Article

The Demographic and Political Composition of Mechanical Turk Samples

SAGE Open
January-March 2016: 1–17

DOI: 10.1177/2158244016636433
sgo.sagepub.com
SAGE

Kevin E. Levay[1], Jeremy Freese[2], and James N. Druckman[1]

Abstract

One of the most notable recent developments in survey research is the increased usage of online convenience samples drawn from Amazon's Mechanical Turk (MTurk). While scholars have noted various social and political differences (e.g., age, partisanship) between MTurk and population-based samples, the breadth and depth of these variations remain unclear. We investigate the extent to which MTurk samples differ from population samples, and the underlying nature of these differences. We do so by replicating items from the population-based American National Election Studies (ANES) 2012 Time Series Study in a survey administered to a sample of MTurk respondents. With few exceptions, we not only find that MTurk respondents differ significantly from respondents completing the 2012 ANES via the Web but also that most differences are reduced considerably when controlling for easily measurable sample features. Thus, MTurk respondents do not appear to differ fundamentally from population-based respondents in unmeasurable ways. This suggests that MTurk data can be used to advance research programs, particularly if researchers measure and account for a range of political and demographic variables as needed.

Keywords

Amazon's Mechanical Turk, survey sampling, research methods, public opinion

Introduction

Surveys remain fundamental to the study of public opinion and political behavior. At the same time, a number of significant developments over the past few decades have transformed survey methodology. One notable development is the rise of Web-based surveys (see Baker et al., 2010). With increased public accessibility of the Internet, researchers collect survey data from thousands of participants at a lower cost and quicker pace than ever before (Couper, 2000, 2011).

Amazon.com's Mechanical Turk (MTurk) is one example of a useful Internet-based resource. MTurk allows researchers to easily recruit participants and administer surveys at little cost: Consequently, it has emerged as a prominent data source among researchers of U.S. public opinion (e.g., Arceneaux, 2012; Gerber, Huber, Doherty, & Dowling, 2011; Huber, Hill, & Lenz, 2012; Huber & Paris, 2013). However, MTurk samples include only individuals who opt into MTurk and complete a given survey, leading to samples that are typically unrepresentative of the U.S. population. A fair amount of work shows that MTurk samples consistently differ from population-based samples of the American public on many demographic and political variables (Berinsky, Huber, & Lenz, 2012; Buhrmester, Kwang, & Gosling, 2011; Huff & Tingley, 2015; Ipeirotis, 2010; Krupnikov & Levine, 2014; Mullinix, Leeper, Freese, & Druckman, 2015).[1] Nevertheless, a growing literature focusing on the generalizability of experimental research using MTurk data finds that with some exceptions, researchers can make credible, generalizable experimental inferences with some confidence (e.g., Berinsky et al., 2012; Krupnikov & Levine, 2014; Mullinix et al., 2015).

We update and expand on this body of work by comparing an MTurk sample and a population-based sample designed to be representative of the U.S. population on a larger range of demographic and political variables than most previous work, namely, a host of attitudinal and behavioral measures. Of greater importance, we address whether easily measurable features can explain raw descriptive sample differences; that is, can differences between the samples in more specific variables be explained by more basic and well-known demographic and political variables? If so, MTurk samples could provide a fruitful resource for hypothesis generation and testing for observational researchers of

[1]Northwestern University, Evanston, IL, USA
[2]Stanford University, CA, USA

Corresponding Author:
James N. Druckman, Department of Political Science, Northwestern University, Scott Hall 601 University Place, Evanston, IL 60208, USA.
Email: druckman@northwestern.edu

public opinion, so long as they account for at least these basic variables where differences usually exist. If not, MTurk respondents may fundamentally differ from those in population-based samples in ways that will misdirect inferences in observational studies and that cannot be corrected simply by weighting the sample by measurable features. In either case, researchers of public opinion, including experimentalists, will gain more clarity about when and how they might adjust their sampling strategy and procedures to meet their particular needs.

To address these issues, we replicate items from the American National Election Studies (ANES) 2012 Time Series Study in a survey administered to a sample of MTurk respondents. We compare MTurk respondents with a population-based sample who completed the 2012 ANES survey via the Web. We document numerous and many less-noted raw differences, though most are consistent with one another and extant research. Yet, we also find that, with some exceptions, once we control for a short set of relevant variables, approximately nine, the effects of being an MTurk respondent diminish considerably. In short, most differences between respondents in our MTurk and population-based samples do not seem to reflect unobservable distinctions, but rather variation on factors that researchers might expect and can easily capture.

Our results suggest that, if used with care, observational researchers can use MTurk for developing a research program in that the data will generate useful inferences on which to build (i.e., they should not generate an unusual number of false negatives or false positives). However, we do not suggest that MTurk is a wholesale substitute for population-based samples, nor that controlling for the covariates we identify here is a universally effective formula. In addition to some evidence showing that MTurk samples might under- and overrepresent some groups differently at distinct points in time, we also find some inexplicable and less well-noted differences. Our point, then, is that if researchers account for most basic sample features, they can effectively use these measures to adjust observational data (e.g., using sampling weights) so that inferences can often, though not always, be generalized to larger populations (see Callegaro et al., 2014, for discussion).[2] In other words, while perhaps not ideal for observational research, with extensive and careful measurement, MTurk can be used with some confidence for theory development.

In addition, our results further clarify when experimentalists expecting heterogeneous treatment effects by some characteristic should be more or less careful when using MTurk. Although by no means definitive, our findings suggest that even where MTurk underrepresents individuals with a given characteristic, these researchers can proceed with some confidence that increasing the number of people with that characteristic in their sample might be sufficient for properly estimating treatment effects across or among those individuals. In any case, researchers of all kinds should always analyze and situate their findings in the context of their sample, both in process and composition.

We begin with a brief discussion of MTurk and the ways in which samples drawn with MTurk most commonly differ from population-based samples. Next, we describe our data—how we recruited and collected data from our MTurk sample, as well as the features of our comparison population-based sample who completed the 2012 ANES. We then present the results of our analysis, showing a number of basic raw observational differences, but also evidence that many, though not all, of these differences are reduced considerably with the inclusion of easily measurable covariates. We conclude with suggestions for researchers interested in using MTurk to conduct surveys who might be concerned with MTurk sample composition.

MTurk and MTurk Samples

MTurk is an online crowdsourcing labor market launched by Amazon.com in 2005. Individuals and organizations (requesters) use MTurk to hire humans (workers) to complete various computer-based tasks (Human Intelligence Tasks or HITs). Requesters post HITs on the MTurk site and can make HITs available to all MTurk workers or only workers who meet a given set of requirements (e.g., an IP address located in a specific geographic area). Upon completion of a HIT, requesters then either reject or accept individuals' work and reward them accordingly. Given that MTurk now boasts a large number of workers willing to complete any number of computer tasks for small sums of money, it has become a popular resource among survey researchers, as surveys can be posted as HITs for completion. For example, a Google Scholar search of "Mechanical Turk" yields 1,080 hits from 2010; 2,750 from 2012; and 5,520 from 2014.

Nevertheless, MTurk respondents include those who opt in to MTurk and a given survey, meaning that MTurk samples are unrepresentative of many specified populations. This also means calculating respondents' chances of being sampled is virtually impossible (see Baker et al., 2010; Couper, 2011). Thus, as reviewed in more detail below, while MTurk samples are usually more representative than other commonly used convenience-based samples, such as college students (Berinsky et al., 2012; Buhrmester et al., 2011; Casler, Bickel, & Hackett, 2013; Paolacci, Chandler, & Ipeirotis, 2010), they consistently and clearly differ from population-based samples aiming to be representative of the U.S. population. Indeed, the use of MTurk has sparked considerable debate, though there is emerging consensus that MTurk is largely suitable for experimental work (Berinsky et al., 2012; Buhrmester et al., 2011; Casler et al., 2013; Krupnikov & Levine, 2014; Mullinix et al., 2015; Paolacci et al., 2010). There is no such consensus when it comes to observational work, however.

How do MTurk samples usually differ from the U.S. population? Extant work shows a number of stable demographic differences between MTurk samples and U.S. population-based samples. MTurk samples tend to have lower average incomes, higher average education levels, lower average ages, and much smaller percentages of most non-White groups, especially Blacks (Berinsky et al., 2012; Buhrmester et al., 2011; Huff & Tingley, 2015; Ipeirotis, 2010; Krupnikov & Levine, 2014; Paolacci & Chandler, 2014).[3] Similarly, differences in political identity appear to be mostly stable, with MTurk samples frequently containing many more Democrats and liberals than population-based samples (Berinsky et al., 2012; Krupnikov & Levine, 2014; Mullinix et al., 2015).

Existing comparisons are useful, but they do not consistently capture all potentially relevant differences between MTurk and U.S. population-based samples, nor do they consider relationships between differences. This is especially true when it comes to specific political attitudes (e.g., government spending preferences, opinions about climate change, gun control, etc.) and/or behaviors (e.g., contacting a government official; although, see Berinsky et al., 2012, who include some political variables, but do not analyze relationships between differences on these and other variables). These gaps are notable, given that researchers using MTurk most often want to make observational and/or experimental inferences about attitudes and behaviors, as well as about and across subgroups with particular characteristics.

