjoint questions (the values of the attributes) differed across respondents; this makes sense, as not everyone would have received, say, a choice between a young and an old candidate*

58. Survey 58 (n = 1,037)

- Same design as Survey 57, but with fixed conjoint profiles: all respondents see the same candidate matchups with identical attribute values and attribute order.
- *Use: Same as Survey 57 but with fixed profiles to enable direct individual-level comparison and to test whether priming effects exist when profile variation is eliminated*

59. Survey 59 (n = 3,521)

- **Treatment Groups 1–10:** S P T1a [B Q1a D1 D2 D3 Q1b T2] T1b
- **Treatment Group 11:** S P [B Q1a D1 D2 D3 Q1b T2],
where T1a and T1b are the same political opinion question asked before and after the conjoint distractor block, and T2 = additional political opinion items (party identification, ideology, 2024 vote, social distance, election legitimacy, conspiracy beliefs, healthcare, military intervention, government spending) embedded within the distractor block. The repeated opinion questions (T1) differ across treatment groups as follows:
 - (a) “In politics people sometimes talk of left and right. Where would you place the Democratic Party on a scale from 0 to 10 where 0 means the left and 10 means the right?” (Slider, 0–10)

- (b) “Some people think we need much tougher government regulations on business in order to protect the environment. Suppose these people are at one end of a scale, at point 1. Others think that current regulations to protect the environment are already too much of a burden on business. Suppose these people are at the other end, at point 7. And, of course, some other people have opinions somewhere in between, at points 2, 3, 4, 5, or 6. Where would you place yourself on this scale, or haven’t you thought much about this?” (Response options: 1–7)
- (c) “I’m going to describe a type of political system. Would you say it is a very good way of governing this country, fairly good, fairly bad, or very bad? *Having experts, not government, make decisions according to what they think is best for the country.*” (Response options: Very good; Fairly good; Fairly bad; Very bad)
- (d) “I’m going to describe a type of political system. Would you say it is a very good way of governing this country, fairly good, fairly bad, or very bad? *Having a strong leader who does not have to bother with parliament and elections.*” (Response options: Very good; Fairly good; Fairly bad; Very bad)
- (e) “I’m going to describe a type of political system. Would you say it is a very good way of governing this country, fairly good, fairly bad, or very bad? *Having a democratic political system.*” (Response options: Very good; Fairly good; Fairly bad; Very bad)
- (f) “Please tell me how much you agree or disagree with the following statement. *Obedience and respect for authority are the most important values children should learn.*” (Response options: Agree strongly; Agree; Disagree; Disagree strongly)
- (g) “Should immigration to the United States be increased, kept the same, or decreased?” (Response options: Decreased; Kept the same; Increased)
- (h) “If China were to invade Taiwan, would you support or oppose the United States sending US troops to Taiwan to help the Taiwanese government defend itself against China?” (Response options: Support; Oppose)
- (i) “Do you approve or disapprove of the way Donald Trump is handling his job as president?” (Response options: Approve; Disapprove; No opinion)
- (j) “What is your gender?” (Response options: Man; Woman; Other)

Treatment Group 11 receives no repeated opinion question and serves as a conjoint-only control. Conjoint profiles are fully randomized (both attribute values and attribute order). The survey includes embedded page-level timing and attention tracking fields.

- *Use: Measure response instability on non-conjoint political opinion questions using the conjoint task as a distractor between repeated measures; additionally provides conjoint instability data via the Q1a/Q1b repeated measure*

References

- Chen, Mon Chu, John R Anderson, and Myeong Ho Sohn (2001). “What can a mouse cursor tell us more? Correlation of eye/mouse movements on web browsing”. In: *CHI’01 extended abstracts on Human factors in computing systems*, pp. 281–282.
- Clayton, Katherine, Yusaku Horiuchi, Aaron R Kaufman, Gary King, and Mayya Komisarich (2025). “Correcting Measurement Error Bias in Conjoint Survey Experiments”. In: *American Journal of Political Science [forthcoming]*. URL: GaryKing.org/conjointE.
- Hauser, David J and Norbert Schwarz (2015). “It’s a trap! Instructional manipulation checks prompt systematic thinking on “tricky” tasks”. In: *Sage Open* 5.2, p. 2158244015584617.
- Huang, Jeff, Ryen White, and Georg Buscher (2012). “User see, user point: gaze and cursor alignment in web search”. In: *Proceedings of the sigchi conference on human factors in computing systems*, pp. 1341–1350.
- Papoutsaki, Alexandra (2015). “Scalable webcam eye tracking by learning from user interactions”. In: *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 219–222.
- Stedman, Richard C, Nancy A Connelly, Thomas A Heberlein, Daniel J Decker, and Shorna B Allred (2019). “The end of the (research) world as we know it? Understanding and coping with declining response rates to mail surveys”. In: *Society & Natural Resources* 32.10, pp. 1139–1154.
- Vos, Myrte, Serge Minor, and Gillian C Ramchand (2022). “Comparing infrared and webcam eye tracking in the visual world paradigm”. In.
- Yang, Xiaozhi and Ian Krajbich (2021). “Webcam-based online eye-tracking for behavioral research”. In: *Judgment and Decision making* 16.6, pp. 1485–1505.