Social Desirability Bias and Polling Errors in the 2016 Presidential Election*

Andy Brownback, Aaron Novotny

Abstract

Social scientists have observed that socially desirable responding (SDR) often biases responses to unincentivized questions. In the final month of the 2016 presidential campaign, we conducted three list experiments to test the effect SDR has on responses to polls of agreement with presidential candidates. We elicit a subject's agreement with either Hillary Clinton or Donald Trump using explicit questioning or an implicit elicitation that allows the subject to veil their individual response. We find marginally significant evidence that explicit polling overstates agreement with Clinton and understates agreement with Trump. Dividing our sample by party affiliation, we find that SDR has a significantly larger effect on statements of agreement with the opposing party's candidate. Democrats, in particular, are significantly less likely to reveal agreement with Trump when asked explicitly. This exaggerates the disagreement between Democrats and Republicans, underestimating the likelihood of large swings in the electorate. We find no evidence that ideological alignment drives SDR. We merge our location-specific survey data to county-level voting data and find suggestive evidence that SDR may originate with local voting patterns.

1 Introduction

Political polls generate sweeping economic and political consequences far in advance of election day. Polling numbers are used to motivate changes in campaign spending, staff deployment, fundraising efforts, and even policy positions. Strong polling numbers, for example, motivated Hillary Clinton's 2016 campaign to forgo campaigning in certain states in the upper Midwest that her opponent, Donald Trump, subsequently won. While the exact causal effect of campaign efforts on election outcomes is debatable, the influence of polling on a campaign's strategic focus is incontestable. Polls play additional roles in winnowing television debate participants and evaluating the viability of a candidate months in advance of election day.[1]

Since an incentive-compatible method of collecting voting preferences would be infeasible—and illegal in most cases—methods that rely on stated preference between candidates have been accepted as viable, second-best alternatives. Critics of polling typically point to its vulnerability to non-response bias and optimism bias (Pew Research Center (2012); Armstrong (2001)). But, research from psychology and

*Brownback: Department of Economics, University of Arkansas (e-mail: abrownback@walton.uark.edu). Novotny: Department of Economics, University of Arkansas (e-mail: anovotny@walton.uark.edu).

[1]For example, Fox News Insider (2016) reports that Fox News limited its 2016 Republican presidential primary debate to the 10 candidates with the highest average polling.

economics offers several new reasons to consider the weak incentive-compatibility of polling dubious. Since Maccoby and Maccoby (1954) and Edwards (1957), social scientists have known that survey participants have a bias towards concealing their true preferences when those preferences are not perceived to be socially desirable. Behavior consistent with "socially desirable responding" (SDR, hereafter) holds in many social, political, and electoral contexts.[2] For example, feelings toward African-American politicians (Heerwig and McCabe (2009), Redlawsk et al. (2010), Stephens-Davidowitz (2014)), female politicians (Streb et al. (2008)), and Jewish politicians (Kane et al. (2004)). SDR has also been shown to influence the expression of beliefs about immigration (Janus (2010)), same-sex marriage (Powell (2013), Lax et al. (2016), Coffman et al. (2016)), and racially charged statements (Krysan (1998)).

With record levels of candidate unfavorable ratings (Enten (2016*a*)), the 2016 presidential election provides optimal conditions under which SDR could threaten the validity of polling results. Moreover, historically high levels of partisanship among the voting bases of each party (Andris et al. (2015); Pew Research Center (2016)) allow us to understand the differential effect of SDR across an increasingly divided electorate.[3] The specific question about the role of SDR in the 2016 election has been addressed in several newspaper articles and in a Morning Consult study offering a more in-depth analysis that explores differential responses to telephone and online polls (Dropp (2015)).[4] These analyses have compiled traditional polls and reanalyzed them to find small but not statistically significant effects of SDR on Donald Trump's polling performance. In contrast with this approach, our paper is, to our knowledge, the first to collect and analyze data for the express purpose of identifying the effect of SDR on candidate polling. Moreover, we cover both telephone and online environments and use a methodology designed explicitly to test for SDR in poll responses.[5]

Our results show marginally significant evidence that SDR may cause explicit polling to understate agreement with Donald Trump and overstate agreement with Hillary Clinton. We decompose our sample by political party and find that SDR has a large and significant effect on the willingness of voters to state agreement with the opposing party's candidate. Additionally, we find that, while the effect of SDR is closely related to party affiliation, it is unrelated to political ideology. That is, SDR is closely tied to the party a voter has chosen but is unrelated to policy preferences that may have driven them to that party.

[2] Paulhus (1984); Droitcour et al. (1991); Fisher (1993); Rudman and Kilianski (2000)

[3] In 1969, Richard Nixon referred to the "silent majority" of people who concealed their support for the Vietnam War. Similarly, the "Bradley effect" was a hypothesized reluctance among voters to reveal that their votes against Tom Bradley were racially motivated. In Great Britain, a similar theory has been labeled the "Shy Tory Factor."

[4] See also, Enten (2016*b*), Connors et al. (2016), Shepard (2016)

[5] All analysis is run within a polling medium to control for medium-specific effects.

We use 3 list experiments (a method sometimes called the "item count technique") to estimate the effect of SDR on political polling. This method was developed by Miller (1984) to understand the ways in which respondents predictably misreported answers to unincentivized polling questions.[6] In a list experiment, subjects are presented with a list of statements and asked to report the *total number* they agree with. Half of the subjects are assigned to the *Implicit* treatment where their list features 5 statements, including a "sensitive" statement of interest. The other half of the subjects are assigned to the *Explicit* treatment where their list consists of the same 4 non-sensitive statements and is followed by a direct—"Yes" or "No"—question about the sensitive statement. Thus, all respondents face the same 5 statements, but the treatment assignment randomly varies the observability of an individual's response to the sensitive statement. Blair and Imai (2012) validate and formalize the analysis of list experiments. Critical to the validity of this methodology is the restriction that only socially undesirable responses be affected. Tsuchiya et al. (2007) and Coffman et al. (2016) use placebo tests to validate the methodology.

Figure 1 displays our Implicit and Explicit elicitations. The first two experiments measure the SDR associated with statements of agreement with presidential candidates. The final experiment tests for a differential effect of economic policy preferences on the SDR associated with each candidate. In all three experiments, subjects are randomly assigned to the Explicit or Implicit treatment and are then presented with a sensitive statement that asks about *agreement with* a presidential candidate. Experiment 1—a live telephone poll of 800 Arkansas residents—elicits responses to the statement, "I often find myself agreeing with Donald Trump." In Experiments 2 and 3—online surveys with approximately 1,000 eligible voters each—we randomly assign subjects to respond to either 1) "I often find myself agreeing with Hillary Clinton" *or* 2) "I often find myself agreeing with Donald Trump."

It is important to note that our sensitive statement does not ask which candidate respondents intend to vote for, but simply asks if subjects "often agree" with a randomly assigned candidate. In this way, we can explore the psychological motivations behind revealing candidate preferences that are not as transparent as candidate choice.

[6] A similar method was proposed in Raghavarao and Federer (1979).

EXPLICIT:

Consider the following list of statements. Below, we will ask how many of the statements you agree with.

- I think small businesses are important for the economy.
- I agree with George H.W. Bush's foreign policy.
- I think the threat of global warming is exaggerated.
- I prefer presidential candidates who oppose the NRA.