How consistent are the various differences of MTurk samples? Do MTurk samples differ in other ways than those that are most often measured? More importantly, do they differ in more specific attitudes and/or behaviors, and if so, can these be explained by more basic demographic and political differences, or are they reflective of fundamental, potentially unmeasurable features of MTurk respondents? If the former, MTurk could be a relatively useful data source even among observational researchers, as they can account for factors that differentiate MTurk respondents from others. If the latter, the ability of researchers using MTurk to make inferences would be more limited, particularly those wishing to make observational inferences of various kinds, but also researchers wishing to draw inferences about specific subgroups or who expect treatment effects to differ among specific subgroups who are less well represented among MTurk respondents.

Data and Method

To address our questions, we turn to the ANES 2012 Time Series Study. The ANES Time Series Study has been conducted in every national election year since 1948 and is one of the most extensive social and political surveys of the American electorate. In 2012, for the first time in its history, the Time Series Study was conducted both face-to-face and over the Internet with two independent samples. The sample who completed the Internet survey were members of GfK's (formerly Knowledge Networks) KnowledgePanel, a sample of U.S. households drawn using random digit dialing and address-based methods. Sampled households were drawn randomly from a listing of addresses from the U.S. Postal Service's Delivery Sequence File and U.S. residential landline telephone numbers. Households without a computer and/or Internet access were offered a laptop and free monthly Internet access from GfK in return for participation. A sample of KnowledgePanel members, who were U.S. citizens and at least 18 years old by the time of the 2012 presidential election (November 6, 2012), was sent invitations to participate in the 2012 ANES. The sample includes 3,860 individuals who completed the 2012 ANES during the 2012 U.S. presidential election in four waves, two conducted prior to the election (October 11, 2012 through November 6, 2012) and two after (November 29, 2012 through January 24, 2013).[4]

The 2012 ANES is well suited for our purposes. To begin, it provides data from a population-based sample that, with the use of weights, is representative of the U.S. population. As important, it offers data on a number of variables of direct relevance to political and social science researchers and that remain largely unexplored with respect to MTurk. Finally, given completion of the survey via Internet by an independently drawn, population-based sample, we can analyze and compare the data from the 2012 Web ANES sample with data from an MTurk sample while controlling for survey mode (see Struminskaya et al., 2014 and Liu & Wang, 2015 on survey mode effects).

We designed and administered a survey on MTurk replicating numerous questions from the 2012 ANES. While limited by timing (i.e., not during or directly after the 2012 campaign; see below), our survey included a range of questions, many of which do not appear in extant work, thereby allowing us to offer one of the broadest sample comparisons to date and to test whether a number of sample differences can be captured by a host of basic demographic and political variables. In addition to a number of commonly measured demographics, such as age, income, and racial and ethnic identification, we included less consistently measured, as well as more specific demographics, such as employment status, home ownership, and both religious identity and group affiliation. We also included a number of opinion measures on salient topics, such as abortion, gun control, government spending in numerous areas, lesbian, gay, bisexual, transgender, queer (LGBTQ) rights, the environment, and the general role of government, as well as behavioral measures, such as contact with an elected official, volunteering, and media consumption habits. All questions were worded and

randomized identically to the 2012 ANES, as well as ordered according to the order in which they appeared in the survey questionnaire.[5]

We made our survey available to MTurk workers on June 10, 2013. Following others (e.g., Berinsky et al., 2012; Huff & Tingley, 2015; Krupnikov & Levine, 2014), we set three requirements for participation: a HIT Approval Rate greater than or equal to 90, at least 50 HITs approved, and an IP address based in the United States. The survey was advertised to workers as a national public opinion survey that would take 20 to 30 minutes to complete, terms similar to that of other work.[6] We closed the HIT after 2 days, on June 12, 2013, when we reached 2,000 unique survey completions.

Given the advertised completion time, and to gather data as quickly as possible, participants were compensated US$2.50 upon completion of the survey. Although relatively high, this amount is smaller than some (e.g., Huff & Tingley, 2015) and the rate per minute based on the advertised survey completion time is roughly comparable with most other work (e.g., Berinsky et al., 2012). Moreover, though not studied systematically, there is little evidence to suggest that higher compensation rates do more than decrease the amount of time it takes to recruit participants (Buhrmester et al., 2011; Mason & Watts, 2009; also see Paolacci & Chandler, 2014). In short, given comparable recruitment methods, we believe that our findings can be reasonably compared with those of extant work.

Below, we first assess the breadth of social and political differences between the samples and then consider the extent to which they can be explained by measured characteristics. We do so using regression analyses, first without covariates so as to calculate the simple differences between MTurk and 2012 ANES Web respondents on various items—which we call the "raw" differences—and then with basic covariates to estimate whether and how much this resolves those differences. For all analyses, we exclude respondents for which no response was given or "Don't Know" was chosen. We also weight the ANES data to be representative of the U.S. population.

Results

Raw Socio-Demographic Differences

Table 1 presents descriptive data for socio-demographic variables among each sample, as well as coefficients for an MTurk dummy variable used in regression models of each variable with no controls. The latter provides the size of a given difference between the samples and whether it is statistically significant by conventional standards. Table 1 mostly confirms differences found in prior work; the MTurk sample is younger, more likely to be unmarried, lower income, and much less racially and ethnically diverse than the ANES sample. The average age of the MTurk sample is almost 20 years younger than the ANES sample (31.6 years vs. 47.4 years). A much larger share of the MTurk sample has also never been married (60.3% vs. 25.7%). The average reported household income of the MTurk sample is about US$10,000 less than the ANES sample (~US$54K vs. ~US$65K) and a significantly smaller percentage of MTurk respondents owns a home (37.8% vs. 72.1%). Finally, compared with the ANES sample, the MTurk sample contains about the same percentage of Whites (71.8% vs. 70.5%), but significantly smaller percentages of Blacks (7.1% vs. 12.2%) and Hispanics or Latino/as (5.6% vs. 11.2%). In addition, many more MTurk respondents report no race or ethnicity at all (7.1% vs. 1.9%) or something other than only those categories (8.6% vs. 3.7%).[7] We also find a much larger percentage of MTurk respondents are students (22.3% vs. 6.4%). Thus, as others find, despite being less well-off economically, the MTurk sample is much more educated than the ANES sample, with 45% of the MTurk sample possessing at least a bachelor's degree versus only 29.5% of the ANES sample.[8]

Interestingly, while some earlier work highlights that MTurk samples underrepresent men (e.g., Ipeirotis, 2010; Paolacci et al., 2010), our data exhibit the opposite pattern (53.9% male). This finding, however, is consistent with more recent work (e.g., Berinsky et al., 2012; Casler et al., 2013; Krupnikov & Levine, 2014). Similarly, though less consistently measured across studies, our data appear to contrast when it comes to religious identity. We find no differences between the samples in any organized religious group identifications (also see Berinsky et al., 2012). However, we do confirm previous findings on religious beliefs and religiosity; a strikingly substantial percentage of the MTurk sample—over 40%—identifies as agnostic or atheist, whereas only 10% of ANES respondents does and many fewer MTurk respondents attend church or religious services (29.4% vs. 54.9%; see Paolacci & Chandler, 2014).

Given comparable recruitment methods, though speculative, the above suggests that the composition of the MTurk worker population might shift over time, at least on some variables. Also, MTurk samples might differ at different points in time on particular variables more than on others, even where measured characteristics might seem related (e.g., wider and more stable differences in religious beliefs/religiosity vs. organized religious group identification). While these are generally beyond researchers' control, as we discuss in the conclusion, it underscores the utility of regularly measuring multiple demographic features in multiple ways when using MTurk.

Raw Political Differences

Tables 2 and 3 include descriptive data for political variables, as well as the MTurk coefficients for regression models of

Table 1. Socio-Demographic Characteristics of MTurk and 2012 ANES Web Samples.

