The (Non) Religion of Mechanical Turk Workers

ANDREW R. LEWIS
Department of Political Science
University of Cincinnati

STEPHEN T. MOCKABEE
Department of Political Science
University of Cincinnati

PAUL A. DJUPE
Department of Political Science
Denison University

JOSHUA SU-YA WU
Department of Political Science
The Ohio State University

Social science researchers have increasingly come to utilize Amazon's Mechanical Turk (MTurk) to obtain adult, opt-in samples for use with experiments. Based on the demographic characteristics of MTurk samples, studies have provided some support for the representativeness of MTurk. Others have warranted caution based on demographic characteristics and comparisons of reliability. Yet, what is missing is an examination of the most glaring demographic difference in MTurk—religion. We compare five MTurk samples with a student convenience sample and the 2012 General Social Survey, finding that MTurk samples have a consistent bias toward nonreligion. MTurk surveys significantly overrepresent seculars and underrepresent Catholics and evangelical Protestants. We then compare the religiosity of religious identifiers across samples as well as relationships between religiosity and partisanship, finding many similarities and a few important differences from the general population.

Keywords: religion, Mechanical Turk, convenience sample, experiment, politics, seculars.

INTRODUCTION

Amazon's Mechanical Turk (MTurk) marketplace has increasingly become a useful resource for experimental researchers seeking adult convenience samples. Because it offers a low-cost, timely turnaround for survey responses (and other tasks), there has been a proliferation of especially experimental work in the social sciences using MTurk data, including work on religion (e.g., Johnson et al. 2012; McLaughlin and Wise 2014). While analyses of MTurk have rightly presented evidence of both its opportunities for social science research and necessary cautions, to date scholars have largely overlooked one aspect of MTurk samples—religion.[1] Turkers, on average, are decidedly more secular and less committed in their religious orientation than the population. Scholars, in particular those who study American religion, need to understand these distributional differences when contemplating turning to MTurk and when discussing their findings. More importantly, researchers need to know if relevant religious subgroups within MTurk samples look and act like their counterparts within the national population. It is to these tasks that we turn in this article.

[1]Berinsky, Huber, and Lenz in table 3 (2012:357) include religious affiliation among several demographic variables on which they compare MTurk samples to other nationally representative samples, but their discussion of religion in the text is limited to the following: "MTurk subjects are more likely to: have never married (51 percent), rent rather than own their home (53 percent), and report no religious affiliation (42 percent)" (2012:358).

Prior studies have supported the quality of what MTurk "workers" produce, finding that the workers are more representative of the national population than other types of convenience samples, including undergraduate student samples, while maintaining a high level of internal validity (Berinsky, Huber, and Lenz 2012; Buhrmester, Kwang, and Gosling 2011; Horton, Rand, and Zeckhauser 2011; Iyengar 2011). Related, MTurk workers have been shown to pay attention to survey items and prompts, passing screeners at high levels (Berinsky, Margolis, and Sances 2014). This is especially important for experimental research. Yet, a recent study suggests that caution is warranted when using MTurk (Krupnikov and Levine 2014). Concerns with MTurk include samples being wealthier, younger, more educated, less racially diverse, and more Democratic than national samples, as well as workers being savvy, frequent survey takers (Berinsky, Huber, and Lenz 2012). Most notably, there are external validity concerns, as results have been shown to diverge from theoretical expectations (Krupnikov and Levine 2014).

In what follows, we compare five MTurk samples to a nationally representative probability sample (the 2012 General Social Survey) and a student convenience sample. We find stark differences in the religious composition of MTurk workers—seculars are overrepresented by a factor of 2.5–3. However, the religious respondents on MTurk are largely representative of their peers in nationally representative samples (with a few exceptions). As such, there is appropriate variation to conduct reliable survey experiments, but where religion is expected to condition experimental treatment effects, scholars using MTurk may need to increase their sample sizes to capture enough religious respondents to enable appropriately powerful tests.