How many of the previous statements do you agree with?
(0) (1) (2) (3) (4)

Do you often find yourself agreeing with Donald Trump?
YES NO

IMPLICIT:

Consider the following list of statements. Below, we will ask how many of the statements you agree with.

- I think small businesses are important for the economy.
- I agree with George H.W. Bush's foreign policy.
- I think the threat of global warming is exaggerated.
- I often find myself agreeing with Donald Trump.
- I prefer presidential candidates who oppose the NRA.

How many of the previous statements do you agree with?
(0) (1) (2) (3) (4) (5)

Figure 1: Examples of both Explicit and Implicit elicitations of support for Donald Trump.

Comparing responses under the Implicit and Explicit treatments yields a clean comparison of otherwise identical environments where the only change is the psychological cost associated with reporting a socially undesirable response. For example, in the Implicit treatment, a response of "3," does not reveal a position on the sensitive statement. Subjects in the Explicit treatment always reveal their position on the sensitive statement. If agreement with one candidate is more stigmatized than another, we expect to find that the polling numbers for the stigmatized candidate improve relative to the non-stigmatized candidate when elicited under the Implicit treatment. Our results indicate that a differential effect of SDR does exist between the two candidates. In particular, we find that the Explicit treatment decreases statements of agreement with Donald Trump by 0.10 and increases statements of agreement with Hillary Clinton by 0.09. After dividing by party affiliation, the effect of SDR grows for the opposing party's candidate and shrinks for the own-party candidate.

To explore the origins of SDR associated with each candidate, we merge our data set to county-level voting data. This merged dataset allows us to address the possibility that these phenomena may derive from the political preferences of a subject's region. We find suggestive but conflicting evidence that county-level voting patterns may predict whether support for a given candidate is strongly affected by SDR.

Other studies have tested the specific claim that SDR may present a threat to the validity of political polling (Traugott and Price (1992), Bishop and Fisher (1995), Keeter and Samaranayake (2007), Hopkins (2009)). Berinsky (1999) shows that poll respondents more often claim to be undecided when answering a poll potentially subject to SDR. Our paper expands on this tradition by collecting primary data to estimate the impact of SDR on political polls directly. Our methodology is related to other techniques used to manipulate the observability of the respondent's answer such as the randomized response technique (Warner (1965)). The utility of these methods is derived from the ability to allow respondents to reveal socially undesirable behavior without direct observation. This motivation relates to results from economics suggesting that anonymity and the ability to excuse behavior affect choices in social interactions (Dana et al. (2007), Andreoni and Bernheim (2009), Charness and Gneezy (2008), Bénabou and Tirole (2011), Exley (2016)). In exposing how political polling can be influenced by these psychological motivations, we hope to shed light on new methods that may present a clearer picture of voter preferences.

2 Experimental Design

Our list experiment is designed to understand the impact of SDR on affirmations of agreement with a political candidate—either "I often find myself agreeing with Hillary Clinton" or "I often find myself agreeing with Donald Trump"—by varying the observability of that affirmation. To experimentally vary response observability, subjects are assigned to either the "Explicit" or "Implicit" treatment, where their responses are directly observed or veiled, respectively.

In the Implicit treatment, the affirmation of agreement is included in a list with four neutral statements. Subjects are asked to respond with the total number—zero to five—of statements they affirm from the list. Call this a subject's "total affirmations." These total affirmations do not reveal agreement with any one statement, thus a subject's response to the affirmation of agreement is concealed.

In the Explicit treatment, subjects see a list with only the four neutral statements and respond with the total number they agree with. They then respond to the affirmation of agreement directly with a "yes" or "no" answer. In this case, call the "total affirmations" the aggregate number of agreements from the list and the directly elicited affirmation of agreement.

In Figure 1, we presented examples of the Explicit and Implicit elicitations for subjects assigned to evaluate their agreement with Donald Trump. Subjects assigned to evaluate their agreement with Hillary Clinton saw an identical list except that the candidate name was changed. The neutral statements of the list are identical in all three experiments. Like the affirmation of agreement, they are political statements. But, since they are presented identically in both the Implicit and Explicit treatments, any influence they have on responses will be constant across treatments. We chose neutral statements that negatively covary to limit the number of responses of zero or five, which transparently reveal the opinions of a subject assigned to the Implicit treatment.[7]

Our study is comprised of the following three list experiments, each successively narrowing in on relevant psychological phenomena:

2.1 Experiment 1: Arkansas Poll

Our first experiment was included in the Arkansas Poll, a live-telephone survey of 800 Arkansas residents between October 18th and October 27th, 2016. 60 percent of respondents were contacted on land-line telephones and 40 percent were contacted on cell phones. The cooperation rate was 29 percent and 25 percent for land-lines and cell phones, respectively. Respondents skewed towards Donald Trump with 45 percent of the sample indicating plans to vote for him compared to 31 percent for Hillary Clinton. The sample was older than the national average with a median age of 63. Detailed demographics for all three experiments can be found in the appendix.

The Arkansas Poll consisted of approximately 50 questions with several possible follow-up questions. Question 9 asked respondents which presidential candidate they intended to vote for. Our experiment took the place of the 29th and 30th questions, depending on treatment assignment.

Due to space limitations, we only explored responses to one affirmation of agreement, "I often find myself agreeing with Donald Trump." Respondents were randomly assigned to the Implicit or Explicit treatment.

2.2 Experiment 2: M-Turk Poll 1

On November 1, 2016, we conducted a second list experiment online with 1,006 eligible American voters using Amazon's Mechanical Turk website. Respondents skewed towards Hillary Clinton. 56 percent our our sample indicated that they intended to vote for Clinton compared to 23 percent for Trump. The sample was disproportionately young, with a median age of 31.

We explored responses to affirmations of agreement with both candidates. Each subject was randomly assigned to the Implicit or Explicit treatment and then assigned to respond to either "I often find myself agreeing with Hillary Clinton" or "I often find myself agreeing with Donald Trump." This gives us a "two by two" randomization design.

[7] We will repeat the analysis with these observable responses dropped from the Implicit treatment for robustness.

2.3 Experiment 3: M-Turk Poll 2

Our final experiment took place on November 7th and 8th of 2016.[8] We recruited 985 eligible American voters again using Amazon's Mechanical Turk website. 57 percent of our respondents indicated plans to vote for Hillary Clinton relative to 27 percent for Donald Trump. The median age was 32.

We again elicited demographic information and which candidate the subject intended to vote for. We used the same randomization design as Experiment 2: assigning subjects to respond about agreement with Hillary Clinton or Donald Trump and assigning them to the Implicit or Explicit treatment.

In this experiment we added 6 questions about economic policy preferences. Three of the questions indicated more conservative economic policy preferences and three indicated more liberal preferences. This elicitation will allow us to perform sub-group analysis in treatment responses by ideology.

3 Results

Our outcome variable of interest will be Total Affirmations. Recall that, for the Implicit treatment, this measure captures the total number of statements from the list that a subject agrees with. For the Explicit treatment, Total Affirmations equals the number of statements from the list that a subject agrees with plus one if the subject also agrees with the affirmation of agreement with the assigned candidate. If the observability of the response is irrelevant to the subject—that is, if SDR is not a motivation—then Total Affirmations should be equal across the two treatments.