	MTurk sample (n=605-2,144)	ANES Web Sample (n=1,944-3,860)	MTurk coefficient
Age (mean number of years)	31.6	47.4	−15.82***
Male	53.9%	47.9%	0.06***
Race and ethnicity[a]			
White	71.8%	70.5%	0.01
Black, non-Hispanic/Latino	7.1%	12.2%	−0.05***
Non-White Hispanic/Latino	5.6%	11.2%	−0.06***
Other	8.6%	4.8%	0.04***
No race given	7.1%	1.4%	0.06***
Income[b] (M)	US$54,257	US$63,565	−9307.01***
Education (mean on 5 point scale, 5 = *graduate degree*)	3.4	2.9	0.52***
Spouse's education (mean on 5 points scale, 5 = *graduate degree*)	3.4	3.0	0.69***
Marital status			
Married	31.2%	53.3%	−0.22***
Divorced	6.9%	13.1%	−0.06***
Separated	0.9%	2.3%	−0.01***
Widowed	0.6%	5.5%	−0.05***
Never married	60.3%	25.7%	0.35***
Living with partner (unmarried)	21.9%	18.4%	0.04*
Own home	37.8%	72.1%	−0.34***
Employment status[c]			
Working now	57.4%	53.3%	0.04**
Unemployed	13.9%	7.8%	0.06***
Temporarily laid off	1.9%	1.1%	0.01*
Retired	1.5%	21.3%	−0.19***
Homemaker	9.2%	9.1%	0.001
Permanently disabled	1.4%	6.3%	−0.05***
Student	22.3%	6.4%	0.16***
Number of hours working per week (if working now)	36.6	38.3	−1.68***
Out of work/laid off in last 6 months (if working now)	13.4%	10.7%	0.03
Number of children in household ages 0-10 (mean on 3 point scale, 2 = *two or more*)	0.3	0.2	0.09***
Number of children in household ages 11-17 (mean on 3 point scale, 2 = *two or more*)	0.2	0.2	0.03
Has served on active duty in the Armed Forces	3.9%	12.5%	−0.09***
Level of satisfaction with life (mean on 5 point scale, 5 = *extremely satisfied*)	2.9	3.2	−0.28***
Self-assessment of health condition (mean on 5 point scale, 5 = *excellent health*)	3.4	3.3	0.12***
Religious identity[c]			
Agnostic	20.7%	5.5%	0.15***
Atheist	23.7%	4.1%	0.19***
Nontraditional	9.3%	11.9%	−0.03**
Progressive	11.7%	16.0%	−0.04***
Secular	6.2%	4.7%	0.02*
Spiritual—not religious	22.1%	28.2%	−0.06***
None	17.4%	35.8%	−0.18***
Religious group			
Protestant	41.6%	38.6%	0.03
Catholic	25.7%	29.0%	−0.03
Jewish	4.5%	3.0%	0.02
Other	28.2%	29.3%	−0.01
Attends religious/church services	29.4%	54.9%	−0.26***

Note. MTurk = Mechanical Turk; ANES = American National Election Studies 2012 Time Series Study. ANES data are weighted for all analyses. All data exclude respondents providing no answer or "Don't Know."

[a]Respondents able to select more than one racial category. Hispanic/Latino was asked as a separate question on ANES and MTurk survey. Response options recoded to reflect the following: White indicates that participant selected only White; Black, non-Hispanic/Latino indicates that participant selected only Black for his or her racial identification and no to Hispanic/Latino; Hispanic/Latino indicates that participant answered yes to identifying as Hispanic or Latino and did not select White; Other indicates participant selected one of the following additional options: Asian, American Indian/Alaska Native, Native Hawaiian/other Pacific Islander, or other or selected multiple categories (excluding those indicating Hispanic/Latino); no race given indicates participant did not answer any questions about racial identification.

[b]Individual-level income data not accessible from ANES at time of analysis; data reflect income midpoints of categorical income variable that was available at the time of this writing.

[c]Respondents were able to select more than one response option so percentage total may equal more than 100. Percentages calculated by dividing number of respondents to select response option out of total possible number of respondents in each sample (2,144 for MTurk and 3,860 for 2012 ANES).

*Significant at the .1 level. **Significant at the .05 level. ***Significant at the .01 level.

Table 2. Political Characteristics of MTurk and 2012 ANES Web Samples.

	MTurk sample (n=1,264-2,085)	ANES Web sample (n=2,511-3,853)	MTurk coefficient
Political ideology (mean on scale of 1-7; 7 = *extremely conservative*)	3.3	4.2	−0.91***
Partisanship			
Democrat	46.1%	36.1%	
Republican	14.6%	29.4%	
Independent	34.3%	31.1%	
Other	4.9%	3.5%	
Identify strongly with Democratic or Republican Party	42.6%	53.9%	−.11***
Partisanship (mean on scale of 1-7; 7 = *strong Republican*)	3.1	3.8	−0.70***
Favor government environmental regulation of business (mean on scale of 1-7; 7 = *more regulation to protect environment and create jobs*)	5.0	4.6	0.41***
Believe government should guarantee jobs and income (mean on scale of 1-7; 7 = *should see to a job and good standard of living*)	4.1	3.7	0.39***
Favor private insurance plans over government insurance plan (mean on scale of 1-7; 7 = *should be private plans*)	3.2	4.3	−1.01***
Government should provide more services and increase spending (mean on scale of 1-7; 7 = *increase spending/provide more services*)	4.4	3.7	0.78***
Federal spending on child care (mean on scale of 0-2; 2 = *increase*)	1.3	1.0	0.24***
Federal spending on environmental protection (mean on scale of 0-2; 2 = *increase*)	1.4	1.2	0.28***
Federal spending on aid to the poor (mean on scale of 0-2; 2 = *increase*)	1.3	1.1	0.172***
Federal spending on public schools (mean on scale of 0-2; 2 = *increase*)	1.7	1.5	0.19***
Federal spending on science and technology (mean on scale of 0-2; 2 = *increase*)	1.6	1.3	0.32***
Support abortion (mean on scale of 1-4; 4 = *permit abortion under all circumstances*)	3.3	2.9	0.35**
Agree that gay and lesbian couples should be allowed to adopt children	82.2%	61.5%	0.21***
Support lesbian and gay marriage (mean on scale of 1-3; 3 = *legal marriage*)	2.6	2.1	0.50***
Favor offshore drilling (mean on scale of 0-2; 2 = *favor*)	0.9	1.4	−0.39***
Agree that global warming is happening	85.7%	78.1%	0.08***
Agree that rising temperatures are good (mean on scale of 0-2; 2 = *good*)	0.2	0.3	−0.14***
Agree that climate change caused mostly by humans (mean on scale of 0-2; 2 = *human activity*)	1.4	1.1	0.28***
Favor checking status of those suspected to be undocumented (mean on scale of 0-2; 2 = *favor*)	0.9	1.3	−0.35***
Favor granting citizenship to unauthorized immigrants (mean on scale of 1-4; 4 = *should remain in country and qualify for citizenship*)	2.7	2.5	0.25***
Increase level of foreigners permitted to enter United States to live (mean on scale of 1-5; 5 = *increase a lot*)	2.8	2.4	0.37***
Favor strict gun control laws (mean on scale of 0-2; 2 = *make it more difficult to buy gun*)	1.5	1.4	0.08***
Support the death penalty	54.4%	74.7%	−0.20***
Feels strongly about position on death penalty	55.8%	66.6%	−0.11***

Note. MTurk = Mechanical Turk; ANES = American National Election Studies 2012 Time Series Study. ANES data are weighted for all analyses.
*Significant at the .1 level. **Significant at the .05 level. ***Significant at the .01 level.

each variable without any controls. Consistent with previous work, our MTurk sample is significantly more liberal and Democratic than the ANES sample. The mean ideology of the MTurk sample is almost a full point lower on a scale of 1 to 7 (3.3. vs. 4.2, where 1 is extremely liberal) with 58.4% of the MTurk sample identifying as between slightly liberal and extremely liberal compared with only 27.8% of the ANES sample. Similarly, a significantly larger percentage of MTurk respondents identifies as Democratic (46.1% vs. 38.1%), and a smaller percentage identifies as Republican (14.6% vs. 27.6%).

Our MTurk sample also differs significantly in political preferences, though in ways that are consistent with other work (e.g., Berinsky et al., 2012) and the samples' broader political differences (i.e., preferences reflect the more liberal, Democratic composition of the MTurk sample). Relative to the ANES sample, the MTurk sample more strongly favors environmental regulation of business, guaranteed jobs and income by government, increased spending and services by the government, and a government health insurance plan, as well as increased government spending in many specific areas (e.g., on science and technology). In addition, the MTurk sample expresses stronger support of gay and lesbian marriage, abortion with no restrictions, and stricter gun control laws than the ANES sample. Finally, more respondents in the MTurk sample oppose offshore

Table 3. Political Behaviors of MTurk and 2012 ANES Web Samples.

	MTurk sample (n=909-2,122)	ANES Web sample (n=1,966-3,857)	MTurk coefficient
Has contacted elected official to express view in past 12 months	20.9%	22.9%	−0.02
Has participated in community work in past 12 months	28.5%	30.1%	−0.02
Agree that he or she has little say in what government does (mean on scale of 1-5)	3.3	3.3	0.01
Number of days spent reviewing news (means on scale of 0-7)			
On Internet	4.9	3.7	1.23***
In print newspaper	1.1	2.1	−1.04***
On television	2.6	3.8	−1.15***
Level of attention to news (means on scale of 0-5; 1 = *none at all*, 5 = *a great deal*)			
On Internet	3.1	2.8	0.26***
In print newspaper	2.7	2.9	−0.82***
On television	2.9	3.2	−0.57***
n	909-2,122	1,966-3,857	

Note. MTurk = Mechanical Turk; ANES = American National Election Studies 2012 Time Series Study. ANES data are weighted for all analyses.
*Significant at the .1 level. **Significant at the .05 level. ***Significant at the .01 level.

drilling, deporting immigrants, increasing immigration restrictions, and the death penalty than in the ANES sample, as well as agree that global warming is happening, rising temperatures are bad, and climate change is caused mostly by humans.

