Supporting Information

SI 1 Detailed Information About MTurk Samples

SI 1.1 Descriptive Statistics on Suspicious IP Addresses

pollname	1	2	3	4	5	6	9
August 2018 Study	1885	20	13	4	1	1	1
June 2020 Study	1424	33	3	0	0	1	0
July 2020 Study	396	5	1	0	0	0	0

Table SI 1.1: Number of Times an IP Address Appears in the Data

pollname	Canada	India	Other	United States	Venezuela
August 2018 Study	6	17	54	1870	42
June 2020 Study	0	12	5	1488	0
July 2020 Study	0	0	3	406	0

Table SI 1.2: Country of Origin

pollname	Buffalo	Chicago	Kansas City	Las Vegas	Los Angeles	Maracaibo	New Yor
August 2018 Study	77	0	28	0	44	31	72
June 2020 Study	0	40	0	35	52	0	31
July 2020 Study	0	0	0	0	0	0	0

Table SI 1.3: City of Origin

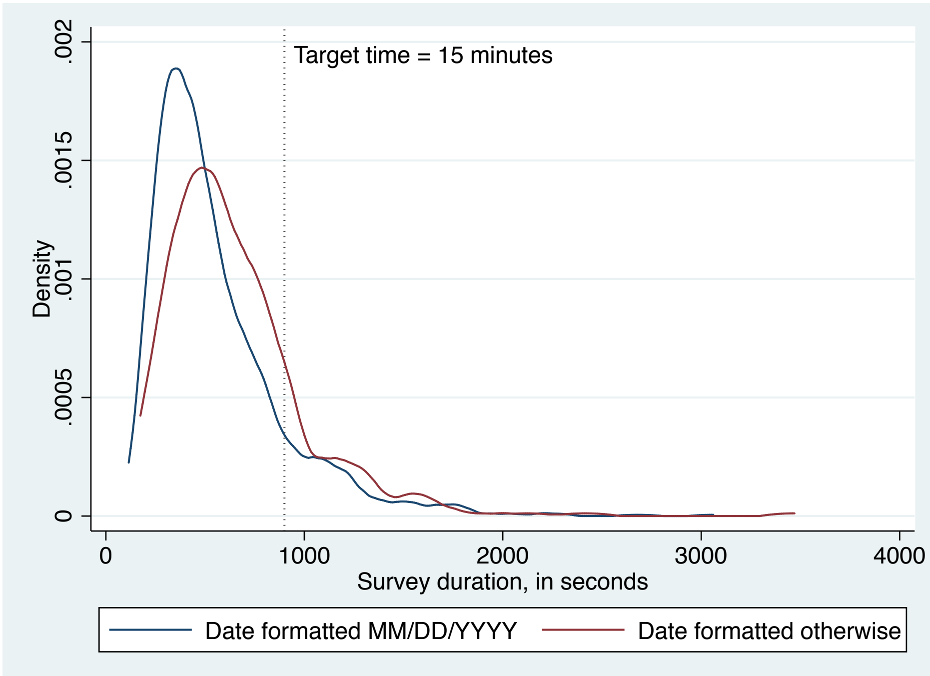
SI 1.2 Probing Duplicate IP Addresses

Duplicate IP addresses present two potential, unique problems. First, and most obviously, the same person may take the survey multiple times. Second, people may take the survey from the same network (e.g., a college campus or a workplace), which especially presents problems if these people are alerting each other to the survey—at a minimum, such a data-generating process will yield larger standard errors than those we calculate naïve to clustering. We make use of our June 2020 survey to assess the degree to which these various processes contribute to the presence of duplicated IP addresses in our data. We index each duplicated IP address and look at the start and end times of each survey from that address to do so. In particular, if the start time of survey j from address i is within ten minutues of survey $j-1$, we classify the duplicates as likely coming from the same individual. (Nearly all of these cases have end/start times within 1-2 minutes of each other.) If there are overlapping start and end times between the duplicated responses, we classify the duplicates as likely coming from a coordinated cluster of individuals. (Note, though, that this could also reflect one individual taking the survey multiple times concurrently on multiple devices.) Accordingly, we estimate that of the 37 duplicate IP addresses in the data, 10 (27%) reflect people filing mulitiple submissions and an additional 21 (57%) coming from coordinated clusters. And even with the remaining 16% of responses from duplicated IP addresses, there is likely significant heterogeneity within those clusters that should be accounted for in analysis.

