

“Who are these people?” Evaluating the demographic characteristics and political preferences of MTurk survey respondents

Research and Politics
July–September 2015: 1–12
© The Author(s) 2015
DOI: 10.1177/2053168015604648
rap.sagepub.com


Connor Huff and Dustin Tingley

Abstract

As Amazon’s Mechanical Turk (MTurk) has surged in popularity throughout political science, scholars have increasingly challenged the external validity of inferences made drawing upon MTurk samples. At workshops and conferences experimental and survey-based researchers hear questions about the demographic characteristics, political preferences, occupation, and geographic location of MTurk respondents. In this paper we answer these questions and present a number of novel results. By introducing a new benchmark comparison for MTurk surveys, the Cooperative Congressional Election Survey, we compare the joint distributions of age, gender, and race among MTurk respondents within the United States. In addition, we compare political, occupational, and geographical information about respondents from MTurk and CCES. Throughout the paper we show several ways that political scientists can use the strengths of MTurk to attract respondents with specific characteristics of interest to best answer their substantive research questions.

Keywords

Survey Research, Mechanical Turk, Experimentation

Introduction

In the last several years Amazon’s Mechanical Turk (MTurk) has surged in popularity in experimental and survey-based social science research (Berinsky et al., 2012; Chandler et al., 2014; Krupnikov and Levine, 2014; Paolacci and Chandler, 2014). Researchers have used the results from MTurk surveys to answer a wide array of questions ranging from understanding the limitations of voters to exploring cognitive biases and the strengths of political arguments (Arceneaux, 2012; Grimmer et al., 2012; Huber et al., 2012). As this type of work has grown in popularity, researchers hear an increasing number of important questions at workshops and conferences about the external validity of the inferences made drawing upon MTurk samples. Questions such as: “Are your respondents all young White males?”, “Do any of them have jobs?” and “Where do these people live?” are rightfully voiced. In this paper we seek to answer some of these questions by unpacking the survey-specific respondent attributes of MTurk samples.

Berinsky et al. (2012) take an important first step in exploring the validity of experiments performed using MTurk. They show that while respondents recruited via MTurk are often more representative of the US population than in-person convenience samples, MTurk respondents are less representative than subjects in Internet-based panels or national probability samples. Berinsky et al. (2012) reach this conclusion by comparing MTurk to convenience samples from prior work (Berinsky and Kinder, 2006; Kam et al., 2007) and the American National Election 2008–2009 Panel Study. In their paper, Berinsky et al. (2012) assess numerous characteristics of MTurk respondents that are of interest to political scientists. These variables include

Department of Government, Harvard University, USA

Corresponding author:

Connor Huff, Department of Government, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138, USA.
Email: cdezzanilhuff@fas.harvard.edu



party identification, race, education, age, marital status, religion, as well as numerous other variables of interest. The comparisons presented by Berinsky et al. (2012) provide an excellent foundation for exploring the relationship between samples drawn from MTurk surveys and other subject pools commonly used by political scientists.¹

In this paper we present a number of results that contribute to a broader goal of understanding survey data collected from platforms such as MTurk. In doing so we provide a framework that will allow social science researchers, who frequently use this platform, to better understand the characteristics of their respondent pools and the implications of this for their research. This paper builds upon Berinsky et al. (2012) to make four contributions. First, in Section 2 we present a new benchmark comparison for MTurk surveys: the Cooperative Congressional Election Survey (CCES).^{2,3} The CCES is a nationally stratified sample survey administered yearly from late September to late October. The survey asks about general political attitudes, demographic factors, assessment of roll call voting choices, and political information.⁴ In this paper we present the results of a simultaneous MTurk and CCES survey.^{5,6} Unlike in Berinsky et al. (2012), this design allows us to focus our comparisons on the similarities and differences between CCES and MTurk samples at a common point in time.

Second, we provide a partial picture of the joint distributions of a number of demographic characteristics of interest to social science researchers. Berinsky et al. (2012) take an important first step toward understanding the racial, gender, and age characteristics of MTurk samples by reporting the percentage of respondents in each of these categories. However, they do not explore the relationship between these key variables of interest. By presenting the joint distributions of several of these variables we are able to analyze the properties of MTurk samples within the United States as they cut across these different categories. For example, we show that MTurk is excellent at attracting young Hispanic females and young Asian males and females. In contrast, MTurk has trouble recruiting most older racial categories and is particularly poor at attracting African Americans. We focus on race, gender, and age as these are some of the most prominent attributes of respondents across which researchers might expect to observe heterogeneous treatment effects.⁷ This means that providing information about the number of respondents within each of these categories of interest, and how this differs from other prominent survey platforms, can assist researchers in both the design and interpretation of their experimental results.

Third, we compare the political characteristics of respondents on MTurk and CCES. In Section 3 we show how the age of respondents interacts with voting patterns, partisan preferences, news interest, and education.⁸ We demonstrate that, on average, the estimated difference between CCES and MTurk markedly decreases when we subset the data to younger individuals. In Section 4, we

compare the occupations of MTurk and CCES respondents. We show that the percentage of respondents employed in a specific sector is similar across both platforms, with a maximum difference of less than 7%. For example, the percentage of respondents employed as “Professionals” is in the range of approximately 12–16% across both surveys. These results show that MTurk and CCES have a similar proportion of respondents across industry. In Section 5, we present geographic information about respondents. We show that the number of respondents living in different geographic categories on the rural–urban continuum is almost identical in MTurk and CCES. Both MTurk and CCES draw approximately 90% of their respondents from urban areas. Using geographic data from the surveys, we map the county-level distribution of respondents across the country.

Finally, we discuss how researchers can build “pools” of prior MTurk respondents, recontact these respondents using the open-source R package MTurkR,⁹ and then use these pools to over-sample and stratify to create samples that have desired distributions of covariates. This is a useful tool for social science researchers because it allows them to directly stratify on key moderating variables.¹⁰ By drawing on the strengths and weaknesses of MTurk samples and cutting-edge research tools such as MTurkR, researchers can use similar sampling strategies to those of professional polling firms to directly address concerns about the external validity of their survey research.

Age, gender, and race: exploring the joint distributions of key demographic characteristics

In this section we compare distributions of basic demographic variables in a CCES team survey¹¹ and a survey conducted on MTurk at the same time during the fall of 2012. The MTurk survey had 2706 respondents and the CCES had 1300. The questions in both surveys were asked in the exact same ways, though the CCES survey respondents were also asked additional questions.¹²

Obtaining a survey sample with the desired racial, age, or gender characteristics is a difficult endeavor that has persistently challenged the external validity of research. For example, scholars have frequently debated the quality of inferences when the results are drawn from college-age convenience samples (Druckman and Kam, 2011; Peterson, 2001). Some argue that research must be replicated with non-student subjects before attempting to make generalizations. Experimentalists push back and invite arguments about why a particular covariate imbalance would moderate a treatment effect. We argue in this paper that insofar as this debate plays out with respect to MTurk, we should have detailed information about what exact covariate imbalances actually exist.