The MTurk sample differs less in terms of political behavior. Similar and statistically indistinguishable percentages of MTurk and ANES respondents are likely to contact elected officials (20.9% vs. 22.9%) and participate in community work (28.5% vs. 30.1%). Relatedly, the samples are similar in terms of whether they have any say in what government does, as indicated by virtually equal means of 3.3 among the samples on a scale of 1 to 5. Although these findings appear to contradict those showing that MTurk respondents are less politically active (e.g., Berinsky et al., 2012), they are likely due to measurement differences. Others have mostly gauged activity with measures of voter turnout and registration, activities for which socio-demographic characteristics, namely age, are both strongly related to being an MTurk respondent and less significant predictors of the political activities measured here (Schlozman, Verba, & Brady, 2012).

That said, the samples differ when it comes to news media consumption, though in ways similar to recent work (Huff & Tingley, 2015). The difference between the samples in the average number of days spent getting news during a typical week is higher for the MTurk sample, but only with respect to Internet news (4.9 vs. 3.7). Alternatively, this average is about a day higher for both print and TV news for the ANES sample (2.1 vs. 1.1, and 3.8 vs. 2.6, respectively). Relatedly, the average level of attention paid to news among those who gather it from print or television at least 1 day or more during a typical week is lower among the MTurk sample (2.7 vs. 2.9 and 2.9 vs. 3.2, respectively, on a scale of 1-5), but higher when it comes to Internet news gatherers (3.1 vs. 2.8 on a scale of 1-5). These findings not only confirm previous work but also underscore that even as it underrepresents some groups, MTurk might be useful for attracting subgroups with more specific qualities. That is, MTurk samples might do relatively well at attracting large numbers of younger people who are highly interested in news and older people who are not, at least compared with population-based samples (see Huff & Tingley, 2015).

Multivariate Analyses of Socio-Economic Differences

We now turn to the important question of whether differences diminish when including measured covariates or reflect some fundamental and unmeasured distinction of MTurk respondents. We select sets of the broadest, most commonly measured covariates for which one can reasonably expect effects on more specific socio-demographic and political variables.

Table 4 reports the coefficients for MTurk sample membership (vs. ANES) from regressions of every socio-demographic variable, adding respondents' age, age^2, gender, and race and ethnicity as covariates; income and education are covariates in all models except for when they are outcomes.[9] For categorical outcomes, reported coefficients represent the average discrete change in the predicted probability resulting from being an MTurk respondent. We also report the percentage reduction of the coefficients from the models without any controls.

A quick glance at Table 4 suggests that differences between the samples are not simply a by-product of their more basic demographic characteristics. Indeed, being an MTurk respondent remains a significant predictor for 21 of

Table 4. Regression Analyses of Socio-Demographic Characteristics of MTurk and 2012 ANES Web Samples.

	MTurk coefficient	% reduction in coefficient
Income[a]	−11000.95***	—
Education (scale of 1-5; 5 = *graduate degree*)[a]	0.42***	19.1%
Spouse's education (scale of 1-5; 5 = *graduate degree*)	0.05	92.1%
Marital status		
Married	−0.09***	57.6%
Divorced	−0.01	85%
Separated	−0.01	40.7%
Widowed	−0.004	91.9%
Never married	0.08***	75.7%
Living with partner (not married)	0.01	76.9%
Own home	−0.20***	40.8%
Employment status		
Working now	−0.07***	—
Unemployed	0.06***	3.6%
Temporarily laid off	0.01***	56.8%
Retired	−0.03*	87%
Homemaker	0.01	910.3%
Permanently disabled	−0.02**	56.4%
Student	0.03***	79.9%
Number of hours working per week (if working now)	−0.81	52.2%
Out of work/laid off in last 6 months (if working now)	0.01	79.9%
Number of children in household ages 0-10 (scale of 0-2; 2 = *two or more*)	−0.01	—
Number of children in household ages 11-17 (scale of 0-2; 2 = *two or more*)	0.05**	94.3%
Has served on active duty in the Armed Forces	−0.04**	57.3%
Level of satisfaction with life (scale of 1-5; 1 = *not at all satisfied*, 5 = *extremely satisfied*)	−0.35***	22.3%
Self-assessment of health condition (scale of 1-5; 5 = *excellent*)	−0.12***	—
Religious identity		
Agnostic	0.12***	24.2%
Atheist	0.12***	40.1%
Nontraditional	−0.02	34.6%
Progressive	−0.04**	4.3%
Secular	0.01	16.4%
Spiritual—not religious	−0.01	84.3%
None	−0.19	2.7%
Religious group		
Protestant	0.05*	69.3%
Catholic	0.03	
Jewish	0.01	43%
Other	−0.07***	484.3%
Attends religious/church services	−0.22***	12.1%

Note. MTurk = Mechanical Turk; ANES = American National Election Studies 2012 Time Series Study. Variables in left-hand side column serve as dependent variable. Coefficients reflect impact of being a member of MTurk sample relative to being a member of 2012 ANES Web sample. Coefficients for categorical dependent variables are average marginal effects. Model 1 includes no controls. Model 2 includes controls for age, age^2, gender, race, income, and education. ANES data are weighted for all analyses. All data exclude respondents providing no answer or "Don't Know."

[a]Model includes controls for only age, age^2, gender, and race.

"—" indicates that coefficient is either unchanged or increases in size across models.

*Significant at the .1 level. **Significant at the .05 level. ***Significant at the .01 level.

35 variables, suggesting that other, unmeasured differences might exist among the samples. However—perhaps more revealingly—there are substantial reductions in the size of the MTurk coefficient across the models for most every variable (i.e., the *substantive difference of being an MTurk respondent becomes quite small*). The main exceptions are for employment status, earnings, and health assessment. Thus, while remaining statistical differences

should not be dismissed, respondents in the MTurk sample diverge much less from those in the ANES sample on most variables once basic demographics are accounted for—differences are more likely to reflect easily measurable features.

Multivariate Analyses of Political Differences

We next reconsider differences in the political attitudes and behaviors presented in Tables 2 and 3. We run regression models for each variable, now adding nine covariates: age, age,[2] gender, race and ethnicity, income, education, whether a respondent has never been married, whether a respondent identifies as atheist or agnostic, political ideology, and partisanship (the latter two are excluded when each is an outcome variable).

The results in Table 5 show that, as with socio-demographic variables, being an MTurk respondent remains statistically significant for most of the variables, 25 of 39. Nevertheless, the size of the coefficients is also *reduced considerably* across the models for nearly every variable, except for the number of days an individual spends gathering news from the Internet during a typical week. Even when statistically important, both the relative substantive impact of being an MTurk respondent is much smaller than other variables and these other variables predict outcomes in expected ways. For example, while statistically significant, an MTurk respondent is only 0.25 points more supportive of government insurance than an ANES respondent on a scale of 1 to 7, a small increase not only by itself, but also compared with the impact of other variables on such support, such as ideology (see the appendix). There are only a few notable exceptions to this pattern, such as death penalty attitudes and Internet news consumption. Overall then, as with our analysis of demographics, MTurk respondents *do not appear to fundamentally differ* (in terms of the substantive size of the effect) on political variables once we control for just a handful of variables, particularly ideology and partisanship.[10]

To summarize, our samples differ in ways that are similar to previous work, though not exactly. In line with past work, for example, our MTurk sample is much younger, less racially and ethnically diverse, and more liberal and Democratic than the ANES Internet sample. However, the sample differs less or differently from previous work when it comes to gender and organized religion. We also identified other less well-explored, though expected differences, namely, attitudes on a number of different political issues and policies. As important, we showed that the importance of being an MTurk respondent is reduced in most instances and negligible in many when controlling for relatively few, broad, and relevant factors. In other words, the MTurk sample differs from the ANES sample but this is explained fairly well by measurable differences.

Conclusion

Amazon's MTurk allows researchers to recruit large samples quickly at low cost. In turn, its presence among many types of behavioral science research has rapidly expanded (Bohannon, 2011). However, the relative representativeness of MTurk samples raises some concerns over inferences based on MTurk data, at least compared with those based on data from population-based samples, namely, generalizability (see Berinsky et al., 2012).

We sought to update and extend our knowledge of MTurk samples by comparing an MTurk sample with a nationally representative one on a large range of demographic and political variables. Our samples differed on variables in similar ways as in past work, but also some not included in extant scholarship. More importantly, these differences are *significantly reduced* when accounting for nine broad, measurable features by which MTurk samples differ starkly: age, gender, race and ethnicity, income, education, marital status, religion, ideology, and partisanship. Importantly, this set of variables is slightly broader than that most commonly identified as most consequential (e.g., age and ideology).