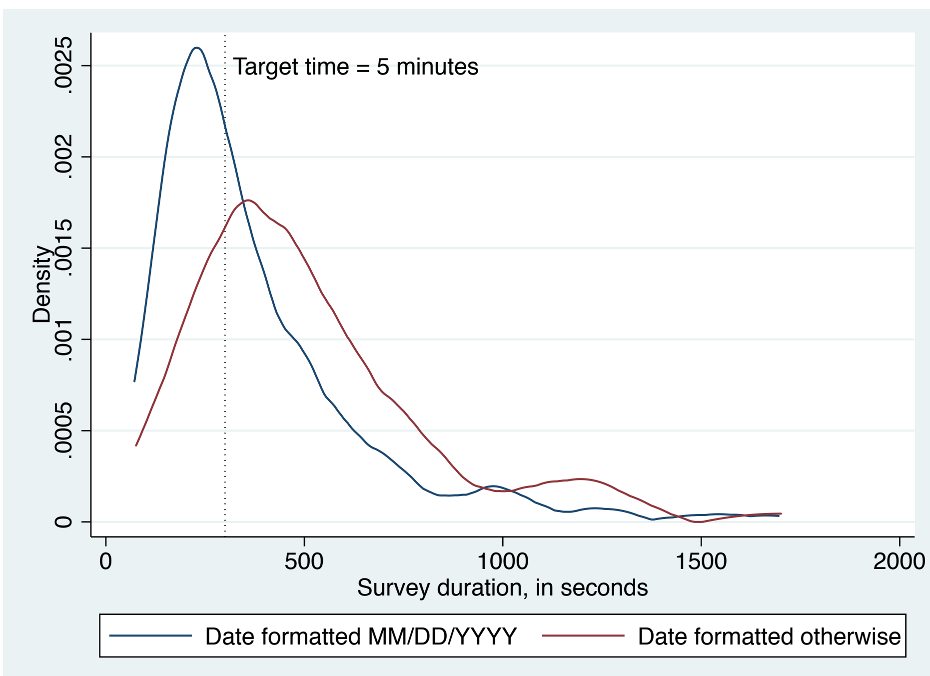
SI 1.3 Probing Foreign IP Addresses

One possibility worth investigating is whether foreign respondents actually try to take the survey genuinely—or, at the very least, spend time reading it. One might expect that people with limited English and/or understanding of American politics would speed through the survey, since they are likely taking it purely for the reward. But this is not what we find. When we use writing the date in DD/MM/YYYY format (asked on the 2020 surveys) as a proxy for being outside the U.S., we find that respondents who write the date in the non-U.S. format actually take longer, on average. The figures below show this, plotting the distributions of completion times by whether the date was written DD/MM/YYYY or MM/DD/YYYY. We find that people who use the U.S. date format complete the survey more quickly—Kolmogorov-Smirnov tests conclude the probability that these sets of response times were drawn from the same distribution is less than .001 for each survey. This suggests that respondents outside the U.S. may actually be trying to read and respond to American MTurk surveys in some meaningful way, despite the fact that they are outside the sampling frame (and are thus undesirable as survey respondents for researchers of U.S. politics). There could be other explanations—slower internet connections, for example—but one possibility is that these respondents take surveys as genuinely as possible so as to avoid detection.