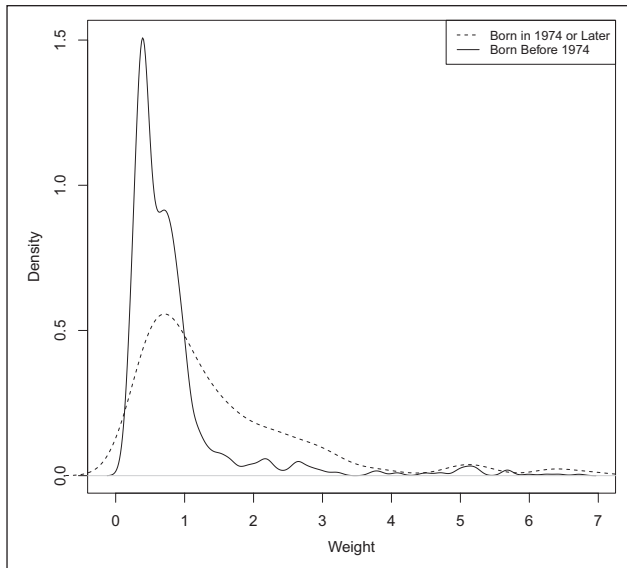


Figure 1. Survey weights for different age cohorts in the CCES data.

In the survey research tradition there are a variety of methods for achieving a “nationally” representative poll. For example, the CCES creates a nationally representative sample of US adults using approximate sample weights from sample matching on registered and unregistered voters. This means that in order to generalize to the target population of US adults the CCES must weight respondents with certain background characteristics more heavily than others.¹³ Figure 1 shows the survey weights placed on individuals in different age brackets. The results demonstrate how the CCES up-weights younger individuals while down-weighting older individuals.¹⁴ The cutpoint for age is found by taking the mean of *all* the data (including CCES and MTurk). This method is used since we want to directly compare individuals in the different age categories across MTurk and CCES. The results do not change when using other similar cutpoints. In the remainder of the paper we will not use the CCES survey weights.¹⁵ Individuals could always construct weights for MTurk samples. By ignoring weights we get to observe the underlying differences in the unweighted samples.

In Figure 2 we get a sense of the joint distributions of three key variables: age, gender, and race. The mosaic plots show, for each racial category, the proportion of respondents that are male or female and young or old. For example, the first row of mosaic plots show for individuals of all races, the proportion that are older females, older males, younger females, and younger males. If the width of a box under female is larger than for male, this means that there is a large proportion of females within that particular race. Similarly, if a box is taller for younger than for older individuals, this means that there is a larger proportion of younger than older individuals of a particular race represented in the sample.

Figure 2 demonstrates that the young individuals weighted most heavily by CCES are often the same categories that MTurk was best at attracting.¹⁶ We can see that approximately 75% of all respondents in CCES and MTurk were White. Figure 2 also demonstrates differences in the CCES and MTurk samples with respect to African American, Hispanic, and Asian respondents. For example, MTurk is able to attract between 2% and 5% more Hispanic and Asian respondents.¹⁷ In contrast, CCES is approximately 6% better at recruiting African-American respondents. We can take this analysis a step further by exploring the joint distributions of age, gender, and race. For example, we can see that in all racial categories MTurk attracts a large number of young respondents with this contrast at its starkest among young Asian males.¹⁸

Researchers could leverage the differential abilities of survey pools to attract respondents with demographic characteristics most suited to answering their theoretical question of interest.¹⁹ Just as scholars select the methodological tools most suited to addressing their question, the same logic can be applied to choosing between survey pools. Recognizing the differential abilities of MTurk and CCES to recruit specific individuals of particular demographic characteristics is an important step. For example, Figure 2 demonstrates that MTurk is an excellent resource for exploring the opinions of Young Asian and Hispanic Males. However, CCES might be a better choice for exploring the opinion of Male African-Americans. As experimental and survey-based research continues to surge in popularity political scientists can and should take advantage of these strengths and weaknesses of MTurk survey pools.

Party ID, ideology, news interest, voting, and education

In this section we explore the interaction between age and several variables commonly used in political science research. These include: (1) voter registration; (2) voter intentions; (3) ideology; (4) news interest; (5) party identification; and (6) education. In doing so, we build upon the work of Berinsky et al. (2012) by exploring the interaction of these variables with age. Using regression we demonstrate that, on average, the estimated difference between CCES and MTurk decreases when we subset the data to younger individuals. This means that when researchers are considering the dimensions along which they might expect to find heterogeneous treatment effects they should be cognizant of the ways in which older respondents differ across survey platforms. The regression estimates with standard errors are presented in Figure 3.²⁰

Figure 3 depicts several differences across the two survey platforms. First, voting registration and intention to turnout patterns among younger respondents are very similar for both CCES and MTurk. In contrast, older respondents in MTurk turnout and vote less than individuals of a