As a whole, these results imply that observational researchers of public opinion can profitably use MTurk, as long as they measure *at least* the nine covariates we identify. Doing so will allow for the development of sample weights commonly employed when using other types of opt-in samples. As Callegaro et al. (2014) explain, those who use non-probability samples (e.g., opt-in) "argue that the bias in samples . . . can be reduced through the use of auxiliary variables that make the results representative. These adjustments can be made with . . . [w]eight adjustments [using] a set of variables that have been measured in the survey" (p. 13). What we have done, then, is identified those variables that allow observational researchers to weight data from MTurk and reduce potential biases.

Still, measuring the covariates identified here is not a universal, foolproof solution for observational researchers interested in using MTurk. We employed a particular recruitment method and while broad in approach, nevertheless explored particular variables. Moreover, differences remain, however negligible, and some can apparently be inconsistent. Taken together, this suggests that additional covariates are more or less important to include depending upon the question at hand and the timing of sampling.

Our results are also useful for experimentalists of public opinion concerned with treatment effects that are expected to differ by particular characteristics. While the

Table 5. Regression Analyses of Political Characteristics and Behaviors of MTurk and 2012 ANES Web Samples.

	MTurk coefficient	% reduction
Political ideology (scale of 1-7; 1 = *extremely liberal*, 7 = *extremely conservative*)	—	—
Partisanship (scale of 1-7; 1 = *strong Democrat*, 7 = *strong Republican*)	—	—
Identify with Democratic or Republican Party strongly	−0.07*	43.1%
Favor government environmental regulation of business (scale of 1-7; 1 = *no regulation*, 7 = *regulate to protect environment and create jobs*)	−0.28***	—
Believe government should guarantee jobs and income (scale of 1-7; 1 = *let each person get ahead on own*, 7 = *see to a job and good standard of living*)	−0.09	—
Favor private insurance plans over government insurance plan (scale of 1-7; 1 = *should be government plan*, 7 = *should be private plans*)	−0.26***	74.8%
Government should provide more services and increase spending (scale of 1-7; 1 = *decrease spending/ fewer services*, 7 = *increase spending/more services*)	0.34***	55.9%
Federal spending on child care (scale of 0-2; 0 = *decrease*, 1 = *keep same*, 2 = *increase*)	0.12**	51.1%
Federal spending on environmental protection (scale of 0-2; 0 = *decrease*, 1 = *keep same*, 2 = *increase*)	0.08**	72.1%
Federal spending on aid to the poor (scale of 0-2; 0 = *decrease*, 1 = *keep same*, 2 = *increase*)	0.08**	53.8%
Federal spending on public schools (scale of 0-2; 0 = *decrease*, 1 = *keep same*, 2 = *increase*)	0.02	91.4%
Federal spending on science and technology (scale of 0-2; 0 = *decrease*, 1 = *keep same*, 2 = *increase*)	0.15***	54.5%
Support abortion (scale of 1-4; 1 = *never permit abortion*, 4 = *permit abortion in all circumstances*; see Table 5)	−.05	—
Agree that gay and lesbian couples should be allowed to adopt children	0.03	86.4%
Support lesbian and gay marriage (scale of 1-3; 1 = *no legal recognition*, 3 = *legal marriage*; see Table 5)	0.11***	77.3%
Favor offshore drilling (scale of 0-2; 0 = *oppose*, 1 = *neither*, 2 = *favor*)	−0.06*	83.7%
Agree that global warming is happening	−0.05	—
Agree that rising temperatures are good (scale of 0-2; 0 = *bad*, 1 = *neither*, 2 = *good*)	−0.06**	56.3%
Agree that climate change caused mostly by humans (scale of 0-2; 0 = *natural causes*, 1 = *both human activity/natural causes*, 2 = *human activity*)	0.01	95.3%
Favor checking status of those suspected to be undocumented (scale of 0-2; 0 = *oppose*, 1 = *neither*, 2 = *favor*)	−0.03	90.6%
Favor granting citizenship to unauthorized immigrants (mean on scale of 1-4; 1 = *make felons and deport*, 4 = *remain in country and qualify for citizenship*; see Table 5)	0.08*	68.7%
Increase level of foreigners permitted to enter the United States to live (scale of 1-5; 1 = *decrease a lot*, 5 = *increase a lot*)	0.04	88.8%
Favor strict gun control laws (mean on scale of 0-2; 0 = *make it easier to buy gun*, 1 = *keep rules the same*, 2 = *make it more difficult to buy gun*)	0.016	79.8%
Support the death penalty	−0.09***	54.1%
Feels strongly about position on death penalty (oppose or support)	−0.04	64.4%
Has contacted elected official to express view in past 12 months	−0.02	13%
Has participated in community work in past 12 months	−0.06**	198.3%
Agree that he or she has little say in what government does (scale of 1-5; 1 = *disagree strongly*, 5 = *agree strongly*)	−0.07	—
Number of days spent reviewing news (scales of 0-7)		
On Internet	1.01***	18.1%
In print newspaper	−0.27***	74%
On television	0.18	—
Level of attention to news (scales of 0-5; 1 = *none at all*, 5 = *a great deal*)		
On Internet	0.25***	5%
In print newspaper	−0.67***	18%
On television	−0.31***	45.7%

Note. MTurk = Mechanical Turk; ANES = American National Election Studies 2012 Time Series Study. Variables in left-hand side column serve as dependent variable. Coefficients reflect impact of being a member of MTurk sample relative to being a member of 2012 ANES Web sample. Coefficients for categorical dependent variables are average marginal effects. Model 1 includes no controls. Model 2 includes controls for age, age^2, gender, race, income, education, whether a respondent has never been married, whether a respondent identifies as atheist or agnostic, and respondent ideology and party ID. ANES data are weighted for all analyses. All data exclude respondents providing no answer or "Don't Know."

"—" indicates that coefficient either is unchanged or increased in size across models.

*Significant at the .1 level. **Significant at the .05 level. ***Significant at the .01 level.

representativeness of MTurk does not threaten experimental inferences in most cases (e.g., Mullinix et al., 2015), one might nevertheless misestimate treatment effects if groups who are expected to respond to an experimental treatment differently than others are underrepresented and/or differ in unmeasurable ways from others like them

who are not included in the sample. Importantly, however, we find not only that the impact of being an MTurk respondent is reduced when controlling for other factors, but also that these factors behave in expected ways. While not definitive by any means, this provides at least some evidence that even while underrepresented, individuals belonging to groups who opt into being an MTurk worker that are underrepresented might not differ all that much from those who do not. As such, researchers can proceed with some confidence that increasing the number of people with a given characteristic in their sample might be sufficient for properly estimating treatment effects across or among those individuals. In any case, researchers should always analyze and situate findings in the context of their sample, both in process and composition.

With the above in mind, as highlighted by others (e.g., Huff & Tingley, 2015), a promising approach for researchers concerned with representativeness or with specific sample needs is to employ software and sampling procedures aimed at creating larger and/or targeted survey pools. While varying in the type and amount of resources required, programs developed by researchers and other programmers, such as MTurk R (see Leeper, 2013) and TurkPrime, as well as implementing multiple demographic surveys on MTurk with unique samples of workers, readily provide a means of creating samples with more diverse and/or specific demographic and political characteristics. Thus, depending upon researchers' needs, these methods can be usefully leveraged to address potential sampling issues introduced by using MTurk.

To be sure, we reiterate that MTurk does not provide a substitute for population-based samples; indeed, as mentioned, even with our set of covariates, sample differences remained. However, with proper care and measurement approaches, researchers can use it to adequately build research programs. Our identification of key sources of sample differences offers a blueprint for what scholars can do in using MTurk and in turn, maximize the likelihood that they can draw useful inferences from MTurk. In sum, as with any source of data, to get the most out of MTurk, researchers should critically examine their sampling strategy and composition in relation to their expectations and the question at hand.

Appendix

Table A1. Distributions of Responses for Select Variables.

	MTurk sample (n=2,056-2,044)	ANES Web Sample (n=3,547-3,853)
Education (mean on scale of 1-5)	3.4	2.9
1. *Less than high school*	1.3%	10.3%
2. *High school credential*	10.0%	29.9%
3. *Some post-high school*	43.8%	30.2%
4. *Bachelor's degree*	35.8%	18.7%
5. *Graduate degree*	9.1%	10.8%
Political ideology (mean on scale of 1-7)	3.3	4.2
1. *Extremely liberal*	11.1%	3.1%
2. *Liberal*	26.0%	11.1%
3. *Slightly liberal*	21.3%	12.6%
4. *Moderate/middle of road*	19.7%	35.9%
5. *Slightly conservative*	9.6%	13.6%
6. *Conservative*	9.5%	19.4%
7. *Extremely conservative*	2.8%	4.4%
Partisan scale (mean on scale of 1-7)	3.1	3.8
1. *Strong Democrat*	20.4%	19.5%
2. *Democrat*	25.8%	16.5%
3. *Lean Democrat*	16.6%	9.6%
4. *Does not identify with either party*	15.7%	15.7%
5. *Lean Republican*	7.0%	9.4%
6. *Republican*	9.1%	13.6%
7. *Strong Republican*	5.5%	15.7%

Note. MTurk = Mechanical Turk; ANES = American National Election Studies 2012 Time Series Study.