Figure SI 1.1: Surveys from Likely Foreign Respondents Take Longer



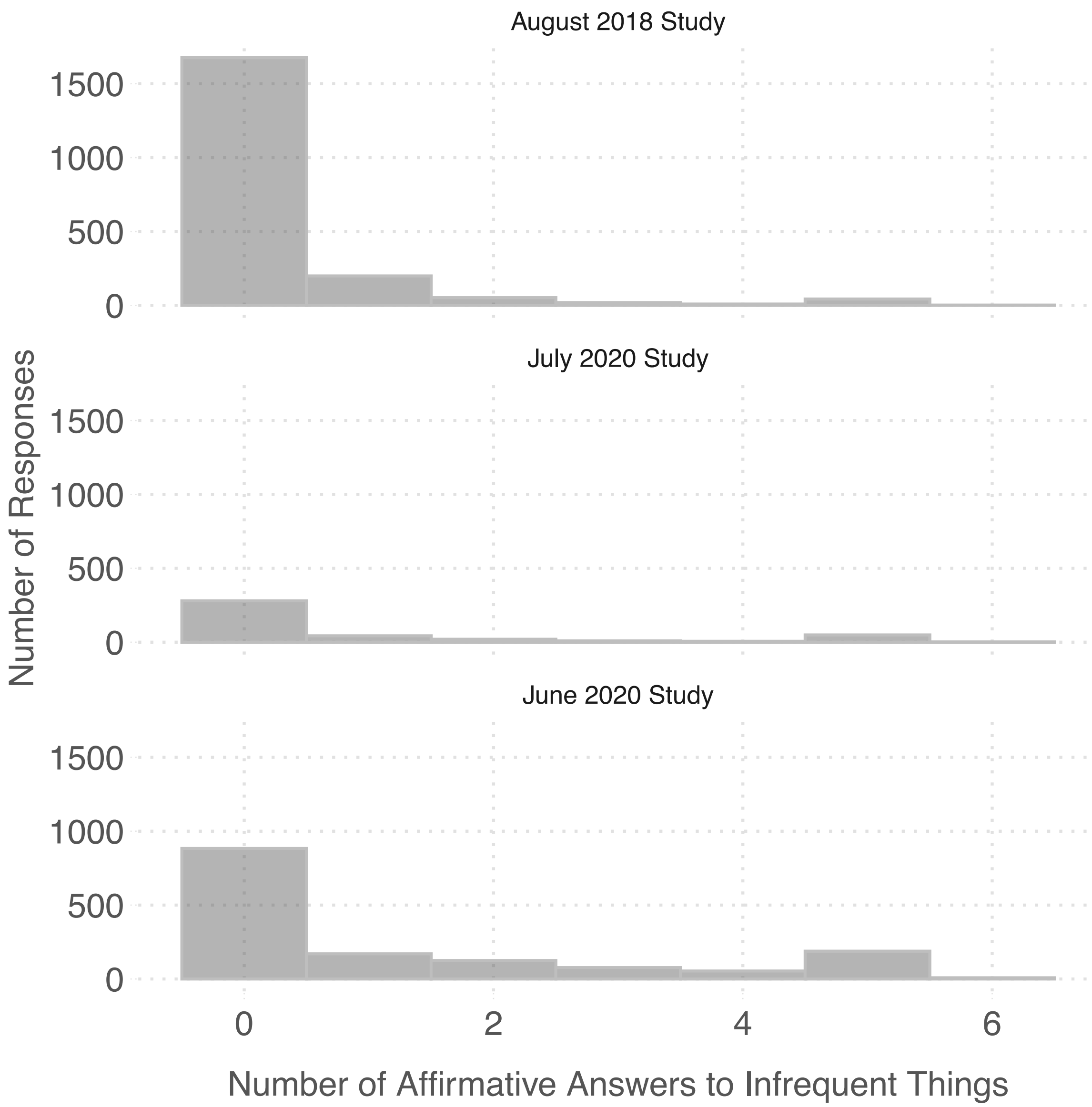
(a) June 2020



(b) July 2020

SI 1.4 Distributions of Counts of Affirmative Responses to Low-Incidence Screener Items

Figure SI 1.2: *Distribution of Counts of Affirmative Responses to Low-Incidence Screener Items Across Surveys*



SI 1.5 Additional Information Regarding Speedy Respondents

Following the same procedure for classifying slow and fast outliers described in the paper, we estimated the proportions of each in Studies 2 and 3.

In Study 2, roughly 6% of the sample is classified as fast outliers, having completed the survey in 232 seconds or less. Roughly 8% of the sample is classified as slow outliers, taking more than 1,112 seconds to finish the survey. While respondents flagged as potential trolls or originating from a suspect IP address are not any less likely to be classified as slow outliers than non-flagged respondents (7.2% vs. 7.6% of the sample, respectively, p(diff)=0.773)) they are significantly less likely to be classified as fast outliers than non-flagged respondents (4.7% vs. 7.1% of the sample, respectively, p(diff)=0.057)).

In Study 3, 14.7% of the sample is classified as fast outliers, having completed the survey in 177 seconds or less. 8.6% are classified as slow outliers—those who took longer than 799 seconds to take the survey. In this survey, 3.5% of respondents flagged as potential bad actors (by virtue of being potential trolls or for having taking the survey from a suspicious IP address) were classified as fast outliers, a figure that dwarfs in comparison to the proportion of non-suspicious respondents classified as such (roughly 18%, p(diff)=0.000)). Suspect respondents are not statistically more likely than non-suspect respondents to be classified as slow outliers (roughly 12% vs. 7.4%, diff, respectively, p(diff)=0.159).