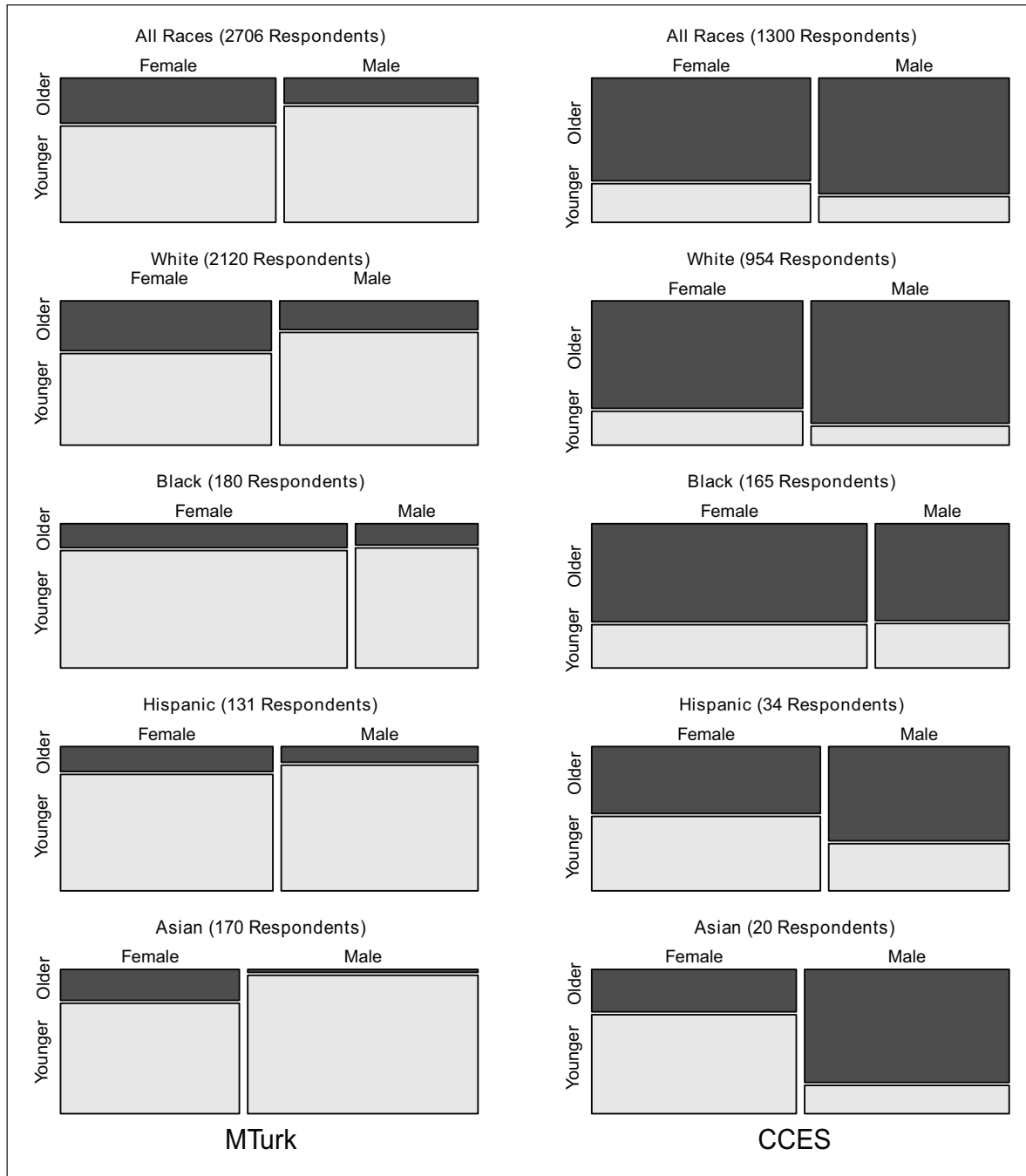


Figure 2. Mosaic plots showing the gender and age composition for different racial categories in the CCES and MTurk modules.

similar age from the CCES. For party identification, which was measured on a seven-point scale ranging from Strong Democrat to Strong Republican, we again observe that younger respondents are more similar for both CCES and MTurk. For older individuals the respondents in MTurk are consistently more liberal than CCES. Somewhat similar trends hold for ideology. The level of news interest, which varies from most of the time to hardly at all, between respondents in MTurk and CCES varies dramatically. Older

individuals in MTurk are less interested in the news than older individuals from CCES. In contrast, younger MTurk respondents are more interested in the news than younger individuals from CCES. Finally, we can see that there are not substantial differences in the levels of education between younger and older MTurk and CCES respondents.

We can draw a number of conclusions from these results. First, the similar registration and intention to vote patterns of CCES and MTurk respondents shows that

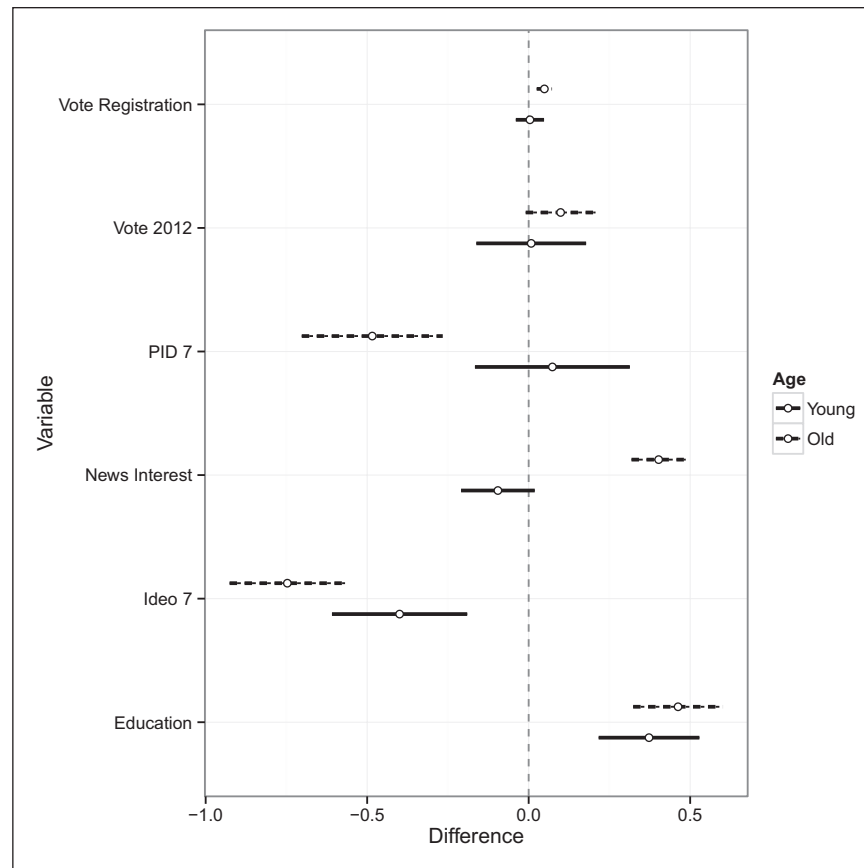


Figure 3. Differences in means with 95% confidence intervals for the proportion of respondents registered to vote, proportion of respondents that intend to vote in 2012, party identification, ideology, level of news interest, and education level in the CCES and MTurk modules. Positive values indicate that MTurk is greater than CCES. Dashed lines correspond with the confidence intervals for older respondents and solid lines for younger.

MTurk could be an excellent means for exploring how experimental manipulations could influence voting tendencies. As we showed in the previous section, these manipulations could be targeted at particular demographic groups such as young Hispanic or Asian respondents. Second, MTurk provides a useful means for attracting young respondents interested in the news. This means that MTurk could be used by political scientists to build upon prior research exploring the complex relationship between news interest, political knowledge, and voter turnout (Philpot, 2004; Prior, 2005; Zaller, 1992). The regression results presented in Figure 3 provide a means for political scientists to more fully understanding the external validity of MTurk surveys and also showing the strengths of MTurk for exploring a number of substantive questions of interest to political science researchers.