Since the Total Affirmations depend on the response to five different questions, it can be thought of as the sum of five random variables. Thus, our experiment requires large sample sizes to find statistical differences between treatments. When possible, we will control for demographic characteristics to improve our statistical power. In the appendix, we will include a robustness check where we drop all responses of zero or five in the Implicit elicitation since they fully reveal preferences.

3.1 Experiment 1: Arkansas Poll

Subjects were randomly and evenly assigned to the Explicit and Implicit treatments before we elicited their Total Affirmations. We will drop the 70 subjects whose responses did not provide enough information to calculate their number of Total Affirmations.[9] In our sub-group analysis, we will drop any subject

[8] November 8th, 2016 was the day of the election, so it is plausible that certain respondents could have already voted. Since our sensitive statement of interest asks about "agreement" with candidates, not voting preferences, we think this does not present a critical problem.

[9] This could indicate attrition from the survey or refusal to answer relevant questions.

whose subgroup could not be determined.

We present mean Total Affirmations across the different treatments in Table 1. We compare responses across treatments to test the impact of SDR on statements of agreement with Donald Trump. The uncontrolled regression reveals that subjects in the Explicit treatment are 3 percentage points less likely to express agreement with Donald Trump ($p = 0.72$). This estimated difference rises to 4.5 percentage points ($p = 0.59$) with the inclusion of demographic controls.

Table 1: Arkansas Poll: Total Affirmations

	Total Affirmations	
Explicit	2.439	2.264
	(0.06)	(0.45)
Implicit	2.469	2.310
	(0.06)	(0.45)
Controls	No	Yes
N	730	730

Heteroskedasticity-robust standard errors.
Controls: gender, age, income, & education.

Figure 2 divides the sample by party affiliation in order to explore heterogeneity in the effect of SDR.[10] Figure 2 shows that Democrats express 0.74 fewer Total Affirmations than Republicans, on average. When asked explicitly, Democrats' Total Affirmations drop by an additional 0.31 ($p = 0.03$) while Republicans increase their Total Affirmations by 0.11 ($p = 0.42$). This difference in differences estimate is significant at the 5% level ($p = 0.038$).

[10] While the level of agreement with a candidate could drive party affiliation, we are exploring the *differential* likelihood of expressing agreement explicitly. We believe this measure is sufficiently exogenous for use as a sub-group selection criteria.

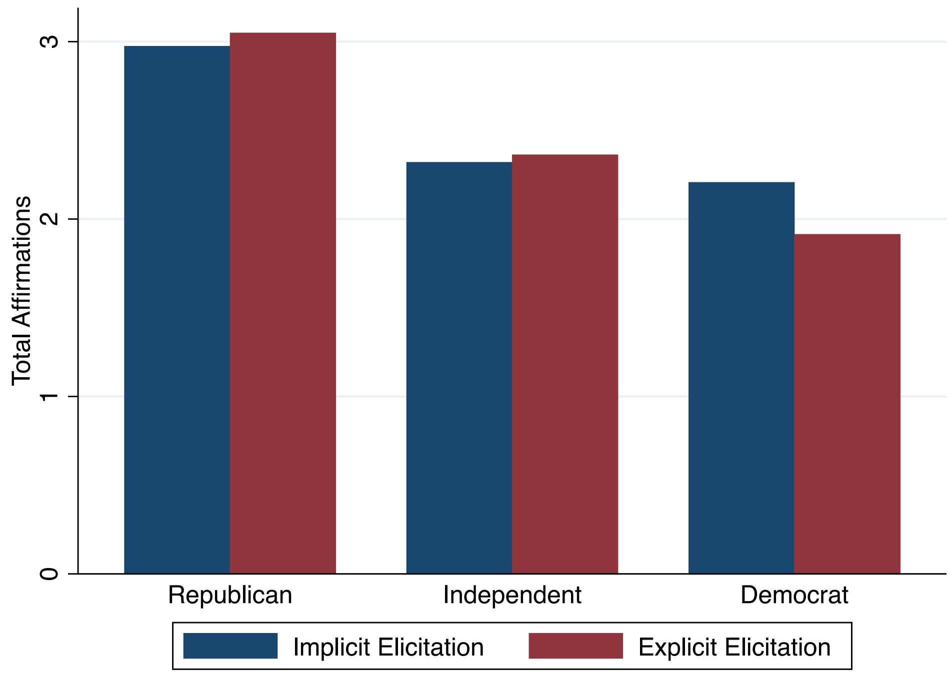


Figure 2: Mean of Total Affirmations split by treatment and party affiliation. Sensitive statement: "I often find myself agreeing with Donald Trump."

3.2 Experiment 2: M-Turk Poll and Candidate Preference

In this experiment, we include affirmations of agreement with Hillary Clinton to compare the effect of SDR across candidates. In Table 2, we present the mean Total Affirmations divided by treatment and randomly assigned candidate. When questioned explicitly, subjects are relatively more likely to report agreement with Hillary Clinton and less likely to report agreement with Donald Trump. Neither candidate shows a statistically significant effect of SDR individually, but when comparing the impact of SDR across the two candidates we estimate a difference in differences of 0.193 ($p = 0.071$). That is, there is a relatively greater effect of SDR on expressions of agreement with Donald Trump.

Dividing our sample based on the subject's party affiliation sheds light on the origin of this differential effect of SDR. Figure 3 shows that the influence of SDR is greater on statements of agreement with the opposing party's candidate. Democrats, in particular, show a significant effect of SDR on their statements of agreement with Donald Trump. Eliciting agreement with Trump explicitly decreases Total Affirmations by 0.180 ($p < 0.01$). SDR has little effect on statements of agreement with Clinton among Republicans. A joint test shows that the effect of SDR increases significantly for statements of agreement with opposing-party candidates relative to own-party candidates. Specifically, the differential effect of the Explicit treatment on Total Affirmations is 0.213 larger when subjects are assigned to their party's

Table 2: M-Turk Experiment 1: Total Affirmations

	Total Affirmations	
Clinton	2.175	2.036
	(0.05)	(0.50)
Clinton × Explicit	0.096	0.090
	(0.08)	(0.08)
Trump	2.000	2.857
	(0.05)	(0.51)
Trump × Explicit	-0.087	-0.103
	(0.08)	(0.08)
Controls	No	Yes
N	1006	1006

Heteroskedasticity-robust standard errors.

Controls: gender, age, and education.

candidate instead of the opposing party's candidate ($p = 0.039$).