Contrary to conventional wisdom, we do not find that respondents who are extraordinarily fast in completing the survey provide low-quality data. Table SI 1.4 models evaluations of the unemployment and inflation rates as a function of the experimental treatment (described in 5), being a fast outlier, and the interaction of the treatment with fast outlier status. The fact that the coefficients on *Out-party treatment * fast* are not substantively or statistically significant at conventional levels suggests that fast outliers do not respond differently to our experiment than respondents not classified as such. (We find similar results for our other experiment detailed in SI 4.4.) It is for this reason that we do not consider fast outliers as a source of low quality data in our broader analysis.

Table SI 1.4: *Impact of Fast Completion Times on Treatment Effects - June 2020 Survey*

	Unemployment DV	Inflation DV
Out-party treatment	-0.097***	-0.063***
	(0.016)	(0.017)
Fast	0.012	-0.004
	(0.047)	(0.049)
Out-party treatment * fast	0.011	0.043
	(0.064)	(0.067)
Constant	0.792***	0.711***
	(0.011)	(0.012)
Observations	1,425	1,425
R-squared	0.027	0.010

Standard errors in parentheses.
***p<0.01, **p<0.05, *p<0.1, two-tailed.

SI 2 Question Wording

SI 2.1 Low Incidence Screener Battery

- Do you use an artificial limb or prosthetic?
  - Yes
  - No
- Are you blind or do you have vision impairment?
  - Yes
  - No
- Are you deaf or do you have hearing impairment?
  - Yes
  - No
- Are you in a gang?
  - Yes
  - No
- Is one or more of your immediate family members in a gang?
  - Yes
  - No

SI 2.2 Sincerity Self-Report

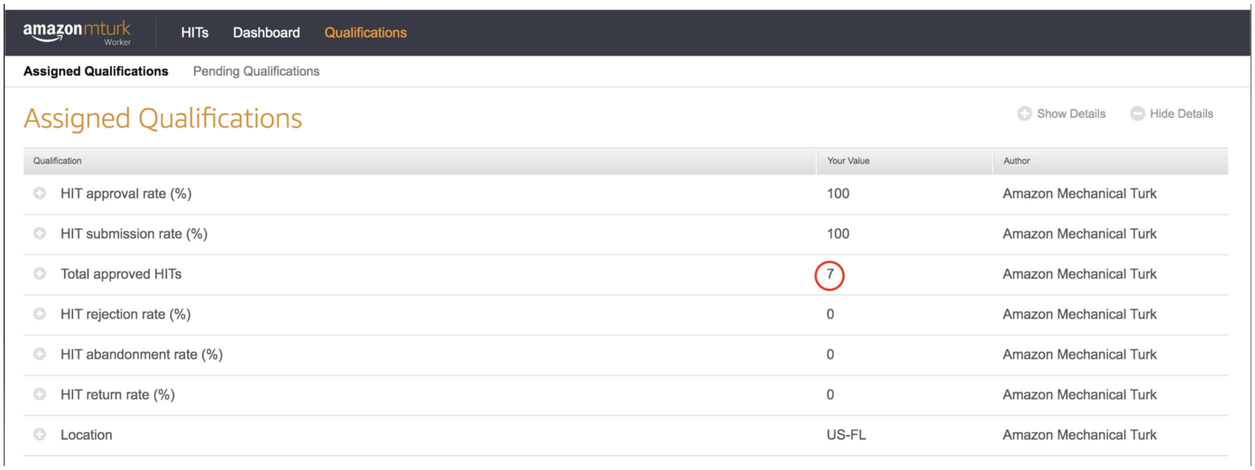
Finally, we sometimes find people don't always take surveys seriously, instead of providing humorous or insincere responses to questions. How often do you do this?

- Never
- Rarely
- Some of the time
- Most of the time
- Always

SI 2.3 Self-Reported Number of HITs Completed

We'd like to know a little more about your participation on MTurk.

To answer the question below, please visit worker.mturk.com/qualifications/assigned and look for your "Total Approved HITs" number (see graphic below). If you cannot find this information, just provide us with your best guess.