What do they do? The occupations of MTurk respondents

One of the most common questions we hear at workshops and conferences is about the occupational categories of

MTurk respondents. Many scholars are rightfully concerned that MTurk respondents might all be unemployed or overwhelmingly draw from a small number of industries. Depending on the particular research question, these differences could interact with our experimental manipulations in significant ways. Thus, the occupation of MTurk respondents would be fundamentally different from that of other sectors of the population about which they are trying to make inferences. However, in this paper we show that the percentage of MTurk respondents employed in specific industries is strikingly similar to CCES.²¹ For example, we can see that the percentage of individuals employed as Professionals ranges from approximately 12% to 16% for CCES and MTurk. Indeed, in the 14 sector-specific occupation categories we compare the maximum difference between MTurk and CCES is less than 6%. We can see this difference in the “Other Service” sector of Table 1 where 16.01% of individuals are employed in “Other Service” in MTurk while there are 21.47% in CCES.²² The results presented in Table 1 should be reassuring to political scientists concerned that the occupation of MTurk respondents is fundamentally different than other survey pools. Table 1

Table 1. The occupation of respondents by survey.

Occupation	CCES (%)	MTurk (%)
Management	15.69	11.94
Independent contractor	4.47	8.72
Business owner	4.91	2.73
Owner–operator	2.45	1.92
Office and administrative support	13.50	17.24
Healthcare support	3.77	4.56
Protective service	1.93	1.22
Food preparation and service	2.80	6.03
Personal care	1.75	2.44
Installation, maintenance and repair	2.89	2.93
Grounds cleaning and maintenance	0.88	0.81
Other service	21.47	16.01
Trade worker or laborer	8.50	9.25
Professional	14.99	14.18

demonstrates the occupational similarities between MTurk and CCES.

Where do respondents live? The urban–rural continuum

Researchers might also be concerned that MTurk respondents are overwhelmingly drawn from either urban or rural areas. This, again, may or may not matter for estimating the effect of an experimental manipulation depending on the research question, but as with employment characteristics it is useful to know. In both the MTurk and CCES data we have self-reported zip codes. We then link this data up with the United States Department of Agriculture (USDA) rural–urban continuum classification scheme to analyze the geographic characteristics of survey respondents.²³ These classification codes range from metro areas coded 1–3 in decreasing population size, to non-metro areas coded from 4 to 9. In Table 2 we show that the number of respondents living in different geographic categories on the rural-urban continuum is almost identical in MTurk and CCES.²⁴ Both MTurk and CCES draw approximately 90% of their respondents from urban areas with the remaining 10% spread across rural areas. For example, we can see that between 52% and 57% of respondents have a rural-urban code of 1 which means they live in counties in metro areas of 1 million or more. In contrast, less than 2% of respondents have a rural-urban code of 9 meaning that they live in a location that is completely rural. The rural-urban comparison of CCES and MTurk is presented in Table 2.

Table 2 shows that MTurk and CCES respondents live in similar geographic locations on the rural-urban continuum. This means that social science researchers should not be concerned that MTurk respondents are overwhelmingly drawn from either urban or rural areas in a way that might bias their results, compared to what they would get from a

Table 2. The percentage of respondents in urban/rural areas by survey.

Urban–rural code	CCES (%)	MTurk (%)
1	53.60	57.13
2	23.03	22.91
3	8.91	7.72
4	4.87	4.76
5	1.48	1.01
6	5.03	2.82
7	1.75	2.53
8	0.36	0.62
9	0.95	0.49

major professional polling firm. In Appendix B we present a map showing the distribution of respondents at the county level in the MTurk sample across the United States (Figure 5). Political scientists can explore the geographic distribution of their respondents using this paper’s replication files cross applied to their own studies. If, for example, an overwhelming number of respondents are drawn from a particular state or county we will be able to view this on the map.

The similarities between the occupation and rural-urban location of respondents from MTurk and CCES has implications for experimental and survey-based research. For example, political economists exploring preferences over trade, immigration, and redistribution, for which occupation and location are of critical importance, can consider using MTurk and not be concerned that their respondents are overwhelmingly drawn from particular occupations or geographic locations that look different from what professional poll sampling would yield. These results provide a first response to questions frequently raised at workshops and conferences about whether the geographic and employment characteristics of MTurk respondents are fundamentally different from other survey pools.

Developing survey pools

Researchers can use MTurk to build “pools” of prior MTurk respondents that they can then use in several different ways for future surveys. This is done by first having a MTurk respondent take a survey where the researcher records variables of interest, such as age, race, gender, and party, and then match these characteristics to the unique identification number possessed by every MTurk respondent. Once this pool is developed researchers can use the open-source R package MTurkR to recontact their prior respondents.²⁵ MTurkR has the potential to revolutionize online experimental and survey research as political scientists can use this package for over-sampling or stratifying on crucial variables of interest such as party or gender.

There are two main techniques researchers can use to build pools. In the first technique, researchers can pool

across respondents from their prior MTurk surveys. Since researchers commonly ask the same battery of questions about the demographic characteristics of their respondents, they can use these characteristics to then stratify on variables of interest. For example, over time, we have collected a large pool of MTurkers that have taken our surveys and told us their gender, ideology, partisan affiliation, and zip code. In this sample of 15,584 MTurkers, 54% were male, on a one- to seven-point ideology scale the average was 3.35, 34% self-identified as Democrat, 22% as Republican, and 26% as independent (the remaining identified with “other” parties), and the average age was 32.²⁶ This technique of pooling across multiple surveys is most useful for researchers that conduct a high volume of surveys on MTurk.²⁷

A recent strain of research exploring the characteristics of MTurk workers argues that what differentiates MTurkers is their status as permanent participants (Chandler et al., 2014; Krupnikov and Levine, 2014; Paolacci and Chandler, 2014). A potential concern with permanent participants is that they have taken a number of similar studies which can then subsequently affect the ways in which they both answer questions and respond to treatment conditions.²⁸ Moreover, the use of high-volume survey takers has the potential to undermine some of the assumptions of experimental research methods.²⁹ We view the ability to build pools of respondents as a way to potentially address this concern.³⁰ Since researchers that build pools have data on the number of times an individual has taken their prior surveys, they could build information into their pool about the types of respondents that are “high-volume” takers and then test for heterogeneous treatment effects. The assumption here is that respondents that take a high-volume of surveys are less likely to be naive workers, and more likely to appear in prior surveys with a higher frequency. Researchers can then incorporate this information about their respondents to test for whether treatment differentially affects MTurk workers that have taken a higher frequency of prior surveys.