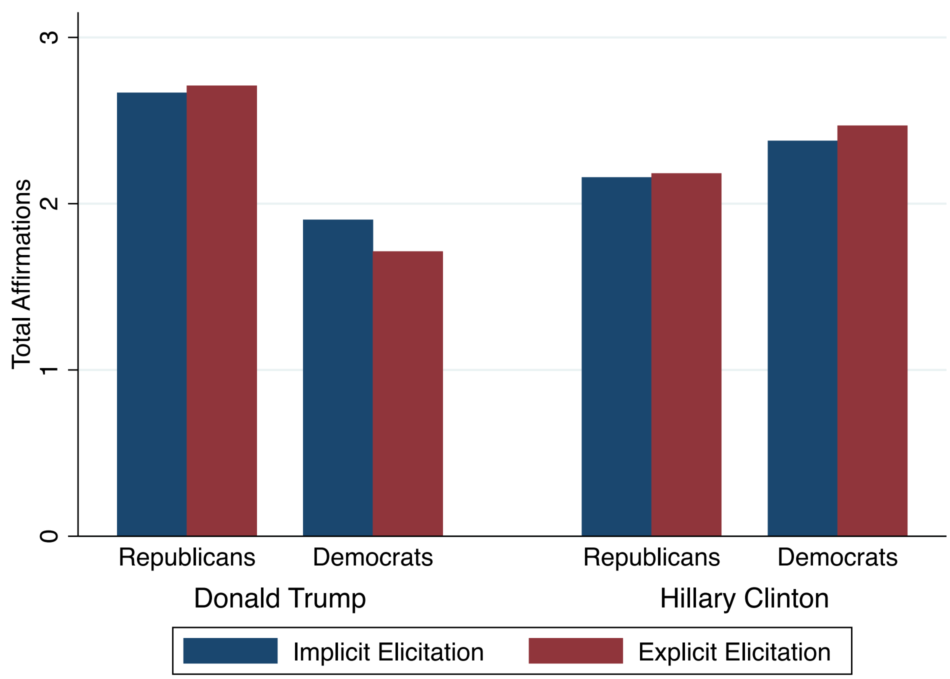


Figure 3: Subjects divided by party affiliation and assigned candidate.

3.3 Experiment 3: M-Turk Poll and Voter Ideology

In this study, we attempt to disentangle the effect of a subject's party affiliation from the ideology that might drive that affiliation. To do so, we use six questions about economic ideology adapted from Halpin and Agne (2009) to divide our sample into conservative and liberal sub-groups.[11] Subjects are again

[11]Subjects responded to each question on a 4-point scale of agreement. We will label a subject conservative (liberal) if the sum of agreement with conservative (liberal) ideological statements exceeds the sum of the agreement with the liberal

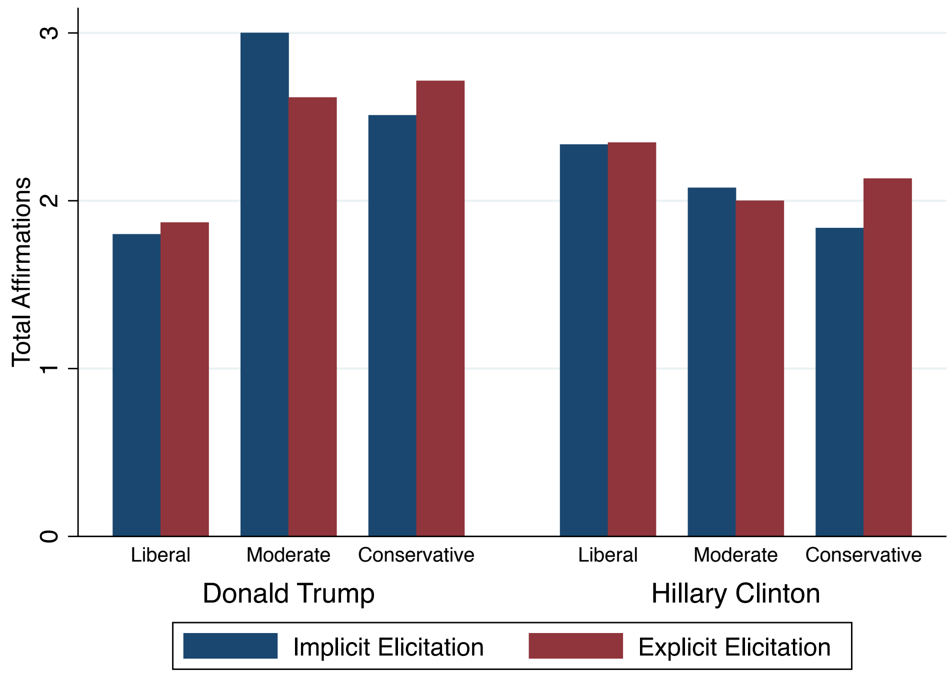


Figure 4: Respondents split by treatment, economic policy position, and randomly assigned candidate.

assigned to evaluate their agreement with a randomly selected candidate.

Figure 4 divides subjects by their economic ideology and assigned candidate and plots Total Affirmations across the Explicit and Implicit treatments.[12] Ideological alignment does drive agreement with a given candidate, the more economically liberal (conservative) is a subject, the more Total Affirmations they report when evaluating Clinton (Trump). However, there is no identifiable pattern to the differential effect of the explicit elicitation across ideologies or candidates. No individual comparisons are statistically significant at conventional levels. Thus, even though Experiment 2 clearly showed an association between party affiliation and SDR, these results suggest that this pattern is not driven by the underlying ideology that drove subjects to join their respective parties.

4 Social-Signaling and County-Level Voting Data

One possible explanation for the origin of social desirability bias is that local, in-person interactions create norms that evolve into socially desirable and undesirable behaviors. We use location data to explore the possibility that the influence of SDR on election polling may have geographic origins. Specifically, we will

(conservative) statements. By this metric, our sample leans liberal—72 percent identify more with liberal economic policy, 21 percent with conservative, and 7 percent are neutral.

[12]Note that sample sizes are not balanced across the ideological bins.

use a subject's geographic location as a proxy for a subject's social setting. Since county-level election results are largely exogenous from the subject's preferences, this approach yields a clean estimate of the influence of a subject's social environment on the SDR associated with their responses to questions about agreement with different candidates.

We leverage location data collected in each survey: The Arkansas Poll collected each subject's county of residence and our Mechanical Turk surveys collected geographic coordinates for each subject's IP address.[13] We merge our survey data with county-level voting data to test if a subject is relatively more likely to *explicitly* state agreement with the candidate who subsequently won the popular vote in the subject's county.[14]

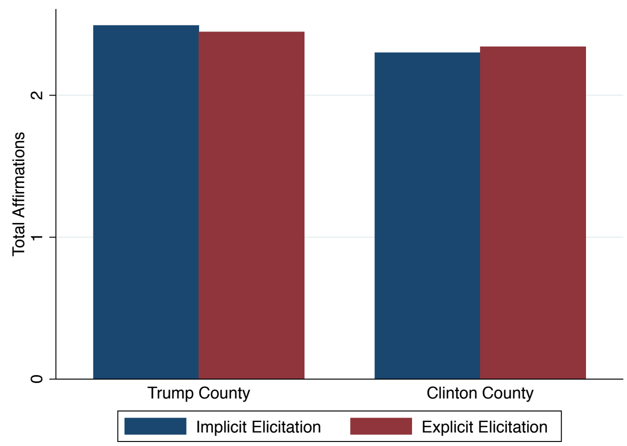
Figure 5 graphs responses to the Arkansas Poll divided by which candidate won the county. In counties that Clinton won, the Explicit elicitation *increases* the likelihood of stating agreement with Trump, while in the counties that Trump won, the Explicit elicitation *decreases* that same likelihood.[15] Neither of these effects approach statistical significance. This could be a result of a relatively small sample—only 14 percent of our sample lives in Arkansas counties won by Clinton, but this pattern is the opposite of what would be expected if SDR were driven by county-level preferences.

Results from the first Mechanical Turk poll are plotted in Figure 6. The larger Mechanical Turk sample conflicts with the Arkansas Poll results and shows that SDR towards Donald Trump is manifest in the responses of subjects from counties that Clinton won. Subjects in these counties reveal 0.23 fewer Total Affirmations when asked about their agreement with Trump explicitly ($p = 0.017$). Comparing this effect to the same effect in counties that voted for Trump does not yield a significant difference in differences ($p = 0.277$). The results do not paint a clear picture, however, since subjects stating agreement with Hillary Clinton explicitly are more likely to do so when they come from counties that Donald Trump won.