Qualification	Your Value	Author
HIT approval rate (%)	100	Amazon Mechanical Turk
HIT submission rate (%)	100	Amazon Mechanical Turk
Total approved HITs	7	Amazon Mechanical Turk
HIT rejection rate (%)	0	Amazon Mechanical Turk
HIT abandonment rate (%)	0	Amazon Mechanical Turk
HIT return rate (%)	0	Amazon Mechanical Turk
Location	US-FL	Amazon Mechanical Turk

About how many HITs have you completed on MTurk?

- Fewer than 100 HITs
- Between 100 and 500 HITs
- Between 500 and 1,000 HITs
- More than 1,000 HITs

SI 2.4 Experimental Item Wording (from Study 2 in June 2020)

Switching gears, we'd like to understand how you think various measures of the economy performed a few years ago, (when Barack Obama was president | when Republicans were in control of both Houses of Congress).

During 2016, (when Barack Obama was president | when Republicans controlled both Houses of Congress), unemployment decreased from 5.0% to 4.8%, a change of 0.2 percentage points. How would you interpret this change? Would you say that unemployment got better, stayed about the same, or got worse?

- Got better
- Stayed the same
- Got worse

In 2016, inflation also decreased from 2.1% to 1.9%, a change of 0.2 percentage points. How would you interpret this change? Would you say that inflation got better, stayed about the same, or got worse?

- Got better
- Stayed the same
- Got worse

SI 3 Experimental Effects by Subgroup

Table SI 3.5: Experimental Effects by Subgroup

	All non-suspicious respondents		All suspicious respondents		Suspicious IPs only		Potential trolls		High HIT (1000+) respondents	
	Unemployment DV	Inflation DV	Unemployment DV	Inflation DV	Unemployment DV	Inflation DV	Unemployment DV	Inflation DV	Unemployment DV	Inflation DV
Out-party treatment	-0.125***	-0.091***	-0.049**	-0.013	-0.029	0.023	-0.043*	-0.017	-0.150***	-0.101***
	(0.020)	(0.021)	(0.023)	(0.026)	(0.044)	(0.048)	(0.026)	(0.028)	(0.025)	(0.026)
Constant	0.775***	0.720***	0.820***	0.698***	0.795***	0.642***	0.819***	0.709***	0.782***	0.743***
	(0.015)	(0.015)	(0.017)	(0.018)	(0.032)	(0.035)	(0.018)	(0.020)	(0.018)	(0.019)
Observations	861	861	564	564	182	182	445	445	505	505
R-squared	0.043	0.022	0.008	0.000	0.002	0.001	0.007	0.001	0.068	0.029

Source: June 2020 study.

Standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.10, two-tailed.

SI 4 A Second Experiment Demonstrating Attenuation of Treatment Effects from Low-Quality Respondents

As an additional means to study how low-quality responses influence the substantive conclusions reached in a study, we embedded an experiment on partisan stereotyping into the August 2018 survey. We replicated a study from Ahler and Sood (2017), examining the degree to which people rely on the representativeness heuristic when making judgments about party composition. Specifically, the study investigates the degree to which people use information about how social groups "sort into" one of the two parties (at the expense of other relevant considerations) to make inferences about aggregate party composition. One way to assess this—specifically, the "at the expense of other relevant considerations" part—is to exploit the *conjunction fallacy*, a cognitive error that occurs when people assert the probability of two events occurring together is greater than the probability of either occurring separately (Tversky and Kahneman 1974).

Ahler and Sood (2017) itself is a modification of Tversky and Kahneman's (1974) "Linda Problem," which presented respondents with the following question:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable?

- Linda is a bank teller.
- Linda is a bank teller and is active in the feminist movement.

The latter option is logically impossible, as the probability that Linda is both a bank teller *and* active in the feminist movement will always be less than or equal to the probability that Linda is a bank teller. Therefore, when respondents select the second option, they commit the conjunction fallacy as a result of their overreliance on representative characteristics.