<div align="center">DATA</div>

Fielding a study on MTurk requires posting a "HIT" (human intelligence task) on the MTurk website. HITs are prepaid according to how much each worker will be paid (typically less than a dollar) and the number of responses needed. For our studies, Amazon charged a 10 percent fee. As of July 2015, however, Amazon now charges a 20 percent fee, plus an additional 20 percent fee for HITs with more than 10 assignments.[2] We marshal data from five separate MTurk surveys conducted in 2013 and 2014, yielding a total of 7,726 respondents across the surveys. Details of the different MTurk surveys are listed in Table 1. Surveys ranged in topic from religious and political hermeneutics, to immigration, culture wars, and foreign policy, and each was conducted under the supervision of the authors. Importantly, all included religion questions, and there is some diversity in the religious topics. Questions covered religious tradition (denominational affiliation), self-identification (born again), behavior (church attendance), and beliefs (biblical interpretation). The surveys also included political questions such as political party identification, ideology, and attitudes toward culture war issues (abortion and same-sex marriage). Finally, each survey asked for demographic data, including age, race, and sex of the respondents.

We compare the MTurk data to two other data sets with extensive religion and political variables. First, we compare the adult convenience samples of MTurk to a 2011 convenience sample of 1,151 undergraduate students from American government courses at four different universities in separate regions (Djupe et al. 2014). This is an important comparison, as student samples continue to be a frequent tool for researchers, and student samples may be the most natural "competitor" to MTurk. Second, we compare all the data to the commonly used and nationally representative 2012 General Social Survey (GSS).

We separate respondents into religious traditions, loosely following prior scholars (Kellstedt et al. 1996; Steensland et al. 2000). The Turk1 survey has a complete battery of denominational

---

[2]Those posting HITs, known as "requesters," are able to specify the kind and location of worker needed. Our studies confined workers to the United States (other countries are available) and then either a base number of HITs they have already completed (at least 100 HITs) or those with a particular satisfaction rate (generally above 95 percent). You can request MT "masters," though Amazon charges a higher rate for access to this specialized worker population.

Table 1: Data sources

Data set	Subject matter	Collected	N	Days open	Average minutes	Minimum HITs completed	User approval rating	Price paid
General Social Survey	General social/ political	Spring 2012	1,966	—	—	—	—	—
Student Survey	Political rights framing	Spring 2011– Fall 2011	1,151	—	—	—	—	—
Turk1	Politics and interpretation	Summer 2014	1,858	29	17	500	95%	$.50
Turk2	Immigration	Spring 2014	607	7	10	100	—	$.75
Turk3	Political polarization	Spring 2014	511	9	12	100	—	$.90
Turk4	Framing and abortion politics	Spring 2013	520	14	8	—	95%	$.72
Turk5	Religion and foreign policy	Summer 2014	4,230	17	7	—	90%	$.50

prompts, most closely mirroring Steensland et al.'s (2000) "RELTRAD" approach of using denominations to aggregate into religious traditions. The Turk2–4 surveys, along with the student survey, have a question about affiliation with general religious traditions, such as "Protestant," "Catholic," "Other Christian," and "None, nothing, secular." We categorize Protestants and Other Christians who also affirm a born-again identity as evangelical Christians. Turk5 takes a slightly different approach, including "evangelical Protestant" as an option in the selection of general traditions. The coding of all variables may be found online in Appendix 1, but we will provide two examples here of our most prominent variables: religious attendance and party identification. We collapse church attendance measures into three separate categories: (1) Never/rare; (2) 1–2 times per month; and (3) About weekly or more. For party identification, we also create three categories: (1) Democrats; (2) Independents; and (3) Republicans. We categorize independent leaners with the party they lean toward.

We expect very little overlap in survey taking in the general population. It turns out, however, that there is substantial overlap in those taking MTurk surveys, with 36% (2,787) taking more than one. This astounding result indicates that the MTurk population is rather limited in size and reinforces the image that these are professional survey takers. But, the overlap is a boon in this case, as it allows us to analyze the consistency of their responses. From our data, we were able to identify 281 matched pairs who were asked the same questions.[3] We performed tetrachoric correlations for the binary variables and polychoric correlations for the ordinal variables. We found strong consistency in responses ($p < .01$) for evangelical (rho = .75), Catholic (rho = .97), secular (rho = .98), and party identifications (rho = .93).[4] We conclude that the MTurk respondents can be trusted to represent their religious and political identifications consistently—they are not simply clicking random buttons to get paid.