[13] IP addresses may not perfectly reflect the subject's place of residence, but should correlate with these, on average.

[14] These data sets were not collected simultaneously, since our experiments occurred before any voting took place. Thus, intervening events could weaken the connection between the two datasets.

[15] Only 8 counties that Hillary Clinton won appear in our data, making this a relatively low-powered test.



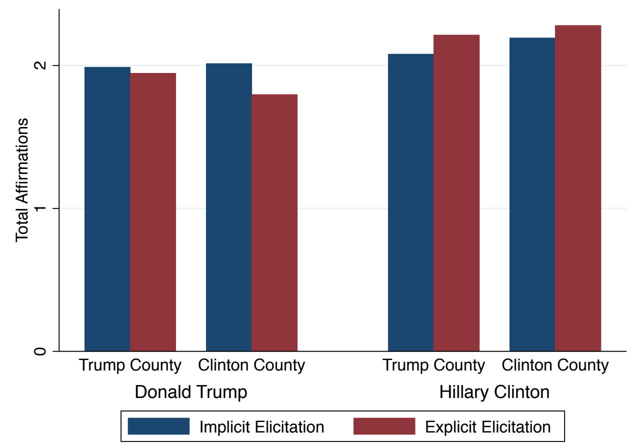


Figure 5: Statements of agreement with Donald Trump divided by county election outcome.

Figure 6: Statements of agreement with Donald Trump and Hillary Clinton divided by county.

5 Discussion and Conclusion

Our results reveal meaningful weaknesses in the current method of eliciting political preferences through explicit polling. In particular, we expose two mechanisms by which socially desirable responding can predictably cause electoral outcomes to deviate from predictions based on explicit poll numbers. First, explicit polling *differentially* underestimates voters' agreement with certain candidates. Second, it exaggerates differences in preferences between the two political parties. This exaggeration gives off the false impression that Democrat and Republican voters have negligible overlap in their agreement with political candidates, leading to an underestimation of the likelihood of large swings in the electorate.

In order to determine the influence SDR may have on policy positions candidates take, we explore the connection between ideological agreement and SDR. We uncover a misalignment between the influence of SDR and a subject's ideological alignment with a candidate. Specifically, our data reject the claim that the influence of SDR decreases as ideological alignment increases. This presents a potential agency problem between the electorate and political candidates. Explicit statements of support for a politician are the primary means by which voters discipline policy choices between elections. So, when a voter's willingness to explicitly reveal support for a candidate fails to respond to changes in the candidate's policy positions, their influence over policy evaporates.

Since we clearly identify a pattern of SDR associated with party affiliation, it is puzzling that ideology plays no important role in SDR, leaving open the question of what aspect of party affiliation drives SDR. Further research is required, but we believe that understanding party identity in the framework of identity economics (Akerlof and Kranton (2000)) may prove fruitful.

Finally, we use county-level voting data to look at the role geography plays in determining the social desirability of candidates. We find suggestive evidence that SDR may be more powerful when a subject is revealing agreement with the candidate who lost the popular vote in their county, though our results are conflicting. The geographic origins of SDR remain a compelling topic for future research, though follow-up studies will need more granular data—for example, at the neighborhood level—to explore a more nuanced concept of a voter's geographic region.

While our results cast doubt on the unbiased nature of explicit polling, our alternative methodology, implicit elicitation, is more complicated to administer and produces noisier estimates that requires larger sample sizes. As such, there are clear limitations to when the implicit elicitation method can substitute for explicit polling. For instance, explicit polling will always be preferred when time or money are of particular concern. Instead of a complete replacement of explicit polling, we propose an alternative approach that respects the speed and simplicity of explicit polling while improving on its accuracy through calibration using implicit polling. Specifically, we suggest that polling organizations conduct occasional, large-sample, implicit elicitations in order to detect bias in their polls. They can then recalibrate their explicit polling results according to the detected bias.

In developing a better understanding of the role SDR plays in political polling, this paper hopes to improve polling methodology and the reliability of forecasts derived from current and future polling methods. Moreover, we hope to provide evidence on the origins of SDR so that future social-science research can take into account respondent characteristics that make SDR an increasingly potent threat to the validity of poll results.

References

Akerlof, George A and Kranton, Rachel E. (2000). 'Economics and identity', *The Quarterly Journal of Economics* 115(3), 715–753.

Andreoni, James and Bernheim, B Douglas. (2009). 'Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects', *Econometrica* 77(5), 1607–1636.

Andris, Clio, Lee, David, Hamilton, Marcus J, Martino, Mauro, Gunning, Christian E and Selden, John Armistead. (2015). 'The rise of partisanship and super-cooperators in the US House of Representatives', *PloS one* 10(4), e0123507.

Armstrong, Jon Scott. (2001), *Principles of forecasting: a handbook for researchers and practitioners*, Vol. 30, Springer Science & Business Media.

Bénabou, Roland and Tirole, Jean. (2011). 'Identity, morals, and taboos: Beliefs as assets', *The Quarterly Journal of Economics* 126(2), 805–855.

Berinsky, Adam J. (1999). 'The two faces of public opinion', *American Journal of Political Science* pp. 1209–1230.

Bishop, George F and Fisher, Bonnie S. (1995). '?Secret ballots? and self-reports in an exit-poll experiment', *Public Opinion Quarterly* 59(4), 568–588.

Blair, Graeme and Imai, Kosuke. (2012). 'Statistical analysis of list experiments', *Political Analysis* 20(1), 47–77.

Charness, Gary and Gneezy, Uri. (2008). 'What's in a name? Anonymity and social distance in dictator and ultimatum games', *Journal of Economic Behavior & Organization* 68(1), 29–35.

Coffman, Katherine B, Coffman, Lucas C and Ericson, Keith M Marzilli. (2016). 'The size of the lgbt population and the magnitude of antigay sentiment are substantially underestimated', *Management Science* .

Connors, Elizabeth, Klar, Samara and Krupnikov, Yanna. (2016). 'There may have been shy Trump supporters after all', *The Washington Post* .
URL: *https://www.washingtonpost.com/news/monkey-cage/wp/2016/11/12/there-may-have-been-shy-trump-supporters-after-all*

Dana, Jason, Weber, Roberto A and Kuang, Jason Xi. (2007). 'Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness', *Economic Theory* 33(1), 67–80.

Droitcour, Judith, Caspar, Rachel A, Hubbard, Michael L, Parsley, Teresa L, Visscher, Wendy and Ezzati, Trena M. (1991). 'The item count technique as a method of indirect questioning: A review of its development and a case study application', *Measurement errors in surveys* pp. 185–210.

Dropp, Kyle. (2015). 'Why does Donald Trump perform better in online versus live telephone polling?', *Morning Consult* .
URL: *https://morningconsult.com/2015/12/21/why-donald-trump-performs-better-in-online-polling/*

Edwards, Allen L. (1957). 'The social desirability variable in personality assessment and research'.