Ahler and Sood (2017) modified the Linda problem by manipulating the characteristics of the target in the vignette (i.e., making the character more or less representative of one of the two parties) to assess which characteristics people weigh most heavily in party stereotypes (Ahler and Sood 2018). To do so, they introduced respondents to a character named James, randomly and independently manipulating particular party-representative characteristics (like gender, race, sexual orientation, and religion) within a vignette. This design is ideal for our purposes here, as the independent manipulation of several features allows for multiple tests of treatment effect attenuation. That is, instead of comparing how suspicious and non-suspicious respondents differ in their response to *one* treatment, we can do so for *multiple* treatments at once, improving statistical power. The vignette read as follows:

> James is a 37-year-old (`white` | `black`) man. He attended the University of Michigan, where he double-majored in economics and political science. While there, James was president of a business and marketing club. He also participated in (`anti-tax demonstrations` | `living-wage demonstrations` | `student government`).
>
> James's co-workers describe him as highly driven, outspoken, and confident. He is married to (`Karen` | `Keith`) and has one son. In James's free time, he (`leads his son's Cub Scouts group, organized through the Baptist Church the family attends` | `leads his son's Junior Explorers group, led through the Secular Families Foundation` | `coaches his son's youth sports teams`).

Following the vignette, we asked respondents what they believe to be most likely among three options: (1) "James is a salesman," (2) "James is a salesman who also supports the Democratic Party," and (3) "James is a salesman who also supports the Republican

Party." In selecting option (2) or (3), respondents commit the conjunction fallacy. In their original study, Ahler and Sood (2017) found, unsurprisingly, that exposure to characteristics that are representative of the Democratic (Republican) Party leads individuals to commit the Democratic (Republican) conjunction fallacy. By including a replication in the present survey, we can examine whether suspicious respondents react differently than traditional survey-takers to an already-validated treatment.

To determine if and how low-quality responses moderate treatment effects, we estimated the *average marginal component effect* (AMCE) of each independently randomized characteristic interacted with an indicator for a low-quality response on the probability that respondents make the Democratic and Republican conjunction fallacies. Since the dependent variable takes on three values—Democratic conjunction fallacy (-1), logically correct response (0), Republican conjunction fallacy (1)—we use an ordered logit model (omitting one value per variable) to analyze the data. Thus, our model takes the following form, with i indexing respondents and j indexing possible values of the dependent variable:

$$p_{ij} = p(y_i = j) = \begin{cases} p(y_i = -1) = p(y_i^* \leq \alpha_{-1}) \\ p(y_i = 0) = p(\alpha_{-1} < y_i^* \leq \alpha_0) \\ p(y_i = 1) = p(\alpha_0 < y_i^*) \end{cases} \tag{2}$$

where y_i^* is the respondent's latent outcome and α_{-1} and α_0 are the model's cutpoints. We model these probabilities as follows:

$$p(y_i = j) \sim \text{logit}^{-1}(\beta_k X_{ik} + \delta LQ_i + \gamma(LQ_i \times X_{ik}) + \varepsilon) \tag{3}$$

where X_k denotes our vector of randomly and independently assigned characteristics of James (his race, sexuality, etc.) and LQ_i is an indicator for `low quality` response. We operationalize `low quality responses` three ways in three different models: first as all respondents flagged for any reason, then as duplicated/blacklisted IP addresses, and finally as respondents flagged for potential trolling.

Full model results are available in SI 4.2. For ease of interpretation, we present marginal effects in Table SI 4.6, specified as the change in the predicted probability of committing the Democratic/Republican conjunction fallacy. We first present results for all non-flagged respondents (column 1) and then all low-quality respondents (duplicated/blacklisted IP addresses and respondents we suspect are non-serious (column 2). Finally, we present the results for flagged IP addresses alone (column 3) and potential trolls alone (column 4).

The first column confirms significant average marginal component effects (AMCEs) of all randomly and independently varied characteristics. Non-suspicious respondents are significantly more likely to commit the Democratic conjunction fallacy when James is described as black, gay, secular, or as having liberal policy preferences; they are also more likely to commit the Republican conjunction fallacy when James is presented as evangelical or as having conservative policy preferences. In sum, people appear to stereotype others as partisan on the basis of social and policy cues, even making illogical inferences in the process.

Column 2 demonstrates that suspicious respondents react differently. AMCEs are generally attenuated among respondents flagged for any reason. The magnitude of this difference is notable: suspicious respondents, for example, are nearly eight percentage points less likely than non-suspicious respondents to make the Democratic conjunction fallacy when James is presented as black. They are almost ten percentage points less likely to make the Democratic conjunction fallacy when James is presented as gay. Oddly, the effect of the conservative cue is substantively larger among suspicious respondents, but this difference from non-suspicious respondents is not precisely estimated.