In the second technique, researchers create a pool by first creating a HIT that oversamples respondents and asks a small battery of questions upon which the researcher would like to subsequently stratify. They then use this new pool to recontact respondents with the desired attributes of interest. This technique is useful to researchers that conduct an infrequent number of surveys as they have likely not built up a pool of adequate size to be able to pool across multiple surveys to directly recontact respondents. Moreover, this two-stage sampling procedure allows researchers to recruit respondents over a relatively short timeframe. Gay et al. (2015) provide a concrete example of how this could be done in practice in order to address concerns about not being able to obtain enough non-White respondents. Using a two-stage sampling procedure, they first recruited 1940 respondents to take a demographic survey. From these 1940 respondents, they then recontacted a sample that included all of the Black, Hispanic, and Asian respondents from this

initial survey, as well as 200 randomly drawn White respondents. This technique allowed Gay et al. (2015) to ensure that they obtained a final sample with variation across their theoretically motivated respondent characteristics of interest.

The ability to over-sample or stratify on variables of interest can be a useful tool for social science researchers. For example, this has been very helpful in our research on climate change politics because we are particularly interested in individuals who deny climate change, which is relatively rare in the liberally oriented MTurk population. Scholars can now ensure that the samples they draw from MTurk satisfy specific criteria of their choosing. Researchers can use this tool to ensure that they obtain a sample with a specified number of Democrats and Republicans. Or researchers could stratify on other questions.³¹ Doing so allows the researcher to obtain larger sample sizes of otherwise hard to reach parts of the population that likely will respond quite differently to experimental manipulations. This then becomes very important for being able to estimate heterogeneous treatment effects which are interesting in their own right. Furthermore, this marks a step toward addressing external validity criticisms of research conducted using MTurk samples since scholars can use similar sampling strategies to those used by professional polling firms. Finally, researchers can create panel surveys by recontacting respondents in much the same way.

Conclusion

In this paper we took a step toward answering the frequently voiced question of “Who are these MTurk respondents?”. In doing so we presented a number of results. First, we compared the joint distributions of key demographic characteristics of interest to political scientists. In doing so we analyzed the strengths and weaknesses of MTurk samples as they cut across these different categories. For example, we showed that MTurk is relatively strong at attracting young Hispanic females and young Asian males and females. Second, we showed how the age of respondents interacts with voting patterns, partisan preferences, news interest, and education. We demonstrated that, on average, the estimated difference between CCES and MTurk decreased when we subset the data to younger individuals. Fourth, we compared the occupations of respondents from MTurk and CCES. We showed that the percentage of respondents employed in a specific sector were very similar, with a maximum difference of less than 7%. Fifth, we showed that the number of respondents living in different geographic categories on the rural–urban continuum is almost identical in MTurk and CCES. Both MTurk and CCES draw approximately 90% of their respondents from urban areas. Finally, we discussed how experimental political scientists can build “pools” of prior MTurk respondents and recontact these respondents using the open-source

package MTurkR. Researchers can use these pools to oversample and stratify to build samples that are balanced on theoretically motivated variables of interest.

The results presented in this paper provide a number of comparisons that could be useful for further understanding the external validity of research relying on MTurk samples. This is important for social science researchers when we have strong theoretical reasons to suspect that our experimental manipulations will interact with characteristics of the sample. We provided several examples of how experimental researches can leverage the strength and weaknesses of MTurk samples to their advantage. For example, we show that MTurk is an excellent resource for attracting young individuals interested in the news, Hispanics and Asian respondents, as well as individuals from a number of industries and geographic locations in ways that parallel other professionally supplied samples. The results demonstrated in this paper show that there are strong reasons for researchers to consider using MTurk to make inferences about a number of broader populations of interest.

There are a number of takeaways from this paper that are useful for both academics and non-academics alike. First, MTurk is a relatively inexpensive and easy to use survey platform that allows researchers in both academia and the private sector to gain access to a large number of survey respondents. This means that MTurk can serve as a “democratizing” force by allowing researchers to field surveys that might otherwise be difficult given the high costs often associated with professional survey firms. Second, respondents on MTurk are not all that different from respondents on other survey platforms. These differences are even smaller as we focus in on certain attributes of the worker pools such as among younger respondents. This means that researchers, policymakers, and journalists reading work that utilizes the MTurk platform should not immediately dismiss the research as being fielded on a non-representative sample, but instead think carefully about how the MTurk worker pool differs from other platforms and how we might theoretically expect this to affect results. Third, the ability to build survey pools and recontact respondents with particular attributes is a useful tool for anyone attempting to survey individuals with a specific set of characteristics. This is useful for researchers in both academia and the private sector as they attempt to gain access to a particular set of respondents.³² As experimental and survey-based research continues to surge in popularity it is important that political scientists, journalists, and policymakers alike continue to ask and answer the important question of “who are these people?”

Acknowledgements

We thank Peter Bucchianeri, Michael Gill, Christopher Lucas, Anton Strezhnev, two anonymous reviewers, and the editors of *Research and Politics* for helpful comments on previous drafts. We also thank Steve Worthington at the Institute for Quantitative Social Science at Harvard University for research support.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Notes

1. The interest in using voluntary labor pools for survey research is also made evident by Berinsky et al. (2013), who looked at the use of mechanisms for detecting “rushing through” surveys. Berinsky et al. (2013) shows that as many as half of all respondents breeze through surveys without paying attention and that Screener passage correlates with politically relevant characteristics, such as education and race. They present a number of ways of overcoming this inattentiveness. This research is parallel to ours as it is an attempt to understand the strengths and weaknesses of new sources of survey data.
2. Berinsky et al. (2012) examine the Current Population Survey (CPS) and American National Election Studies (ANES). Given the expanding set of viable platforms for survey research in political science, contrasting to the CCES is a helpful contribution given the many publications coming out of the CCES project. See <http://projects.iq.harvard.edu/cces/publications> for examples.
3. The refusal rates 1, 2, and 3, are 0.031, 0.033, and 0.070 for CCES, respectively. For more information on the CCES response and refusal rates see Ansolabehere (2012). Given that we used MTurk we cannot report a response rate for this sample.
4. The CCES occurs in two waves during election years. We focus on the first wave since this part of the survey asks about the political attitudes and demographic factors which are crucial in assessing the quality/characteristics of the CCES and MTurk samples.
5. MTurk survey participants were recruited to take a short survey (10–15 questions) on political attitudes using a pay rate of \$3. We required that respondents had a previous approval rating of at least 95% and were registered in the United States.
6. We assume in this paper that the MTurk respondents we compare to CCES are representative of the broader population of MTurk respondents. To check this we compared the demographic characteristics of the sample presented in this paper with a second set of simultaneous MTurk respondents recruited using a similar advertisement and found them to be extremely similar.
7. This paper is part of a growing strain of research seeking to understand the characteristics of the MTurk worker pool (Chandler et al., 2014; Krupnikov and Levine, 2014; Paolacci and Chandler, 2014; Ross et al., 2010). Understanding the demographic characteristics and political preferences of MTurk workers provides a useful stepping stone for further research exploring the costs and benefits of conducting survey research on MTurk.
8. Our motivation for interacting age with other variables is that MTurk is commonly criticized for drawing from a different range of ages than other survey pools. Understanding these interactions is crucial for the continued development of experimental political science.
9. See <http://thomasleeper.com/MTurkR/http://thomasleeper.com/MTurkR/>