Table A2. Full Regression Results, Socio-Demographic Characteristics.

	Income	Education	Spouse's education	Married	Divorced	Separated	Widowed	Never married	Live with partner	Own home	Working now	Unemployed	Laid off	Retired	Homemaker	Disabled	Student	Number of hours working per week	If working-laid off last 6 months
Constant	26245.18***	2.18***	7.11***	−5.47***	−9.10***	−4.69***	−7.87***	5.75***	−1.99***	−2.48***	−5.96***	−2.11***	−8.33***	−17.98***	−2.38***	−7.22***	7.35***	−11.95***	−0.78
MTurk respondent	−11000.95***	0.42***	0.05	−0.45***	−0.10	−0.50*	−0.14	0.55***	0.05	−1.04***	−0.38***	0.66***	0.90***	−0.50**	0.18	−0.58**	0.47***	−0.80	0.05
Male	4,042.37***	0.07**	0.09	−0.09	−0.33***	0.11	−1.28***	0.58***	−0.48***	−0.01	0.51***	0.23**	0.54*	0.03	−2.99***	0.10	0.12	5.83***	−0.14
Age	1,489.55***	0.04***	0.02	0.19***	0.29***	0.15***	0.09**	−0.24***	0.12***	0.07***	0.27***	0.05**	0.22***	0.35*	0.11***	0.29***	−0.44***	0.96***	−0.02
Age^2	−16.58***	−0.0005***	−0.0001	−0.002***	−0.003***	−0.001***	0.0002	0.002***	−0.002***	−0.0004**	−0.003***	−0.0008***	−0.002***	−0.0009	−0.001***	−0.002***	0.004***	−0.009***	0.00007
Education			0.99***	−0.04	0.18***	−0.12	−0.20**	0.07	−0.09	0.05	0.43***	−0.29***	−0.36**	−0.02	−0.34***	−0.48***	0.16**	0.89***	−0.03
Income			$7.69 * 10^{-6}$***	0.00001***	−0.00001***	−0.00002***	−0.00001***	$-0.7*10^{-6}$***	$3.1*10^{-6}$**	0.00001***	$3.3*10^{-6}$***	$-7.2*10^{-6}$**	−0.00002***	$-1.3*10^{-6}$	$2.97*10^{-6}$***	−0.00002***	$6.8*10^{-7}$	0.00002***	−0.00001***
White	10281.27**	0.07	0.59	0.66***	−0.07	−1.49***	−0.12	−0.56**	−0.91***	0.39	0.22	0.01	0.84	−0.37	−0.07	−0.62	−0.79***	−1.51	0.31
Black	−5538.21	−0.26**	0.03	−0.28	0.01	−0.66	0.39	0.34	−0.83***	−0.49*	−0.23	0.79**	0.92	−0.18	−0.80*	−0.38	−0.36	−2.73	0.31
Hispanic/ Latino/a	−3529.67	−0.44***	−0.60	0.46*	0.31	−0.79	−0.06	−0.57*	−0.74**	−0.18	−0.07	0.37	0.62	−0.57	0.04	−0.51	−0.64*	−3.29	0.48
Other race	13549.4**	0.24*	0.95***	0.49*	−1.14***	−0.75	−0.11	−0.14	−1.24***	−0.03	−0.17	0.17	1.44	−0.57	−0.15	−0.37	−0.44	−1.16	−0.44

Table A2. Full Regression Results, Socio-Demographic Characteristics.

	Number of children age 0-10 (0 = *none*, 2 = *two or more*)	Number of children age 11-17 (0 = *none*, 2 = *two or more*)	Served in Armed Forces	Life satisfaction (1 = *not at all satisfied*, 5 = *extremely satisfied*)	Health assessment (1 = *poor*, 2 = *excellent*)	Attends religious services	Protestant	Catholic	Jewish	Other religious group	Agnostic	Atheist	Nontraditional	Progressive	Secular	Spiritual-not religious	No religious identity
Constant	0.18**	−0.19**	−5.49***	3.47***	3.28***	−0.31	−2.39***	−1.78***	−0.89	0.95	−2.74***	−0.89*	−2.45***	−1.94***	−3.55***	−2.84***	0.51
MTurk respondent	−0.01	0.05***	−0.53***	−0.35***	−0.12***	−0.96***	0.23**	0.16	0.26	−0.42***	1.23***	1.39***	−0.17	−0.34***	0.24	−0.05	−1.01***
Male	−0.07***	−0.03***	2.53***	−0.08***	0.01	−0.23***	0.19**	0.18*	−0.03	−0.43***	0.49***	0.61***	0.09	−0.09	0.08	−0.14*	−0.29***
Age	0.01***	0.02***	0.02	−0.04***	−0.02***	−0.02	0.008	0.006	−0.11***	0.04*	−0.04*	−0.09***	0.01	−0.01	−0.04	0.10***	−0.04***
Age^2	−0.0002***	−0.0002***	0.0003	0.0003***	0.0001***	0.0002**	0.0001	0.00005	0.001***	−0.0008***	0.0002	0.0007***	−0.00006	0.0002	0.0004	−0.0009***	0.0003**
Education	0.003	−0.04***	0.08	0.11***	0.18***	0.13***	0.28***	−0.07	0.26*	−0.34***	0.17***	0.19***	0.02	0.15***	0.31***	−0.04	−0.09**
Income	$5.07*10^{-7}$***	$5.98*10^{-7}$***	$-4.51*10^{-7}$	$3.62*10^{-6}$***	$2.8*10^{-6}$***	$3.9*10^{-7}$	$1.6*10^{-6}$**	$4.8*10^{-6}$***	$4.6*10^{-6}$***	$-4.6*10^{-6}$***	$1.3*10^{-6}$	$2.4*10^{-7}$	$-7.2*10^{-8}$	$1.9*10^{-7}$	$1.7*10^{-6}$	$4.7*10^{-7}$	$-6.0*10^{-7}$
White	0.02	0.06*	−0.31	0.07	0.13	0.16	0.59*	0.22	−1.18**	−0.52	0.28	−0.29	−0.02	−0.005	0.13	−0.22	0.24
Black	0.03	0.13***	0.09	0.01	0.06	1.14***	0.32	−1.47***	−1.73**	0.76**	−0.45	−1.59***	−0.32	0.49	0.21	−0.06	0.26
Hispanic/ Latino/a	0.09	0.11***	−0.29	0.11	0.08	0.40	−0.74*	1.66***	−1.94**	−0.96**	−0.29	−0.90***	0.09	−0.03	0.09	−0.26	0.46
Other race	−0.05	0.07*	−0.30	−0.09	0.04	0.03	0.09	−0.19	−1.37	0.47	−0.28	−0.45	−0.09	0.06	−0.03	−0.19	0.33

Note. Variables in left-hand side column serve as independent variable. Variables in top row serve as dependent variable. Coefficients reflect impact of being a member of MTurk sample relative to being [a] member of 2012 ANES Web sample. Log-odds coefficients reported for categorical dependent variables. ANES data are weighted for all analyses. All data exclude respondents providing no answer or "Don't Know."
*Significant at the .1 level. **Significant at the .05 level. ***Significant at the .01 level.

Table A3. Full Regression Results, Political Characteristics and Behaviors.