Averaging these differences in treatment effects (weighted inversely by their estimated standard errors) yields a difference in average treatment effects between suspicious and non-suspicious respondents of 3.7 percentage points (95% confidence interval (CI): [0.10, 6.5]).

Table SI 4.6: Impact of Low-Quality Responding on Treatment Effects - Marginal Effects

	Non-flagged respondents ($n = 1,507$)		All low-quality respondents ($n = 484$)		Flagged IPs only ($n = 397$)		Non-serious respondents only ($n = 125$)	
When James is described as...	More likely to make Dem. CF by	More likely to make Rep. CF by	More likely to make Dem. CF by	More likely to make Rep. CF by	More likely to make Dem. CF by	More likely to make Rep. CF by	More likely to make Dem. CF by	More likely to make Rep. CF by
Black (vs. white)	**14.0%**	**-9.7%**	*5.1%*	*-3.7%*	**7.5%**	**-5.3%**	*-4.6%*	*3.6%*
Gay (vs. straight)	**19.1%**	**-13.2%**	***9.3%***	***-6.8%***	**11.6%**	**-8.3%**	*-0.6%*	*0.4%*
Evangelical (vs. nothing)	**-5.8%**	**4.2%**	1.3%	-0.9%	-1.9%	1.4%	4.9%	-3.7%
Secular (vs. nothing)	**6.6%**	**-4.5%**	7.3%	-5.2%	5.6%	-3.9%	12.4%	-9.0%
Liberal (vs. nothing)	**9.4%**	**-6.4%**	2.6%	-1.9%	5.5%	-3.8%	*1.0%*	*-0.8%*
Conservative (vs. nothing)	**-8.3%**	**5.9%**	**-11.6%**	**9.0%**	**-12.4%**	**9.3%**	-16.4%	13.7%

Estimates in **bold** are significantly different from zero ($p < 0.05$).

Estimates in *italics* are significantly different from those in the non-suspicious respondents column ($p < 0.05$).

When we calculate a precision-weighted average difference between treatment effects in the entire sample and those among non-suspicious respondents, we observe an attenuation effect of roughly 0.9 percentage points [95% CI: [0.3, 1.6]). We can contextualize this attenuation effect by putting it in percentage terms: the observed precision-weighted average treatment effect among non-suspicious respondents is 8.9 percentage points, and the presence of suspicious respondents (and their noisy data) attenuates this estimated effect by 10.1% (see SI 4.3 for more on this estimation procedure).

Estimates are generally attenuated among responses with flagged IPs (column 3), but we find more puzzling results among trolls or satisficers (column 4). These potentially non-serious respondents were significantly more likely to profess James to be a *Democratic* salesman when James was described as evangelical, and more likely to commit the *Republican* conjunction fallacy when James had liberal views. Oddly, however, the effects of the secular and conservative cues were substantively large within this group—larger than those observed for non-suspicious respondents—and in the correct direction, albeit imprecisely estimated because of the small number of potential trolls. While potential trolls appear to mostly add noise to our data, these respondents may pose a larger problem if they respond more systematically to other treatments in a way that differs from non-suspicious respondents—and these results do not allow us to rule that possibility out.

SI 4.1 Question Wording for the "James" Experiment

Experimental Manipulation

Please read the descriptions of recent college graduates on this screen and the next and answer the related questions.

James is a 37-year-old (`white` | `black`) man. He attended the University of Michigan, where he double-majored in economics and political science. While there, James was president of a business and marketing club. He also participated in (`anti-tax demonstrations` | `living-wage demonstrations` | `student government`).

James's co-workers describe him as highly driven, outspoken, and confident. He is married to (`Karen` | `Keith`) and has one son. In James's free time, he (`leads his son's Cub Scouts group, organized through the Baptist Church the family attends` | `leads his son's Junior Explorers group, led through the Secular Families Foundation` | `coaches his son's youth sports teams`).

GPA Guess

What do you think James' GPA was in college?

- 3.80 - 4.00
- 3.50 - 3.79
- 3.00 - 3.49
- 2.50 - 2.99
- 2.49 or below

Conjunction Fallacy

Which of the following do you think is most likely?