RESULTS

We begin with descriptive comparisons of religious affiliations. Figure 1 shows that the MTurk samples have far fewer religious respondents and more than twice as many seculars as compared to the GSS. Using RELTRAD, evangelicals make up 24 percent of the GSS sample, while evangelicals are only 10–14 percent of the five MTurk samples. These results are consistent whether we classify evangelicals via denominational affiliation (Turk1), the combination of Protestant/Other Christian affiliation and born-again identity (Turk2–Turk4), or Evangelical Protestant identity (Turk5). Though they are different than in the GSS, the number of MTurk evangelicals is similar to our student sample (15 percent). The MTurk samples also include about half as many Catholic respondents, 13–17 percent as compared to 25 percent in the GSS. The percentage of Catholics in our student sample (27 percent) is also much greater than the MTurk samples, largely matching the GSS. While evangelicals and Catholics are underrepresented on MTurk, seculars are overrepresented. The percentage of those without a religious affiliation on MTurk ranges from 45 to 54 percent, more than double the GSS (22 percent) and our student sample (23 percent). In our surveys of over 7,700 "Turkers," nearly half identify as religiously

---

[3]The number of matched pairs we use is lower than the full overlap because the Turk1 survey did not record their Worker ID during the survey. Therefore, while we were able to identify the MTurk workers that took the survey, we were unable to match them to their responses. Nevertheless, our 281 matched pairs provided a strong test for response consistency.

[4]The evangelical correlation, while still statistically significant, is lower than the others. We attribute this to variance in question wording, as our MTurk surveys used denominational affiliation, evangelical Protestant identity, or a combination of Protestant and born- again identity to classify evangelicals.

Figure 1
Representation of religious traditions in MTurk samples
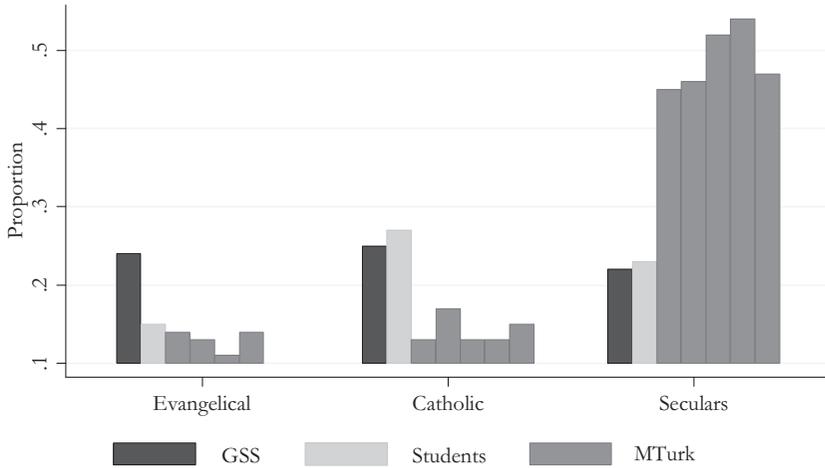
unaffiliated. These differences in religious affiliation are statistically ($p < .01$) and substantively significant and should be considered carefully by scholars doing research on MTurk.[5]

As one would expect with a large percentage of seculars, Turkers report lower levels of religious behaviors and beliefs (see Figure 2). In the GSS and the student sample, 31 percent of respondents reported attending religious services weekly, but this percentage drops to between 10 and 22 percent for the MTurk samples. Rare and nonattendance is about 20–30 percent higher on MTurk, ranging from 71 to 84 percent.[6] Lower rates of attendance on MTurk mean that samples are less likely to have been exposed to messaging from religious elites and have religious networks, two religious factors that have been shown to affect opinion and behavior (Djupe and Gilbert 2009). Predictably, the MTurk sample is also more skeptical of the Bible ($p < .01$). Only one of our surveys, Turk1, measured biblical literalism, but 56 percent view the Bible as either "Written by Men" or a "Book of Fables" compared to 21 percent of the GSS sample that selected "Book of Fables." Low levels of religiosity permeate MTurk samples.