Enten, Harry. (2016*a*). 'Americans Distaste For Both Trump And Clinton Is Record-Breaking', *FiveThirtyEight* .
URL: *http://fivethirtyeight.com/features/americans-distaste-for-both-trump-and-clinton-is-record-breaking/*

Enten, Harry. (2016*b*). "Shy' voters probably arent why the polls missed Trump', *FiveThirtyEight* .
URL: *https://fivethirtyeight.com/features/shy-voters-probably-arent-why-the-polls-missed-trump/*

Exley, Christine L. (2016). 'Excusing Selfishness in Charitable Giving: The Role of Risk', *Review of Economic Studies* 83(2), 587–628.

Fisher, Robert J. (1993). 'Social desirability bias and the validity of indirect questioning', *Journal of consumer research* 20(2), 303–315.

Fox News Insider. (2016), 'See Which Candidates Qualified for the Fox News-Google GOP Debates'.
URL: *http://insider.foxnews.com/2016/01/26/lineup-republican-candidates-fox-news-google-gop-debates*

Halpin, John and Agne, Karl. (2009). 'State of American political ideology, 2009', *Center for American Progress* .

Heerwig, Jennifer A and McCabe, Brian J. (2009). 'Education and social desirability bias: The case of a black presidential candidate', *Social Science Quarterly* 90(3), 674–686.

Hopkins, Daniel J. (2009). 'No more wilder effect, never a Whitman effect: When and why polls mislead about black and female candidates', *The Journal of Politics* 71(3), 769–781.

Janus, Alexander L. (2010). 'The Influence of Social Desirability Pressures on Expressed Immigration Attitudes', *Social Science Quarterly* 91, 928–946.

Kane, James G, Craig, Stephen C and Wald, Kenneth D. (2004). 'Religion and presidential politics in Florida: A list experiment', *Social Science Quarterly* 85(2), 281–293.

Keeter, Scott and Samaranayake, Nilanthi. (2007). 'Can you trust what polls say about Obama's electoral prospects', *Pew Research Center* .
URL: *http://www.pewresearch.org/2007/02/07/can-you-trust-what-polls-say-about-obamas-electoral-prospects/*

Krysan, Maria. (1998). 'Privacy and the expression of white racial attitudes: A comparison across three contexts', *Public Opinion Quarterly* pp. 506–544.

Lax, Jeffrey R, Phillips, Justin H and Stollwerk, Alissa F. (2016). 'Are Survey Respondents Lying About Their Support for Same-Sex Marriage? Lessons from a List Experiment', *Public opinion quarterly* p. nfv056.

Maccoby, Eleanor E and Maccoby, Nathan. (1954). 'The interview: A tool of social science', *Handbook of social psychology* 1, 449–487.

Miller, Judith Droitcourt. (1984). 'A new survey technique for studying deviant behavior', *Unpublished doctoral dissertation, George Washington University, Washington, DC* .

Paulhus, Delroy L. (1984). 'Two-component models of socially desirable responding.', *Journal of personality and social psychology* 46(3), 598.

Pew Research Center. (2012), 'Assessing the Representativeness of Public Opinion Surveys'.
URL: *http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/*

Pew Research Center. (2016), 'Partisanship and Political Animosity in 2016'.
URL: *http://www.people-press.org/2016/06/22/partisanship-and-political-animosity-in-2016/*

Powell, Richard J. (2013). 'Social desirability bias in polling on same-sex marriage ballot measures', *American Politics Research* 41(6), 1052–1070.

Raghavarao, Damaraju and Federer, Walter T. (1979). 'Block total response as an alternative to the randomized response method in surveys', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 40–45.

Redlawsk, David P, Tolbert, Caroline J and Franko, William. (2010). 'Voters, emotions, and race in 2008: Obama as the first black president', *Political Research Quarterly* .

Rudman, Laurie A and Kilianski, Stephen E. (2000). 'Implicit and explicit attitudes toward female authority', *Personality and social psychology bulletin* 26(11), 1315–1328.

Shepard, Steven. (2016). 'GOP insiders: Polls don't capture secret Trump vote', *Politico* .
URL: *http://www.politico.com/story/2016/10/donald-trump-shy-voters-polls-gop-insiders-230411*

Stephens-Davidowitz, Seth. (2014). 'The Cost of Racial Animus on a Black Candidate: Evidence Using Google Search Data', *Journal of Public Economics* 118, 26–40.

Streb, Matthew J., Burrell, Barbara, Frederick, Brian and Genovese, Michael A. (2008). 'Social Desirability Effects and Support for a Female American President', *Public Opinion Quarterly* 72, 76–89.

Traugott, Michael W and Price, Vincent. (1992). 'A review: exit polls in the 1989 Virginia Gubernatorial race: Where did they go wrong?', *The Public Opinion Quarterly* 56(2), 245–253.

Tsuchiya, Takahiro, Hirai, Yoko and Ono, Shigeru. (2007). 'A study of the properties of the item count technique', *Public Opinion Quarterly* 71(2), 253–272.

Warner, Stanley L. (1965). 'Randomized response: A survey technique for eliminating evasive answer bias', *Journal of the American Statistical Association* 60(309), 63–69.

Appendix

A1: Arkansas Poll Demographics

Table 1: **Race**

Race	Number	Percent
White	653	82
Black	72	9
Hispanic	5	1
Asian	1	0
Native American	14	2
Multi-ethnic	21	3
Something else	10	1
Don't know	4	0
Refused	20	2
Total	800	100

*Source:*Arkansas Poll 2016

Table 2: **Political Affiliation**

Affiliation	Number	Percent
Republican	232	29
Democrat	199	25
Independent	295	37
Other	20	2
Don't Know	28	4
Refused	26	3
Total	800	100