10. For examples of the potential moderating effects of race, age, gender, political party, and occupation see Fischer and Shaw (1999), Heo et al. (2005), Dreher and Cox (2000), Philpot (2004), Zaller (1992), and Bamundo and Kopelman (1980).
11. We obtain similar results if we analyze the collection of team surveys which has a sample size of over 50,000 respondents. We analyze the team survey because its sample size is more reflective of what an individual researcher might be able to obtain either on MTurk or through other survey sources for a single survey.
12. This feature is unavoidable. The CCES asks a "Common Content" section of additional questions. While our survey asked some of these questions, we did not ask others such as questions about the individual's own elected representatives.
13. Additionally, YouGov/Polimetrix initially draws a larger sample than is required for the target sample size. To minimize the size of the survey weights, some individuals are excluded from the final sample.
14. It is also notable that in a sample of 1300 people there are 32 with a survey weight over 5. Compared to others in the sample, these individuals are extremely heavily weighted. Some suggest excluding these observations from analysis.
15. Contrasts using the CCES survey weights are presented in Appendix A.
16. These results hold across a second simultaneous sample of CCES and MTurk respondents.
17. This percentage is calculated by dividing the total number of individuals within a particular racial category by the number of respondents in the sample.
18. This trend is depicted in Figure 2 by the large height of the light gray boxes in the column on the left.
19. Researchers should use weights that are calibrated to best represent the populations for which they are attempting to make inferences. We recommend that when using weights researchers present both unweighted and weighted estimates. See Ansolabehere (2012) for how CCES approaches this problem.
20. Regression results weighting CCES respondents are presented in Figure 4 in Appendix A.
21. These percentages are calculated within each survey by dividing the total number of individuals employed in a particular sector by the total number of individuals in all sectors. Table 1 shows these sector-specific percentages.
22. In a second simultaneous study comparing respondents from MTurk and CCES the percentage of respondents employed in "Other Services" was extremely similar at roughly 19%.
23. The rural-urban codes can be found at <http://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx#Uy7xQvldV8E>. In order to calculate the proportion of respondents in a particular urban rural code we first merged the zip code data with FIPS codes. We then aggregated this new dataset sorted at the FIPS level according to the urban-rural codes.
24. The percentages are calculated for a particular sample by dividing the number of individuals within an urban-rural category by the number of respondents in the sample.
25. This R package, which was developed by Thomas Leeper, provides easy access to the MTurk Requester API and a range of functionalities Leeper (2015), Leeper (2013). Other tools are available for a similar purposes, but this particular package is highly developed and within a software framework many political scientists know.
26. In contrast, the 2012 Common Content CCES survey had 54,535 respondents. Unweighted, 47% were male, on a one to seven point ideology scale the average was 4.27, 37% self-identified as Democrat, 27% as Republican, and 30% as independent (the remaining identified with "other" parties), and the average age was 52.
27. This might not be as useful for researchers that conduct surveys infrequently due to the high turnover of respondents on MTurk and the inability to build large pools if the researcher does not attract a large number of respondents.
28. As noted by Paolacci and Chandler (2014) this is in part driven by the fact that workers choose the tasks they work on. This means that workers might have incentives to participate in studies using procedures that they have already seen.
29. As highlighted by Chandler et al. (2014), these assumptions include that observations are randomly assigned and independent from each other.
30. It should be noted that this technique will only be successful if what differentiates MTurk respondents from those in other survey platforms is an observable characteristic. For example, Krupnikov and Levine (2014) observed different results when replicating the same experiment on MTurk and other survey platforms even after accounting for factors such as partisanship and savviness. While we view our paper as a step toward better understand the moderating variables that differentiate respondents on MTurk from other survey pools, further research is needed to both theorize and empirically test other potential unobserved moderating variables.
31. The decision to respond to the invitation will vary by compensation level. Needless to say we and others have been satisfied with the experience.
32. For example, private firms could build a pool of respondents all of whom have previously used their product, and then ask them a targeted set of questions of interest. The ability to build a pool and recontact respondents would allow the firm to do this in a cost-effective way that would be difficult on other types of survey platforms.
33. We were unable to generate maps for Alaska and Hawaii due to technical issues with the way the maps are generated.
34. Available at <http://www.census.gov/popest/data/cities/totals/2012/SUB-EST2012.html> <http://www.census.gov/popest/data/cities/totals/2012/SUB-EST2012.html>

Supplementary material

The replication files are available at: <http://thedata.harvard.edu/dvn/dv/researchandpolitics>

References

- Ansolabehere S (2012) Guide to the 2012 cooperative congressional election survey. Harvard University. Available at: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/21447>
- Arceneaux K (2012) Cognitive biases and the strength of political arguments. *American Journal of Political Science* 56: 271–285.