It might be proffered that the low levels of religiosity on MTurk are attributable to demographic differences, especially since Turkers are known to be younger than the general population (Berinsky, Huber, and Lenz 2012). If this is the case, then controlling for these demographic differences could yield a comparable sample. Like others have found, our MTurk samples are younger and more male than nationally representative samples. The average age for the MTurk surveys is 32–36 years old, at least 25 percent younger than the GSS mean (48). In the GSS and the student survey, 44 and 43 percent of the respondents, respectively, are male, while the MTurk samples range from 57 to 61 percent. The percentages of whites and nonwhites on MTurk, however, are comparable to those in the GSS.

While these demographic differences are important and should be considered when analyzing MTurk data, the vast differences in religiosity cannot be accounted for by demographics alone. We can measure this by modeling evangelical identity using age in the GSS. The GSS predicts a 21.4 percent probability of being evangelical at age 32 (mean age for Turk3–Turk5), a

---

[5]We use 2-tailed t-tests and z-tests and effects size calculations to compare the results of the MTurk surveys with the GSS. The p-values are indicated in parentheses. A table of all the p-values and the Cohen's d statistics are available in Appendix 2 of the online supplementary materials.

[6]The differences in the average attendance rates between each MTurk survey and the GSS are statistically significant ($p < .01$).

Figure 2
Sample-level worship attendance rates and partisanship in MTurk and other samples
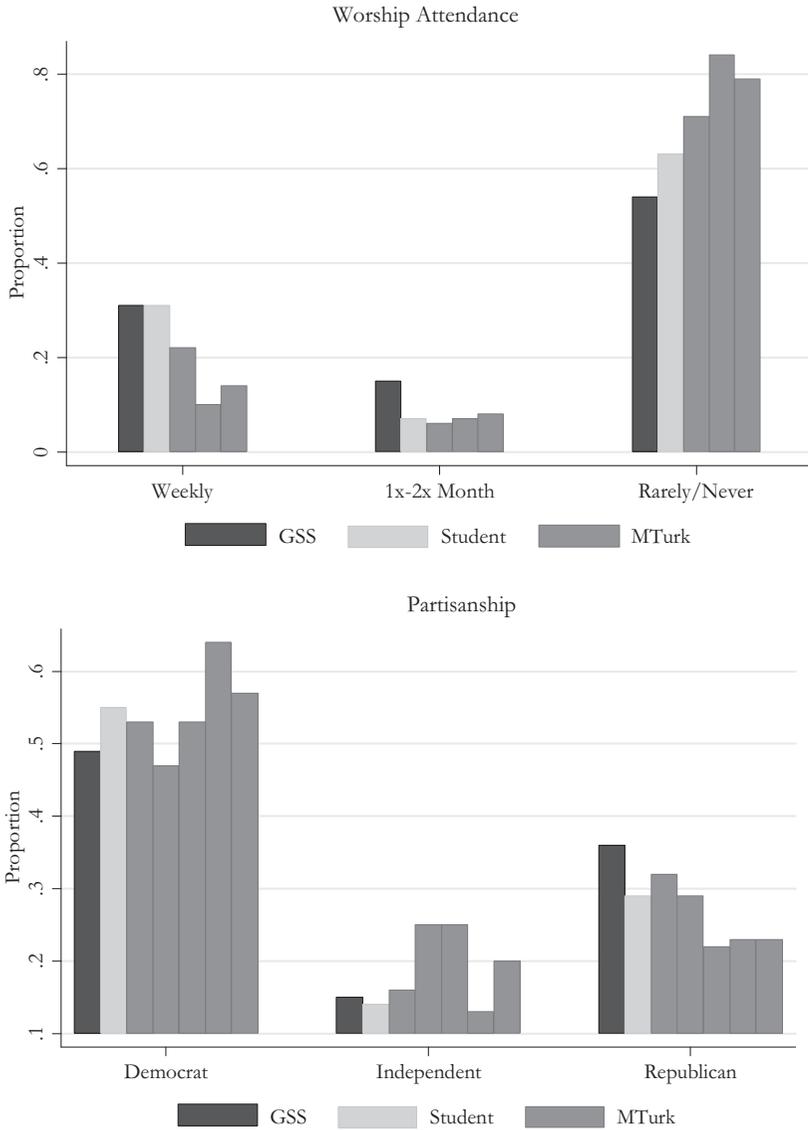
22.1 percent probability at age 36 (mean age for Turk1 and Turk2), and a 24 percent probability of being evangelical at age 48 (mean age for the GSS). Thus, applying age differences in evangelical identity from the 2012 GSS serves to reduce the percentage of evangelicals by about 1.6 points, accounting for only about 10 percent of the 10–14 point difference in evangelical affiliation seen between the GSS and the MTurk samples. Likewise, women are 6 percent more likely to be evangelical than men in the GSS; correcting the MTurk samples for the underrepresentation of women by about 12–16 percent only adds about a percentage point to the evangelical column.[7]