	Govt. regulate business (1-7; 7 = *regulate*)	Govt. guarantee jobs/ income (1-7; 7 = *guarantee*)	Private vs. govt. health insurance (1-7; 7 = *private plan*)	Govt. services/ spending (1-7; 7 = *increase*)	Fed. spending on child care (0-2; 2 = *increase*)	Fed. spending on environment (0-2; 2 = *increase*)	Fed. spending on aid to poor (0-2; 2 = *increase*)	Fed. spending on public schools (0-2; 2 = *increase*)	Fed. spending on science and tech (0-2; 2 = *increase*)	Support abortion (1-4; 4 = *permit in all circumstances*)	Gay and lesbian adoption (0/1; 1 = *support*)	Gay and lesbian marriage (1-3; 3 = *support legalization*)	Favor offshore drilling (0-2; 2 = *favor*)	Global warming happening (0/1; 1 = *agree*)	Favor checking undocumented (0-2; 2 = *favor*)	Favor full citizenship for immigrants (1-4; 4 = *favor*)
Constant	6.57***	6.36***	−1.89***	5.99***	1.87***	2.05***	1.89***	2.49***	1.45***	3.09***	4.33***	3.10***	0.18	2.72***	0.33**	3.20***
MTurk respondent	−0.28***	−0.09	−0.26***	0.34***	0.12***	0.08***	0.08***	0.02	0.14***	−0.05	0.17	0.11***	−0.06**	−0.04	−0.03	0.08**
Male	−0.13**	−0.11**	−0.03	−0.17***	−0.04*	−0.08***	−0.07***	−0.10***	0.12***	−0.03	−0.43***	−0.05**	0.15***	−0.09	0.02	−0.04
Age	0.03***	0.01	−0.03***	0.02***	−0.004	0.001	0.01***	−0.006	−0.01**	0.02***	−0.03*	−0.009**	0.009**	0.02	0.01**	−0.02***
Age^2	−0.0004***	−0.0002	0.0003**	−0.0002**	0.00003	−0.00002	−0.0001***	$-1.8*10^{-6}$	0.0001***	−0.0002***	0.0002	0.00007	−0.00005	−0.0002	−0.00007	0.0002***
Education	0.08***	−0.11***	0.014	−0.10***	−0.02**	−0.002	−0.08***	−0.02*	0.08***	0.07***	0.15***	0.07***	−0.03***	0.14***	−0.08***	0.09***
Income	$-3.72*10^{-7}$	$-2.8*10^{-6}$***	$2.4*10^{-6}$***	$-1.6*10^{-6}$***	$7.5*10^{-7}$***	$2.9*10^{-8}$	$1.3*10^{-6}$***	$-1.6*10^{-7}$	$8.0*10^{-7}$***	$1.9*10^{-6}$***	$1.7*10^{-6}$*	$3.3*10^{-7}$	$7.1*10^{-7}$***	$1.4*10^{-6}$	$2.3*10^{-7}$	$4.9*10^{-7}$*
White	−0.07	−0.08	−0.11	−0.2	0.009	0.02	−0.04	0.05	0.09	0.01	−0.23	0.09	0.10	0.18	0.04	−0.01
Black	−0.18	0.14	−0.01***	0.29*	0.24***	0.02	0.22***	0.12	0.12	−0.12	−1.13***	−0.28***	0.09	0.09	−0.09	0.05
Hispanic/Latino/a	−0.03	0.22	−0.24***	0.03	0.11	0.06	0.03	0.03	0.11	−0.18	−0.89***	0.03	0.04	0.24	−0.44***	0.36***
Other race	−0.13	0.21	−0.05	−0.06	0.03	0.08	−0.0006	0.05	0.15*	−0.12	−1.06***	−0.09	0.06	0.05	−0.11	−0.08
Agnostic	0.005	−0.05	−0.20**	−0.08	0.009	−0.05	−0.03	−0.06	0.11***	0.36***	1.00***	0.22***	−0.08*	0.25	−0.09**	0.02
Atheist	0.13	−0.17**	−0.38***	0.04	−0.007	0.02	−0.06	−0.02	0.22***	0.47***	0.91***	0.19***	−0.15***	0.49**	−0.18***	0.11**
Never married	0.09	0.13*	−0.10	0.17***	−0.003	0.07**	0.11***	−0.06**	0.01	0.22***	0.22*	0.07**	−0.09***	0.25*	−0.03	−0.05
Ideology (1-7; 7 = *extremely conservative*)	−0.38***	−0.35***	0.49***	−0.38***	−0.11***	−0.15***	−0.14***	−0.11***	−0.09***	−0.20***	−0.54***	−0.17***	0.14***	−0.38***	0.15***	−0.11***
Partisanship (1-7; 7 = *strong Rep.*)	−0.24***	−0.18***	0.24***	−0.19***	−0.05***	−0.07***	−0.06***	−0.06***	−0.03***	−0.07***	−0.14***	−0.06***	0.05***	−0.23***	0.07***	−0.04***

Note. Variables in left-hand side column serve as independent variable. Variables in top row serve as dependent variable. Coefficients reflect impact of being a member of MTurk sample relative to being [a]member of 2012 ANES Web sample. Log-odds coefficients reported for categorical dependent variables. ANES data are weighted for all analyses. All data exclude respondents providing no answer or "Don't Know."
*Significant at the .1 level. **Significant at the .05 level. ***Significant at the .01 level.

Table A3. Full Regression Results, Political Characteristics and Behaviors.

	Increase permissible level of foreigners in United States (1-5; 5 = *increase a lot*)	Favor stricter gun control (0-2; 2 = *make more difficult to buy gun*)	Support death penalty (0/1; 1 = *support*)	Strongly oppose/ support death penalty (0/1; 1 = *yes*)	Contacted elected official in past year (0/1; 1 = *made contact*)	Participated in community work in past year (0/1; 1 = *participated*)	Have little say in what govt. does (1-5; 5 = *agree strongly*)	No. of days/ week spent reviewing news on Internet	No. of days/ week spent reviewing news in print media	No. of days/ week spent reviewing news on TV	Level of attention when viewing news on Internet (1-5; 5 = *great deal*)	Level of attention when viewing news in print media (1-5; 5 = *great deal*)	Level of attention when viewing news on TV (1-5; 5 = *great deal*)
Constant	3.24***	1.71***	−0.56	−0.79*	−1.76***	−0.04	3.92***	1.57***	0.89**	0.57	2.24***	2.69***	2.43***
MTurk respondent	0.04	0.02	−0.47***	−0.17*	−0.13	−0.31***	−0.07	1.01***	−0.27***	0.18*	0.25***	−0.67***	−0.31***
Male	0.11***	−0.17***	0.16**	0.31***	0.18**	0.13	0.02	0.55***	0.16**	0.07	0.32***	0.17***	0.15***
Age	−0.03**	−0.002	0.05***	0.04***	−0.0003	−0.03*	0.009	0.05***	−0.04***	0.05***	−0.0002	−0.01	−0.003
Age^2	0.0002***	−0.00005	−0.0005***	−0.0004***	0.0001	0.0002	−0.0002*	−0.0005***	0.001***	0.0002	0.00007	0.0003***	0.0002
Education	0.15***	0.04***	−0.17***	−0.10***	0.25***	0.26***	−0.13***	0.38***	0.07*	−0.14***	0.16***	0.09***	0.09***
Income	$9.4*10^{-7}$**	$4.2*10^{-6}$*	$8.1*10^{-7}$	$-7.7*10^{-7}$	$8.4*10^{-7}$	$5.9*10^{-7}$	$-9.9*10^{-7}$**	$3.1*10^{-6}$***	$2.4*10^{-6}$***	$9.6*10^{-7}$	$5.5*10^{-8}$	$7.6*10^{-9}$	$-3.2*10^{-7}$
White	0.003	0.14*	−0.09	0.14	−0.53**	−0.73***	−0.11	−0.35	0.12	0.28	−0.16	−0.08	0.02
Black	0.12	0.16**	−0.55**	0.03	−0.35	−0.30	−0.39**	−0.33	−0.0006	0.89***	0.02	0.08	0.29*
Hispanic/ Latino/a	0.47***	0.24***	−0.52*	0.05	−0.61**	−0.99***	−0.49***	−0.40	0.03	0.79***	0.03	0.01	0.10
Other race	0.17***	0.27***	−0.07	0.11	−0.79***	−0.74***	−0.12	−0.45	0.0001	0.59**	−0.17	0.03	0.03
Agnostic	0.07	−0.05	−0.41***	−0.08	0.21	−0.19	−0.006	0.21	−0.10	−0.28**	0.08	−0.02	−0.08
Atheist	0.18***	−0.04	−0.29***	0.09	0.24*	−0.37***	0.15**	0.27**	−0.24**	−0.38***	0.12***	−0.15**	−0.03
Never married	−0.009	0.02	−0.11	−0.13	−0.005	−0.06	−0.09	−0.12	0.24**	0.11	−0.08	0.03	−0.03
Ideology (1-7; 7 = *extremely conservative*)	−0.12***	−0.08***	0.19***	0.09***	−0.06	−0.03	−0.03	−0.07	0.02	0.05	−0.05**	−0.05	−0.01
Partisanship (1-7; 7 = *strong Rep.*)	−0.03**	−0.06***	0.11***	0.04	0.01	0.008	0.05**	0.01	−0.009	0.002	0.04***	0.03*	0.03**

Note. Variables in left-hand side column serve as independent variable. Variables in top row serve as dependent variable. Coefficients reflect impact of being a member of MTurk sample relative to being [a] member of 2012 ANES Web sample. Log-odds coefficients reported for categorical dependent variables. ANES data are weighted for all analyses. All data exclude respondents providing no answer or "Don't Know."
*Significant at the .1 level. **Significant at the .05 level. ***Significant at the .01 level.

Table A4. Models of Health Assessment, Currently Working, Number of Children, and Other Religious Group ID.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Self-assessment of health condition (mean on scale of 1-5; 1 = *poor*, 5 = *excellent*)	0.12***	0.12***	−0.05	−0.164***	−0.117***	−0.121***
Employment Status: working now	0.04**	0.06***	−0.02	−0.07***	−0.07***	−0.07***
Number of children in household ages 11-17 (scale of 0-2; 0 = *none*, 2 = *two or more*)	0.03	0.03*	0.01	0.03	0.04**	0.05**
Other religious group	−0.01	−0.01	−0.09***	−0.05*	−0.06**	−0.07***

Note. Variables in left-hand side column serve as dependent variable. Coefficients reflect impact of being a member of MTurk sample relative to being a member of 2012 ANES Web sample. Model 1 includes no controls. Model 2 includes a control for gender. Model 3 includes controls for gender, age, and age^2. Model 4 includes controls for gender, age, age^2, and education. Model 5 includes controls for gender, age, age^2, education, and income. Model 6 includes controls for gender, age, age^2, education, income, and race. ANES data are weighted for all analyses. All data exclude respondents providing no answer or "Don't Know."
*Significant at the .1 level. **Significant at the .05 level. ***Significant at the .01 level.