- Bamundo PJ and Kopelman RE (1980) The moderating effects of occupation, age, and urbanization on the relationship between job satisfaction and life satisfaction. *Journal of Vocational Behavior* 17: 106–123.
- Berinsky AJ, Huber GA and Lenz GS (2012) Evaluating online labor markets for experimental research: Amazon.com's mechanical Turk. *Political Analysis* 20: 351–368.
- Berinsky AJ and Kinder DR (2006) Making sense of issues through media frames: Understanding the Kosovo crisis. *Journal of Politics* 68: 640–656.
- Berinsky AJ, Margolis MF and Sances MW (2013) Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American Journal of Political Science* 58: 739–753.
- Chandler J, Mueller P and Paolacci G (2014) Nonnaveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods* 46: 112–130.
- Dreher GF and Cox TH (2000) Labor market mobility and cash compensation: The moderating effects of race and gender. *Academy of Management Journal* 43: 890–900.
- Druckman JN and Kam CD (2011) Students as experimental participants. In *Cambridge Handbook of Experimental Political Science*. Cambridge: Cambridge University Press, pp. 41–57.
- Fischer AR and Shaw CM (1999) African Americans' mental health and perceptions of racist discrimination: The moderating effects of racial socialization experiences and self-esteem. *Journal of Counseling Psychology* 46: 395.
- Gay C, Hochschild J and White A (2015) Americans belief in linked fate: A wide reach but limited impact. *Journal of Race, Ethnicity, and Politics*, in press.
- Grimmer J, Messing S and Westwood SJ (2012) How words and money cultivate a personal vote: The effect of legislator credit claiming on constituent credit allocation. *American Political Science Review* 1: 1–17.
- Heo M, Pietrobelli A, Fontaine KR, Sirey JA and Faith M (2005) Depressive mood and obesity in US adults: Comparison and moderation by sex, age, and race. *International Journal of Obesity* 30: 513–519.
- Huber GA, Hill SJ and Lenz GS (2012) Sources of bias in retrospective decision making: Experimental evidence on voters' limitations in controlling incumbents. *American Political Science Review* 106: 720–741.
- Kam CD, Wilking JR and Zechmeister EJ (2007) Beyond the "narrow data base": Another convenience sample for experimental research. *Political Behavior* 29: 415–440.
- Krupnikov Y and Levine AS (2014) Cross-sample comparisons and external validity. *Journal of Experimental Political Science* 1: 59–80.
- Leeper T (2013) Crowdsourcing with R and the MTurk API. *The Political Methodologist* 20(2): 2–7.
- Leeper TJ (2015) MTurkR: Access to Amazon Mechanical Turk Requester API via R. R package version 0.6.5.1.
- Paolacci G and Chandler J (2014) Inside the turk understanding mechanical turk as a participant pool. *Current Directions in Psychological Science* 23: 184–188.
- Peterson RA (2001) On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research* 28: 450–461.
- Philpot TS (2004) A party of a different color? Race, campaign communication, and party politics. *Political Behavior* 26: 249–270.
- Prior M (2005) News vs. entertainment: How increasing media choice widens gaps in political knowledge and turnout. *American Journal of Political Science* 49: 577–592.
- Ross J, Irani L, Silberman M, Zaldivar A and Tomlinson B (2010) Who are the crowdworkers? Shifting demographics in Mechanical Turk. In *CHI'10 extended abstracts on human factors in computing systems*. New York: ACM Press, pp. 2863–2872.
- Zaller J (1992) *The nature and origins of mass opinion*. Cambridge: Cambridge University Press.

Appendix A: Contrasts using CCES survey weights

In Figure 4 we use regression with the CCES survey weights to contrast several key variables of interest to political scientists. For the CCES respondents we use the CCES weights while the weights are set at 1 for all MTurk respondents. The results presented in Figures 3 and 4 are extremely similar for Party ID, News Interest, Ideology, and Education. In contrast, there is a marked difference for whether individuals were registered to vote and whether they intended to vote in 2012. This means that the CCES is more heavily weighting individuals with voting patterns that most closely resemble the population about which they are making inferences.

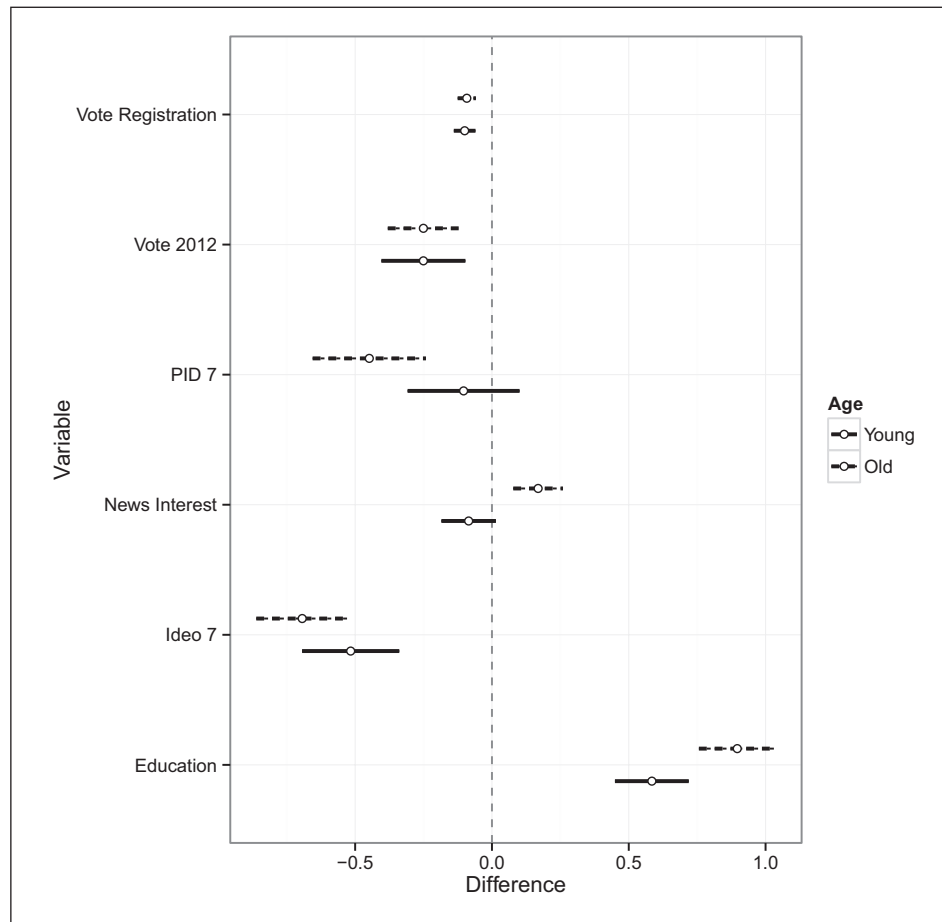


Figure 4. Differences in means using the weighted values for CCES with 95% confidence intervals for proportion of respondents registered to vote, proportion of respondents that intend to vote in 2012, party identification, ideology, level of news interest, and education level in the CCES and MTurk modules. Positive values indicate that MTurk is greater than CCES. Dashed lines correspond with the confidence intervals for older respondents and solid lines for younger.