[7]Related, the MTurk samples also have higher levels of education than the GSS. For example, the Turk1 sample has 19 percent more college-educated individuals than the GSS. If Turk1 had the same percentage of college graduates as the GSS, the percentage of evangelicals should climb approximately 2 percent, as non-college-educated individuals are

Beyond the religious and demographic differences, there are important political differences between the MTurk and GSS samples. MTurk workers are more liberal than the GSS sample, though comparable to the student sample. The lower panel in Figure 2 shows that, with the exception of Turk2, the MTurk samples have more Democrats and fewer Republicans than the GSS ($p < .01$). The differences are particularly evident in the social issue questions. In Turk1 and Turk4, 69–77 percent of the respondents favor abortion rights, compared to only 55 percent in the GSS. Even more starkly, 69 percent (Turk1) and 83 percent (Turk4) of MTurk respondents favor same-sex marriage, compared to only 50 percent in the 2012 GSS. All of these differences are statistically significant ($p < .01$).

The MTurk samples are clearly less religious and more liberal than both nationally representative and student samples. Yet if the religious respondents from MTurk bear similar characteristics to the religious respondents in more representative surveys, then MTurk can still be useful for research on religion, albeit with larger samples to provide sufficient subsamples of religious groups to analyze. To investigate this, we examine the religiosity and political opinions of evangelical, Catholic, and secular respondents from MTurk, the student sample, and the GSS.

For religiosity, the results for seculars are consistent across the MTurk, GSS, and student surveys, with the exception of more GSS respondents having a more traditional view of the Bible. But, as Figure 3 shows, MTurk evangelicals and Catholics have lower levels of religiosity than their GSS counterparts. In the GSS, 53 percent of evangelicals and 30 percent of Catholics attend church weekly, while the MTurk samples range from 39 to 43 percent for evangelicals and 12–31 percent for Catholics. On MTurk, there is also a 30–40 percent increase in those evangelicals and Catholics who rarely or never attend church. Though still significant in two out of three MTurk surveys ($p < .01$), the in-group differences between MTurk and the GSS are less stark than when comparing the complete samples, but the MTurk intratradition samples are closer to the student sample in their religious attendance.

Evangelical and Catholic MTurk workers also have a less traditional view of the Bible than the GSS respondents (evangelicals: $p < .01$; Catholics: $p = .05$). The majority of the GSS evangelicals (59 percent) view the Bible as the "actual word of God," whereas the majority of Turk1 evangelicals view the Bible as "inspired, but not literal" (55 percent). For Catholics, the differences are at the poles. In the GSS, 23 percent of Catholics view the Bible as being the "actual word of God," and only 15 percent view it as being "written by men." Those percentages are essentially reversed among Turk1 Catholics, with 10 percent taking the literalist view and 28 percent the humanist view.