Table A5. Models of Support For Government Regulation of Business.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Favor government environmental regulation of business (scale of 1-7; 1 = *no regulation*, 7 = *regulate to protect environment and create jobs*)	0.41***	0.42***	0.25**	0.21**	0.18*	0.23**	0.03	0.01	−0.24***	−0.28***

Note. Variable in left-hand side column serves as dependent variable. Coefficients reflect impact of being a member of MTurk sample relative to being a member of 2012 ANES Web sample. Model 1 includes no controls. Model 2 includes a control for gender. Model 3 includes controls for gender, age, and age^2. Model 4 includes controls for gender, age, age^2, and education. Model 5 includes controls for gender, age, age^2, education, and income. Model 6 includes controls for gender, age, age^2, education, income, and race. Model 7 includes controls for gender, age, age^2, education, income, race, whether a respondent identifies as atheist, and whether a respondent identifies as agnostic. Model 8 includes controls for gender, age, age^2, education, income, race, whether a respondent identifies as atheist, whether a respondent identifies as agnostic, and whether a respondent reports having never been married. Model 9 includes controls for gender, age, age^2, education, income, race, whether a respondent identifies as atheist, whether a respondent identifies as agnostic, whether a respondent reports having never been married, and respondent ideology. Model 10 includes controls for gender, age, age^2, education, income, race, whether a respondent identifies as atheist, whether a respondent identifies as agnostic, whether a respondent reports having never been married, and respondent ideology and party ID. ANES data are weighted for all analyses. All data exclude respondents providing no answer or "Don't Know."
*Significant at the .1 level. **Significant at the .05 level. ***Significant at the .01 level.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research and/or authorship of this article: The authors acknowledge support from a National Science Foundation grant for Time-Sharing Experiments in the Social Sciences (SES-1227179).

Notes

1. Mutz (2011) defines a population-based sample as one that is representative of the target population of interest for a particular theory.
2. The importance of having a probability-based panel (population) sample for conducting observational research as opposed to any type of opt-in non-probability sample remains debated (Baker et al., 2010). What we have in mind, however, is that researchers could use a wide range of measured variables to adjust opt-in samples to more closely resemble population-based samples (e.g., Gelman & Rothschild, 2014).
3. Recent evidence, however, suggests that MTurk samples might represent some, more specific, non-White subgroups fairly well, such as younger, Hispanic females and Latinas, as well as members of certain occupational categories (Huff & Tingley, 2015).
4. For more information, consult the 2012 American National Election Studies (ANES) Time Series Study User Guide and Codebook at http://electionstudies.org/studypages/anes_timeseries_2012/anes_timeseries_2012_userguidecodebook.pdf
5. The ANES questionnaire can be accessed via Internet here: http://electionstudies.org/studypages/anes_timeseries_2012/anes_timeseries_2012.htm
6. Keywords describing the Human Intelligence Task (HIT) included survey, opinion, politics, public opinion, policy, and demographics. The HIT description read as follows: *Please participate in the National Public Opinion Survey. It should take about 20-30 minutes to complete the survey*. The introductory text to the survey read as follows: *For this study, we would like you to complete the National Public Opinion Survey. It has taken others 20-30 minutes on average to complete. In appreciation of your participation, we can give you $2.50 for completing the survey. To receive compensation for taking the survey, you must complete the entire survey. This requires you to both: Enter your Amazon Worker ID at the designated spot at the end of the survey, and enter a unique numeric code (that we will provide you with) into the designated Amazon Mechanical Turk screen.*
7. Respondents were able to select more than one racial category. We recoded the data so that White includes respondents who

selected only White, regardless of Hispanic/Latino/a identification; Black includes only respondents who selected Black *and* answered no to identifying as Hispanic or Latino/a; Hispanic/Latino/a includes only respondents who answered yes to identifying as Hispanic or Latino/a and did *not* select White racial identification; Other includes only respondents who selected at least one of the following: Asian, American Indian/Alaska Native, Native Hawaiian/other Pacific Islander, other, or multiple categories (excluding those who identify as Hispanic/Latino/a); no race includes only respondents who did not indicate any racial or Hispanic/Latino/a identification.

8. Tables only report means for ease of interpretation and space. Full distributions of responses for variables discussed are available in Appendix Table A1. Distributions for all other variables are available from authors upon request.
9. We also ran models for income and education with controls for age, age^2, gender, and race/ethnicity. Full regression results available in the appendix.
10. Interestingly, in one instance, support for regulation of business to protect the environment, the MTurk coefficient remains statistically significant but the direction changes; MTurk respondents are less favorable once we control for differences in partisanship and ideology. Nevertheless, as with most other variables, the size of the coefficient indicates minimal substantive effect.

References

Arceneaux, K. (2012). Cognitive biases and the strength of political arguments. *American Journal of Political Science*, *56*, 271-285.

Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., . . . Zahs, D. (2010). Research synthesis: AAPOR report on online panels. *Public Opinion Quarterly*, *74*, 711-781.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*, 351-368.

Bohannon, J. (2011). Social science for pennies. *Science*, *334*, 307.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*, 3-5.

Callegaro, M., Baker, R., Bethlehem, J., Göritz, A. S., Krosnick, J. A., & Lavrakas, P. J. (2014). Online panel research: History, concepts, applications, and a look at the future. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 1-22). West Sussex, UK: John Wiley.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*, 2156-2160.

Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, *64*, 464-494.

Couper, M. P. (2011). The future of modes of data collection. *Public Opinion Quarterly*, *75*, 889-908.

Gelman, A., & Rothschild, D. (2014, August 4). Modern polling needs innovation, not traditionalism. *The Monkey Cage*. Retrieved from https://www.washingtonpost.com/news/monkey-cage/wp/2014/08/04/modern-polling-requires-both-sampling-and-adjustment/

Gerber, A. S., Huber, G. A., Doherty, D., & Dowling, C. M. (2011). Citizens' policy confidence and electoral punishment: A neglected dimension of electoral accountability. *Journal of Politics*, *73*, 1206-1224.

Huber, G. A., Hill, S. J., & Lenz, G. S. (2012). Sources of bias in retrospective decision making: Experimental evidence on voters' limitations in controlling incumbents. *American Political Science Review*, *106*, 720-741.

Huber, G. A., & Paris, C. (2013). Assessing the programmatic equivalence assumption in question wording experiments: Understanding why Americans like assistance to the poor more than welfare. *Public Opinion Quarterly*, *77*, 385-397.

Huff, C., & Tingley, D. (2015). "Who are these people?": Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research and Politics*, *2*, 1-12.

Ipeirotis, P. G. (2010). *Demographics of Mechanical Turk* (CeDER Working Paper No. 10-01). New York University. Retrieved from http://www.ipeirotis.com/wp-content/uploads/2012/02/CeDER-10-01.pdf

Krupnikov, Y., & Levine, A. S. (2014). Cross-sample comparisons and external validity. *Journal of Experimental Political Science*, *1*, 59-80.

Leeper, T. J. (2013). Crowdsourcing with R fand the MTurk API. *The Political Methodologist*, *2*, 2-7.

Liu, M., & Wang, Y. (2015). Data collection mode effect on feeling thermometer questions: A comparison of face-to-face and web surveys. *Computers in Human Behavior*, *48*, 212-218.

Mason, W. A., & Watts, D. J. (2009). Financial incentives and the "performance of crowds." *Association for Computing Machinery Explorations Newsletter, 11*, 100-108.

Mullinix, K. J., Leeper, T. J., Freese, J., & Druckman, J. N. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, *2*, 109-138.

Mutz, D. C. (2011). *Population-based survey experiments*. Princeton, NJ: Princeton University Press.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, *23*, 184-188.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*, 411-420.

Schlozman, K. L., Verba, S., & Brady, H. E. (2012). *The unheavenly chorus: Unequal voice and the broken promise of American democracy*. Princeton, NJ: Princeton University Press.

Struminskaya, B., Kaczmirek, L., Schaurer, I., & Bandilla, W. (2014). Assessing representativeness of a probability-based online panel in Germany. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 61-85). West Sussex, UK: John Wiley.

Author Biographies

Kevin E. Levay is a PhD candidate in the Department of Political Science at Northwestern University. He currently is completing a dissertation on the role of race, class, and gender in the politics of incarceration in the United States.

Jeremy Freese is professor of Sociology at Stanford University. He currently is co-Principal Investigator of Time-Sharing Experiments for the Social Sciences (TESS) and the General Social Survey (GSS).

James N. Druckman is the Payson S. Wild Professor of Political Science and faculty fellow at the Institute for Policy Research at Northwestern University. He currently is co-principal investigator of Time-Sharing Experiments for the Social Sciences (TESS) and co-author of the recent book *Who Governs? Presidents, Public Opinion, and Manipulation* (University of Chicago Press).