*Source:*Arkansas Poll 2016

Table 3: **Gender**

Gender	Number	Percent
Male	357	45
Female	443	55
Total	800	100

*Source:*Arkansas Poll 2016

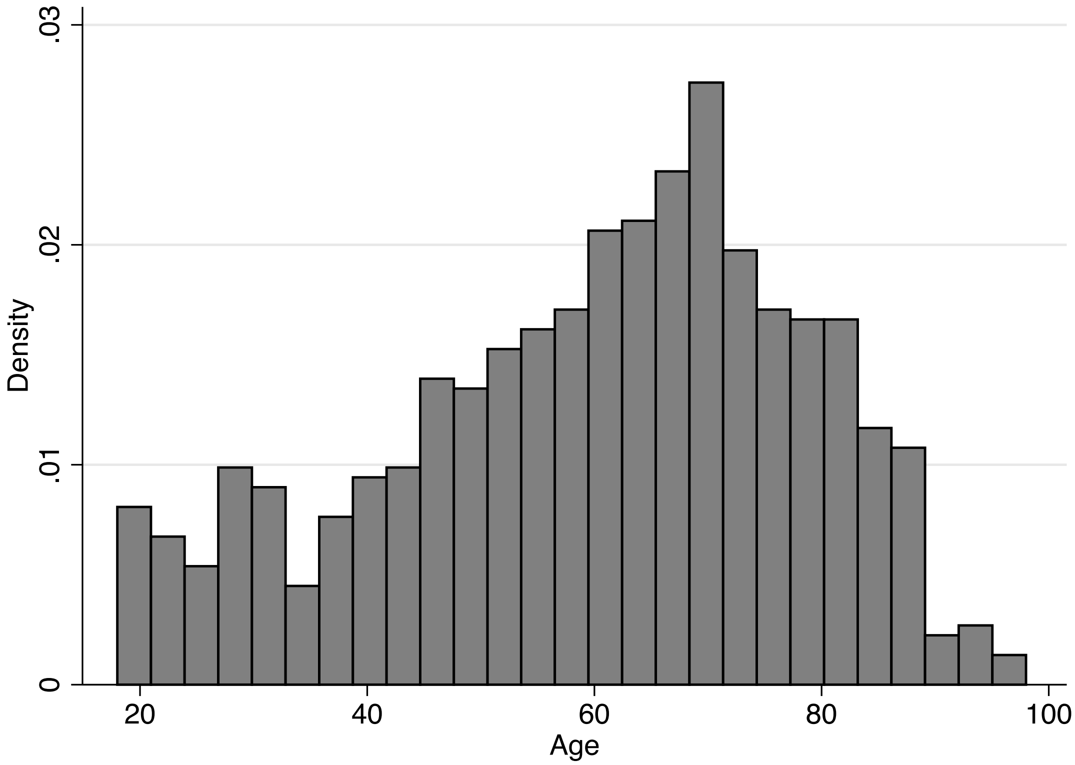


Figure 1: Source: Arkansas Poll 2016

Table 4: **Education**

Education	Number	Percent
No High School	12	2
Some High School	75	9
High School Graduate	218	27
Some College Including Business or Trade School	193	24
College Graduate	150	19
Some Graduate School	28	4
Graduate or Professional Degree	104	13
Don't Know	3	0
Refused	17	2
Total	800	100

Source: Arkansas Poll 2016

Table 5: **Income**

Income	Number	Percent
$7,500 or less	58	7
$7,501 to $15,000	57	7
$15,001 to $25,000	70	9
$25,001 to $35,000	76	10
$35,001 to $50,000	99	12
$50,001 to $75,000	94	12
$75,001 to $100,000	67	8
$100,001 or over	69	9
Don't Know	51	6
Refused	159	20
Total	800	100

*Source:*Arkansas Poll 2016

A2: Mechanical Turk Survey 1 Demographics

Table 6: **Gender**

Gender	Number	Percent
Male	536	58
Female	388	42
Total	924	100

Source: Mechanical Turk: Nov. 1 Survey

Table 7: **Party**

Affiliation	Number	Percent
Democrat	343	37
Lean Democrat	307	33
Lean Republican	197	21
Republican	77	8
Total	924	100

Source: Mechanical Turk: Nov. 1 Survey

Table 8: **Education**

Education	Number	Percent
High School	2	0
Some College	81	9
College Degree	236	26
Some Graduate	88	10
Graduate or Professional Degree	382	41
Refused	135	15
Total	924	100