- James works in sales
- James works in sales and is an active supporter of the Democratic Party
- James works in sales and is an active supporter of the Republican Party

SI 4.2 Results of Fully Specified Ordered Logit Model

Table SI 4.7: Impact of Low-Quality Responses on Treatment Effects - Full Ordered Logit

	All respondents	Suspicious IPs	Non-serious respondents
Low-quality response	-0.15	-0.11	-0.28
	(0.26)	(0.29)	(0.59)
Black	-0.62	-0.61	-0.61
	(0.10)	(0.10)	(0.10)
Black * LQ	0.41	0.28	0.85
	(0.20)	(0.23)	(0.42)
Gay	-0.83	-0.83	-0.82
	(0.10)	(0.10)	(0.10)
Gay * LQ	0.46	0.34	0.95
	(0.20)	(0.22)	(0.42)
Evangelical	0.26	0.26	0.26
	(0.12)	(0.12)	(0.12)
Evang. * LQ	-0.31	-0.24	-0.77
	(0.24)	(0.27)	(0.54)
Atheist/agnostic	-0.31	-0.29	-0.29
	(0.24)	(0.13)	(0.13)
AA * LQ	0.00	0.06	-0.42
	(0.25)	(0.28)	(0.53)
Liberal	-0.42	-0.41	-0.41
	(0.13)	(0.13)	(0.13)
Lib. * LQ	0.31	0.31	0.95
	(0.24)	(0.27)	(0.51)
Conservative	0.36	0.36	0.36
	(0.12)	(0.12)	(0.12)
Con. * LQ	0.10	0.09	0.21
	(0.24)	(0.27)	(0.51)
Cut 1	-0.60	-0.59	-0.59
	(0.13)	(0.13)	(0.13)
Cut 2	0.67	0.65	0.65
	(0.13)	(0.13)	(0.13)
Pseudo R^2	0.04	0.05	0.05
n	1,991	1,866	1,594

NOTE: "LQ" is an indicator for "low-quality." Its exact operationaliztion changes from model to model. In Column 1, LQ == 1 includes all respondents flagged for any reason. In Column 2 we drop likely non-serious respondents so that LQ == 1 only includes respondents flagged for suspicious IP addresses. Finally, in Column 3 we drop respondents flagged for suspicious IP addresses so that LQ == 1 only includes respondents flagged as potential trolls.

SI 4.3 Calculating Attenuation Effects

From the data and the ordered logistic regression model specified above, we estimate the average change in respondents' predicted probability of committing the Democratic and Republican conjunction fallacies when they see that James has k_1 attribute instead of some omitted category k_0. (For example, k could be race, with k_1 meaning that James is black and k_0 that he is white.)

We estimate these average changes in the effect of attributes k among: (1) the full sample, (2) non-suspicious respondents, and (3) suspicious respondents. From there, we calculate the average difference in treatment effects, weighted inversely by the standard errors of those estimated differences, between pairs of these three groups. The difference between groups 1 and 2 is the average attenuation effect as a percentage. We can further contextualize this difference by dividing the estimated effects of k in group 1 by the estimated effects in group 2, which yields the relative size of the observed effect to the "real" effect (i.e., the effect among non-suspicious respondents only)—the *attenuation ratio*. We calculate an average attenuation ratio, weighted again by the inverse of the standard error of these estimated differences. Subtracting the attenuation ratio from 1 yields the attenuation effect in percentage point terms.

SI 4.4 Additional Information Regarding Speedy Respondents

Echoing the results presented in SI 1.5, we do not find that fast outliers react differently to experimental treatments than respondents who are neither extraordinarily fast or slow. In only one out of six cases do they appear to respond significantly differently—the atheist/agnostic cue ($p = .09$)—but the coefficient is incorrectly signed for our hypothesis; fast outliers are slightly more responsive to this cue than slower non-suspicious respondents are.

Table SI 4.8: Impact of Fast Completion Times on Treatment Effects - Full Ordered Logit

	DV: James Experiment
Fast outlier	0.55
	(1.04)
Black	-0.62***
	(0.10)
Black * fast	0.97
	(0.97)
Gay	-0.83***
	(0.10)
Gay * fast	-0.13
	(0.86)
Evangelical	0.26**
	(0.12)
Evang. * fast	-0.52
	(1.00)
Atheist/agnostic	-0.26
	(0.13)
AA * fast	-1.84*
	(1.08)
Liberal	-0.41***
	(0.13)
Lib. * fast	-0.55
	(1.06)
Conservative	0.37
	(0.12)
Con. * fast	-0.99
	(0.96)
Cut 1	-0.57
	(0.14)
Cut 2	0.65
	(0.14)
Pseudo R^2	0.05
n	1,507