

Are samples drawn from Mechanical Turk valid for research on political ideology?

Research and Politics
October-December 2015: 1–9
© The Author(s) 2015
DOI: 10.1177/2053168015622072
rap.sagepub.com


Scott Clifford, Ryan M Jewell and Philip D Waggoner

Abstract

Amazon's Mechanical Turk (MTurk) is an increasingly popular tool for the recruitment of research subjects. While there has been much focus on the demographic differences between MTurk samples and the national public, we know little about whether liberals and conservatives recruited from MTurk share the same psychological dispositions as their counterparts in the mass public. In the absence of such evidence, some have argued that the selection process involved in joining MTurk invalidates the subject pool for studying questions central to political science. In this paper, we evaluate this claim by comparing a large MTurk sample to two benchmark national samples – one conducted online and one conducted face-to-face. We examine the personality and value-based motivations of political ideology across the three samples. All three samples produce substantively identical results with only minor variation in effect sizes. In short, liberals and conservatives in our MTurk sample closely mirror the psychological divisions of liberals and conservatives in the mass public, though MTurk liberals hold more characteristically liberal values and attitudes than liberals from representative samples. Overall, our results suggest that MTurk is a valid recruitment tool for psychological research on political ideology.

Keywords

Mechanical Turk, sample comparison, external validity

Researchers are increasingly turning to Amazon's Mechanical Turk (MTurk) to recruit subjects for public opinion research (e.g., Ahler, 2014; Arceneaux, 2012; Clifford, 2014; Grimmer et al., 2012; Huber and Paris, 2013; Johnston et al., 2015). MTurk allows the rapid recruitment of a diverse sample of subjects at a dramatically lower cost than professional online panels (Berinsky et al., 2012). Researchers have taken numerous approaches to validating MTurk as a sample recruitment tool. Experimental results have been replicated on MTurk across a variety of topics, including framing effects (Berinsky et al., 2012; Weinberg et al., 2014), decision-making biases (Goodman et al., 2013; Paolacci et al., 2010), economic games (Horton et al., 2011), and cognitive psychology tasks (Crump et al., 2013). Others have shown that data from MTurk samples meet common psychometric standards (Buhrmester et al., 2011; Shapiro et al., 2013). MTurk respondents also pay as much or more attention than respondents from other populations (Hauser and Schwarz, forthcoming; Paolacci et al., 2010; Weinberg et al., 2014).

Yet, there are a number of concerns about the use of MTurk as well (e.g., Chandler et al., 2014; Krupnikov and Levine, 2014). Much of the debate over the validity of MTurk as a recruitment tool has followed the discipline's "near obsession" with the external validity of a sample (McDermott, 2002: 334). Researchers consistently find that MTurk samples tend to be more politically liberal, younger, less religious, and less racially diverse than the U.S. population (Berinsky et al. 2012; Huff and Tingley, 2015).

While scholars have documented many of the differences and similarities between MTurk samples and the national public, some critics maintain that the sample is invalid for political research.¹ According to this view, the selection process that produces disproportionately liberal samples implies

University of Houston, USA

Corresponding author:

Scott Clifford, University of Houston, 447 Philip Guthrie Hoffman Hall, Houston, TX 77204, USA.
Email: sclifford@uh.edu



that the conservatives who opt into MTurk differ from other conservatives in psychological dispositions central to their identities. If this claim were true, it may render MTurk samples invalid for studying political and ideological divides. This would be particularly worrisome for research using ideology or partisanship as a moderator of experimental treatment effects (e.g., Bullock, 2011; Druckman et al., 2013; Nyhan and Reifler, 2010) or examining psychological differences between liberals and conservatives (e.g., Feldman and Johnston, 2014). Importantly, the lack of substantial systematic evidence for whether the selection process generates samples that are valid for psychological research on ideological divisions in the mass public has limited this debate.

In this paper, we first review existing evidence on the psychological differences between respondents drawn from MTurk and alternative populations. We then compare psychological models of political ideology across American National Election Studies (ANES) and MTurk samples. We find that conservatives recruited from MTurk look nearly identical to their counterparts in national samples, though MTurk liberals tend to have more liberal dispositions than ANES liberals. We next model political ideology as a function of personality and values, and find strikingly similar results across samples. Overall, our findings suggest that research into the psychological dispositions behind political ideology would reach largely the same conclusions regardless of the sample chosen. Our results suggest, with some caveats, that MTurk is a valid recruitment tool for psychological research on political ideology.

Psychological differences between liberals and conservatives

Liberals and conservatives differ in a variety of psychological dispositions, most notably in their personality traits and value orientations. Decades of research has established the Big Five personality traits—Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability—as a useful framework for understanding stable dispositions that underlie individual behavior (Costa and McCrae, 1992). Two of the Big Five traits, in particular, are consistently strong predictors of ideology. Liberals tend to score higher in Openness, demonstrating greater interest in new information and ideas (Gerber et al., 2013). Conservatives tend to score higher in Conscientiousness, demonstrating greater dependability and self-discipline (Gerber et al., 2013). Findings for other traits are less consistent, but there is some evidence that Extraversion is associated with conservatism (e.g., Gerber et al., 2010; Mondak and Halperin, 2008), and that Agreeableness predicts liberalism (Mondak, 2010).