While there are intratradition religiosity differences in the MTurk samples, the political results are largely similar to the GSS and the student sample, with no statistically significant differences ($p \leq .05$). The upper-right panel of Figure 3 compares the party identification of evangelicals, where MTurk Democrats range from 27 to 35 percent and compare favorably to 31 percent in the GSS. The student sample is the outlier, with 51 percent identifying with the Democrats. Party identification is also fairly consistent among Catholics, averaging 42–50 percent Democrats on Turk and 51 percent in the GSS.

Views toward abortion are also fairly consistent across the samples for evangelicals, with 66 percent of GSS evangelicals and 62–68 percent of MTurk evangelicals taking a pro-life stance. The smaller MTurk sample (Turk4) has a significant difference for evangelicals ($p = .02$). The bigger differences are for Catholics, who are less pro-life on MTurk—32–39 percent as compared to 47 percent in the GSS ($p < .01$ for both samples). The biggest difference, however, is regarding evangelicals' views on same-sex marriage, where the MTurk samples are at least 40 percent more likely to support same-sex marriage ($p < .01$). Twenty-five percent of 2012

---

11 percent more likely to be evangelicals. Demographic factors account for some, but not all, of the religious difference between MTurk and the GSS.

Figure 3
Comparison of the partisanship and worship attendance among evangelicals and Catholics from
MTurk and other samples
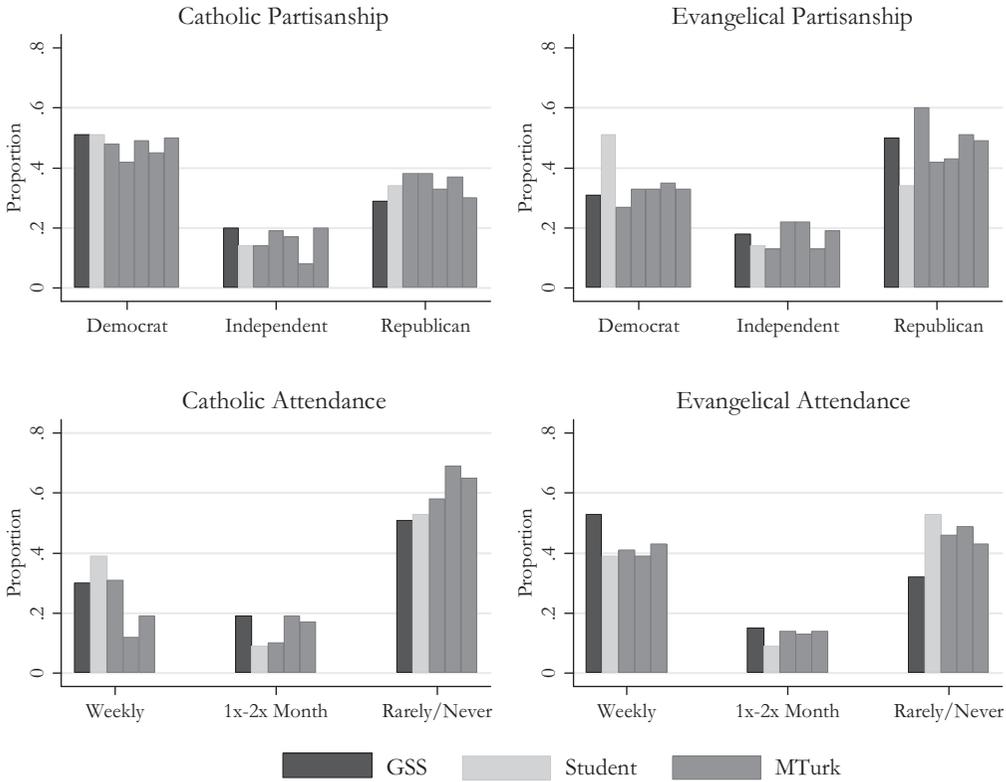
GSS evangelicals support gay marriage, while 35 percent of evangelicals in Turk1 and 42 percent in Turk4 support it. For MTurk Catholics there remain differences ($p < .01$), though less stark. Approval of same-sex marriage is between 61 and 82 percent in the MTurk samples, while 57 percent of GSS Catholics support gay marriage. Therefore, while the political identities within religious traditions are comparable between MTurk and the GSS, there is a cleavage regarding same-sex marriage.