Appendix B: Geographic distribution

Figures 5 and 6 show the proportion of survey respondents by county across the United States.³³ Figure 5 presents the proportions for the MTurk sample that was compared against CCES throughout this paper. Figure 6 shows the proportions from the large pool of prior MTurk respondents. The proportions were calculated by dividing the number of respondents in a county by the total number of respondents in the sample. Red points denote the 20 most populous cities within the United States³⁴ as well as state capitals. These points clearly show the urban clustering we presented in Table 2. For example, we can see the high proportion of respondents around large cities such as Los Angeles, New York, Philadelphia, and Houston. Indeed, the urban clustering and the proportions of respondents in the cities represented in these figures stays relatively constant across both the CCES and MTurk surveys. Our replication code will enable researchers to map the geographic distributions of respondents for their own research.

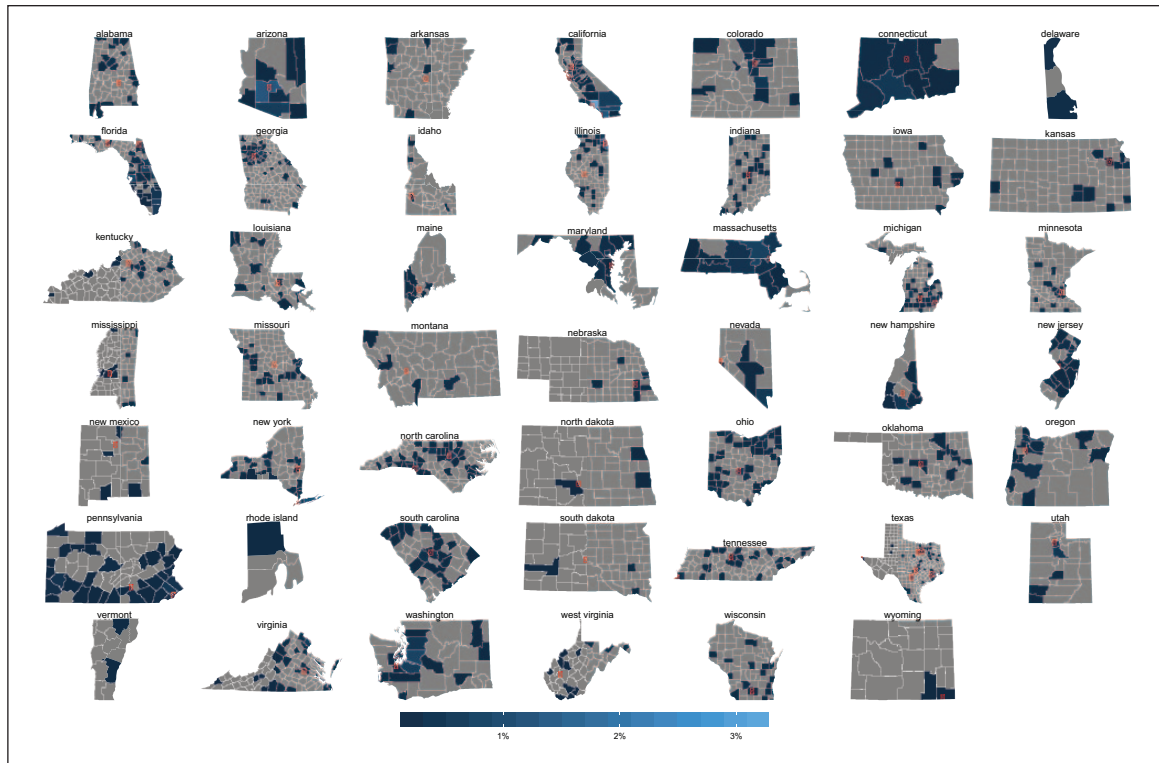


Figure 5. The distribution of respondents from the MTurk survey by county. The percentages are calculated by dividing the number of respondents in a particular county by the number of respondents in the sample.

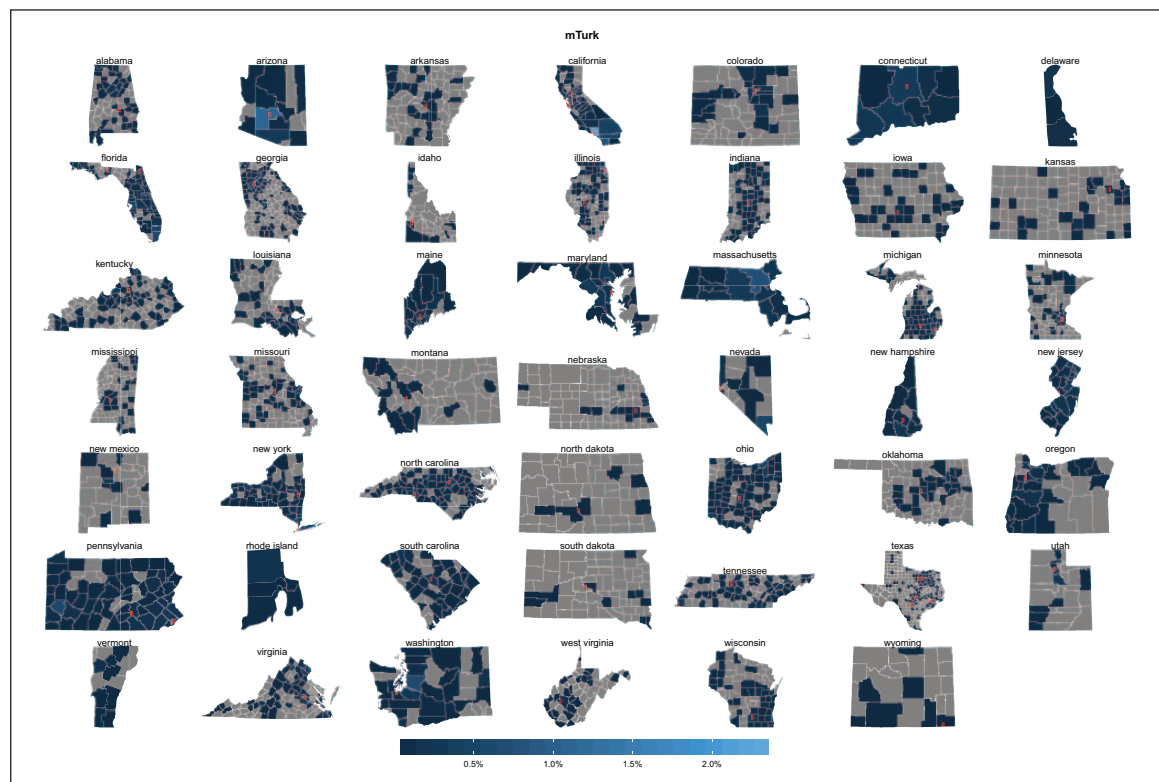


Figure 6. The distribution of respondents from the larger pool of MTurk respondents by county. The percentages are calculated by dividing the number of respondents in a particular county by the number of respondents in the pool.