Source: Mechanical Turk Survey: Nov. 1

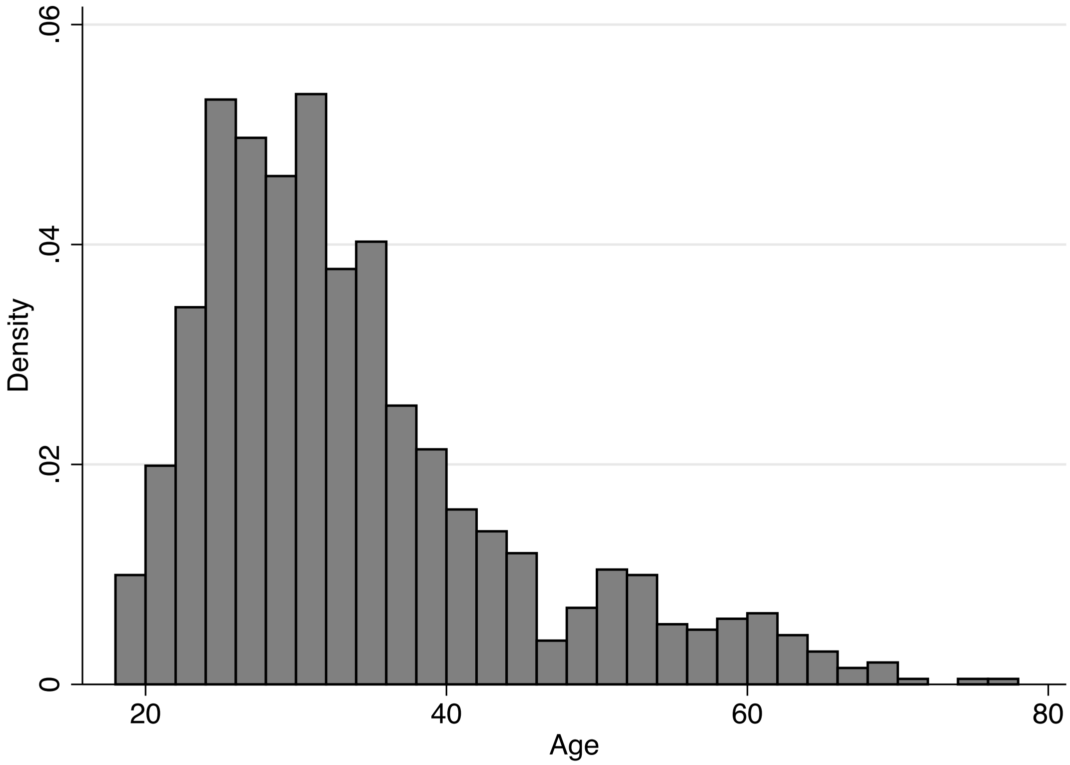


Figure 2: *Source:* Mechanical Turk: Nov. 1 Survey

A3: Mechanical Turk Survey 2 Demographics

Table 9: **Gender**

Gender	Number	Percent
Male	544	55
Female	441	45
Total	985	100

Source: Mechanical Turk: Nov. 8 Survey

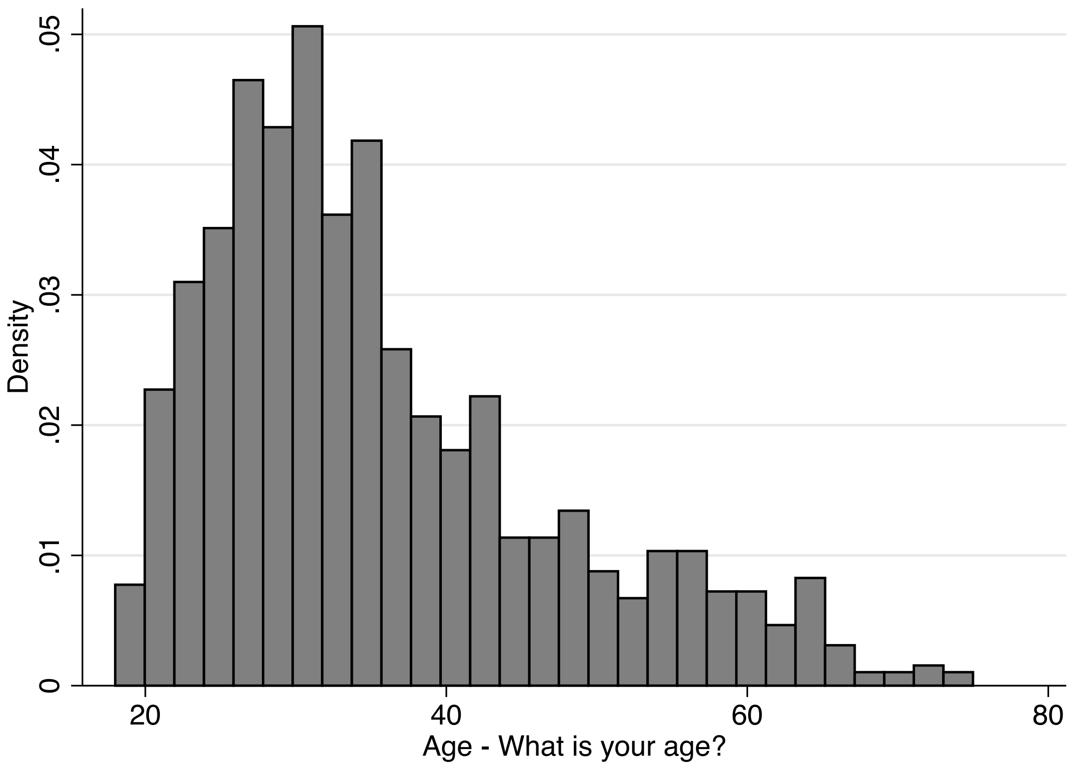


Figure 3: *Source:* Mechanical Turk: Nov. 8 Survey

Table 10: **Education**

Education	Number	Percent
Some High School	11	1
High School Degree	108	11
Some College	280	29
Associate's Degree	101	10
Bachelor's Degree	383	39
Graduate Degree	102	10
Total	985	100

Source: Mechanical Turk Survey: Nov. 8

Table 11: **Party**

Affiliation	Number	Percent
Democrat	353	36
Lean Democrat	303	31
Lean Republican	211	21
Republican	118	12
Total	985	100

Source: Mechanical Turk: Nov. 8 Survey

A4: Regressions from the graphical analysis

Table 12: Arkansas Poll: Total Affirmations by Party Affiliation

	Total Affirmations	
Democrat	2.207	2.571
	(0.12)	(0.41)
Democrat × Explicit	-0.293**	-0.307**
	(0.14)	(0.14)
Republican	2.974	3.309
	(0.10)	(0.41)
Republican × Explicit	0.075	0.114
	(0.14)	(0.14)
Independent	2.320	2.666
	(0.09)	(0.39)
Independent × Explicit	0.042	-0.000
	(0.14)	(0.14)
Controls	No	Yes
N	675	675

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Heteroskedasticity-robust standard errors.

Controls: gender, age, income, & education.

A5: Regressions with revealing responses removed from Implicit treatment

Table 13: Total Affirmations by Assigned Candidate, Party Affiliation, and Treatment

	Total Affirmations	
Trump × Democrat	1.844	1.736
	(0.05)	(0.77)
Trump × Democrat × Explicit	-0.166**	-0.180***
	(0.07)	(0.07)
Trump × Republican	2.346	2.233
	(0.11)	(0.78)
Trump × Republican × Explicit	0.084	0.063
	(0.16)	(0.16)
Clinton × Democrat	2.244	2.146
	(0.07)	(0.77)
Clinton × Democrat × Explicit	0.108	0.102
	(0.09)	(0.09)
Clinton × Republican	2.025	1.903
	(0.09)	(0.78)
Clinton × Republican ×Explicit	0.044	0.034
	(0.14)	(0.14)
Controls	No	Yes
N	1006	1006

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Heteroskedasticity-robust standard errors.

Controls: gender, age, and education.

Table 14: Total Affirmations by Assigned Candidate, Economic Ideology, and Treatment

	Total Affirmations	
Trump × Liberal	1.800	1.506
	(0.06)	(0.14)
Trump × Liberal × Explicit	0.070	0.078
	(0.08)	(0.08)
Trump × Moderate	3.000	2.703
	(0.34)	(0.36)
Trump × Moderate × Explicit	-0.385	-0.403
	(0.40)	(0.39)
Trump × Conservative	2.509	2.207
	(0.14)	(0.18)
Trump × Conservative × Explicit	0.205	0.180
	(0.20)	(0.19)
Clinton × Liberal	2.335	2.025
	(0.06)	(0.14)
Clinton × Liberal × Explicit	0.011	-0.002
	(0.09)	(0.09)
Clinton × Moderate	2.077	1.759
	(0.20)	(0.25)
Clinton × Moderate ×Explicit	-0.077	-0.029
	(0.34)	(0.34)
Clinton × Conservative	1.837	1.553
	(0.12)	(0.17)
Clinton × Conservative ×Explicit	0.295*	0.242
	(0.15)	(0.15)
Controls	No	Yes
N	985	985

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Heteroskedasticity-robust standard errors.

Controls: gender, age, and education.

Table 15: Total Affirmations by County Voting Patterns

	Arkansas Poll		M-Turk Poll 1	
Trump × Clinton-County	2.300	1.919	2.014	1.032
	(0.14)	(0.51)	(0.07)	(0.40)
Trump × Clinton-County × Explicit	0.041	0.045	-0.217**	-0.235**
	(0.16)	(0.18)	(0.10)	(0.10)
Trump × Trump-County	2.492	2.142	1.988	0.993
	(0.07)	(0.48)	(0.09)	(0.40)
Trump × Trump-County × Explicit	-0.046	-0.067	-0.042	-0.045
	(0.08)	(0.07)	(0.15)	(0.15)
Clinton × Clinton-County			2.193	1.220
			(0.07)	(0.39)
Clinton × Clinton-County × Explicit			0.087	0.080
			(0.09)	(0.09)
Clinton × Trump-County			2.080	1.109
			(0.10)	(0.39)
Clinton × Trump-County ×Explicit			0.133	0.116
			(0.13)	(0.13)
Controls	No	Yes	No	Yes
N	721	721	893	893

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Standard errors clustered by county where possible, otherwise by city.

Controls: age, gender, education. Income control added for Arkansas Poll.

Table 16: Arkansas Poll: Total Affirmations

	Total Affirmations
Explicit	2.439
	(0.06)
Implicit	2.451
	(0.05)
Controls	No
N	692

Heteroskedasticity-robust standard errors.

Table 17: Arkansas Poll: Total Affirmations by Preferred Candidate

	Total Affirmations
Clinton–Voters	2.088
	(0.09)
Clinton–Voters × Explicit	-0.190
	(0.12)
Trump–Voters	2.805
	(0.07)
Trump–Voters × Explicit	0.240**
	(0.10)
Controls	No
N	566

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Heteroskedasticity-robust standard errors.

Table 18: M-Turk Experiment 1: Total Affirmations

	Total Affirmations
Clinton	2.199
	(0.05)
Clinton × Explicit	0.072
	(0.07)
Trump	2.016
	(0.05)
Trump × Explicit	-0.103
	(0.07)
Controls	No
N	999

Heteroskedasticity-robust standard errors.

Note: The difference in differences estimate is 0.17 (p=0.097).

Table 19: Total Affirmations by Assigned Candidate, Candidate Preference, and Treatment

	Total Affirmations
Trump × Clinton-Voter	1.929
	(0.06)
Trump × Clinton-Voter × Explicit	-0.274***
	(0.09)
Trump × Trump-Voter	2.644
	(0.09)
Trump × Trump-Voter × Explicit	0.136
	(0.13)
Clinton × Clinton-Voter	2.464
	(0.06)
Clinton × Clinton-Voter × Explicit	0.078
	(0.09)
Clinton × Trump-Voter	2.000
	(0.09)
Clinton × Trump-Voter ×Explicit	0.017
	(0.13)
Controls	No
N	797

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Heteroskedasticity-robust standard errors.