Though the distributions in MTurk samples may not perfectly match the nation, it is still possible that the *relationships* between variables within MTurk samples parallel those in representative national samples. We assessed this with a simple ordered logit model of partisanship that employs "model-based weighting" (Berinsky 2006), including age (in years), gender, and race (white vs. nonwhite), so we can reliably assess the effects of worship attendance and an evangelical identity. The effects of the full range of these variables on the probability change of each partisan category for a GSS and an MTurk sample is shown in Figure 4. There, the results from the two data sets are broadly compatible; the effects of attendance, evangelical identity, and gender are somewhat more pronounced in the MTurk sample, whereas the differences between whites and nonwhites are less stark in the MTurk data. The age effects are inconsistent but also insignificant in both data sets. Thus, the results are consistent if not a perfect match. The ideal test of the relative efficacy of MTurk, however, would be to compare the results of an identical experiment in nationally representative data.

Figure 4
Comparing effects of independent variables on partisanship in the GSS and an MTurk sample
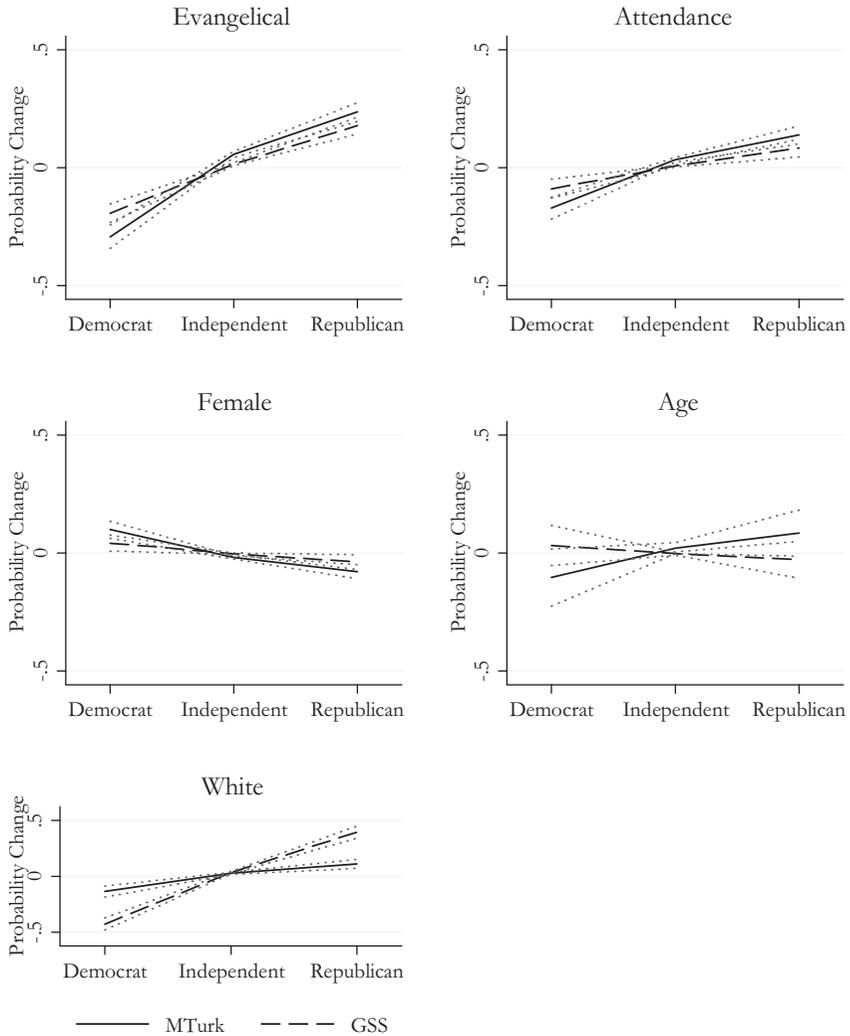
CONCLUSION

While Amazon's Mechanical Turk marketplace has many advantages, scholars need to understand and account for its limitations regarding religion, the most glaring demographic difference. Across our five studies, approximately half of all respondents are secular, more than double the percentage of seculars in the GSS and a large student sample. As such, there is a nonreligious bias on MTurk for which scholars need to account. If researchers expect effects of their treatments to be conditional on religion, then samples need to be larger to capture more religious individuals. While there are necessary cautions about the religious makeup of MTurk samples, the religious respondents on MTurk are similar in their politics to religious respondents in the GSS and student samples, with the notable exception of attitudes about same-sex marriage. Therefore, with appropriate controls, the relationships we find between religion and political variables are not far off the mark of a national sample. In addition, we found that the samples are reasonably stable representations of religious individuals on MTurk and MTurk workers represent their religious

identifications reliably. Together, these facts suggest there is enough reliable variation for survey experiments, which remain the best use of MTurk.

Finally, it should be noted that the large proportions of seculars in MTurk samples could facilitate research on the attitudes and behavior of the religiously unaffiliated. Because MTurk samples consistently should be expected to yield large numbers of secular respondents, it is efficient to conduct experiments on MTurk that aim to study this growing segment of the public. Moreover, the high percentage of seculars who participate in MTurk experiments can enable the use of experimental design features such as blocking (random assignment of respondents to treatment and control groups within each religious tradition) that allow for more efficient identification of treatment effects and better ensure covariate balance between respondents in treatment and control groups.

REFERENCES

Berinsky, Adam J. 2006. American public opinion in the 1930s and 1940s: The analysis of quota controlled sample survey data. *Public Opinion Quarterly* 70(4):499–529.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis* 20(3):351–68.

Berinsky, Adam J., Michele F. Margolis, and Michael W. Sances. 2014. Separating the shirkers from the workers? Making sure survey respondents pay attention on self-administered surveys. *American Journal of Political Science* 58(3):739–53.

Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. Amazon's MechanicalTurk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science* 6(1):3–5.

Djupe, Paul A., and Christopher P. Gilbert. 2009 *The political influence of churches*. New York: Cambridge University Press.

Djupe, Paul A., Andrew R. Lewis, Ted G. Jelen, and Charles D. Dahan. 2014. Rights talk: The opinion dynamics of rights framing. *Social Science Quarterly* 95(3):652–68.

Horton, John J., David G. Rand, and Richard J. Zeckhauser. 2011. The online laboratory: Conducting experiments in a real labor market. *Experimental Economics* 14(3):399–425.

Iyengar, Shanto. 2011. Laboratory experiments in political science. In *Handbook of experimental political science*, edited by James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia, pp. 73–88. New York: Cambridge University Press.

Johnson, Megan K., Jordan Paul Labouff, Wade C. Rowatt, Julie A. Patock-Peckham, and Robert D. Carlisle. 2012. Facets of right-wing authoritarianism mediate the relationship between religious fundamentalism and attitudes toward Arabs and African Americans. *Journal for the Scientific Study of Religion* 51(1):128–42.

Kellstedt, Lyman A., John C. Green, James L. Guth, and Corwin E. Smidt. 1996. Grasping the essentials: The social embodiment of religion and political behavior. In *Religion and the culture wars: Dispatches from the front*, edited by John C. Green, James L. Guth, Corwin E. Smidt, and Lyman A. Kellstedt, pp. 174–92. Lanham, MD: Rowman and Littlefield.

Krupnikov, Yanna, and Adam Seth Levine. 2014. Cross-sample comparisons and externalvalidity. *Journal of Experimental Political Science* 1(1):59–80.

McLaughlin, Bryan, and David Wise. 2014. Cueing God: Religious cues and voter support. *Politics and Religion* 7(2):366–94.

Steensland, Brian, Jerry Z. Park, Mark D. Regnerus, Lynn D. Robinson, W. Bradford Wilcox, and Robert D. Woodberry. 2000. The measure of American religion: Toward improving the state of the art. *Social Forces* 79(1):291–318.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix 1. Variable Coding
Appendix 2. Turk Samples Compared to GSS, z-Tests and Effects Sizes