Liberals and conservatives differ even more strongly in their values. We focus on four values commonly used in research on political attitudes and identities: egalitarianism, moral traditionalism, authoritarianism, and racial resentment. The first three values represent the preference for equal

opportunity, traditional family structures, and conformity and obedience, respectively. The fourth, racial resentment, represents a blend of racial prejudice and the values of individualism and self-reliance (Henry and Sears, 2002).

Psychological characteristics of MTurk respondents

We know that personality affects self-selection into online panels (Brüggen and Dholakia, 2010; Dollinger and Leong, 1993; Rogelberg et al., 2003), however, we do not have evidence of any selection differences between MTurk and other panels. Though we are not aware of any study analyzing the determinants of participation on MTurk, it has been shown that MTurk samples tend to differ from other samples in personality and other dispositions. Compared to an adult community sample, MTurk workers have lower self-esteem, and lower levels of Extraversion, Emotional Stability, and Openness (Goodman et al., 2013). MTurk subjects also tend to score higher in Need for Cognition and Need to Evaluate than national samples (Berinsky et al., 2012). In short, there is some evidence that MTurk samples differ from other populations on personality traits and other psychological dispositions.

We have little evidence, however, regarding how the self-selection process might affect political divides on MTurk. Scherer et al. (2014) find the same partisan differences in System Justification in an MTurk sample and a nationally representative sample. Grimmer et al. (2012) find that the associations between partisanship, ideology, and feelings towards Barack Obama are substantively identical across MTurk and nationally representative samples. However, no work has systematically investigated whether liberals and conservatives on MTurk share the same personality traits and values as their counterparts in the mass public. In other words, do the same values and personality traits that motivate ideological differences in other types of samples also divide liberal and conservative subjects on MTurk?

Data and methods

We follow the logic of parallel studies (Clifford and Jerit, 2014; Hainmueller et al., 2015; Jerit et al., 2013) and compare a large sample recruited from MTurk to two national benchmarks. Our benchmark data comes from the ANES 2012 Time Series Study conducted before and after the 2012 US presidential election, which recruited 1413 respondents for face-to-face interviews (FTF) and 3860 respondents for a web-based survey (Web). The FTF sample was collected using computer-assisted self-interviewing and relied on an address-based sampling frame. The Web sample was recruited from GfK Knowledge Networks' address-based sampling frame. Our MTurk survey was posted online in June 2015 and described as a "Personality and Values Survey." The Human Intelligence Task's (HIT)

description read: “Answer short survey. Should take 8 to 10 minutes. This project has been approved by the University of Houston Committee for the Protection of Human Subjects.” The keywords tied to the HIT were survey, demographics, politics, personality, psychology, and values. It was made available to 1500 US residents with an acceptance rate of at least 95%. Subjects were paid US\$0.40 for completion.² Demographics are shown in Table A1 in the online Appendix.

Measures

All of our measures are drawn directly from the ANES, so we discuss them only briefly here (see online Appendix for full question wording).³ Political ideology and partisanship are both measured using standard seven-point scales. Social and economic ideology are each measured as indices of policy attitudes following the approach of Feldman and Johnston (2014). All dependent variables are coded such that higher values are more conservative.

We measure the Big Five personality traits using the Ten-Item Personality Inventory (TIPI; Gosling et al., 2003). Moral traditionalism, racial resentment, and egalitarianism are measured as the average agreement with 4–6 statements. Authoritarianism is measured using the child traits battery (Feldman and Stenner, 1997). Our demographic controls include age, education, gender, income, and religiosity (following Feldman and Johnston, 2014).

Results

We begin by graphically displaying the levels of each independent variable across each measure of ideology. Results for partisanship were substantively identical and are shown in the online Appendix (Figures A1–A3). Figure 1 displays the average level of each personality trait across self-placement ideology for each sample (ANES FTF, ANES Web, and MTurk). The means and sample differences are shown in Table A2 in the online Appendix.

Starting with the top-left panel, Extraversion is uncorrelated with ideology across all samples. However, ANES FTF respondents are significantly more extraverted than both ANES Web and MTurk samples, and MTurk subjects are less extraverted than the FTF sample (Cohen’s $d = .59$) and the Web sample ($d = .30$). Cohen’s d is the difference between the means of two groups divided by their pooled standard deviation, which provides a measure of effect size relative to the variation in the dependent variable (Cohen, 1988). Turning to the top-right panel, Agreeableness is not significantly related to political ideology in any of the samples, nor are there large differences in means between the samples. Conscientiousness is positively correlated with conservatism across all three samples (all $ps < .01$). Emotional Stability is positively correlated with ideology in all samples, but this relationship is statistically

significant only in the MTurk sample ($p < .01$). Finally, Openness is negatively correlated with ideology in all samples (all $ps < .01$). The MTurk sample scores notably higher in Openness than the Web sample ($d = .39$), but similar to the FTF sample ($d = .09$). Overall, correlations between the Big Five traits and political ideology are substantively identical across samples with the exception of Emotional Stability. The MTurk sample is also lower in Extraversion than both ANES samples.

Figure 2 displays the mean level of each value orientation across self-reported ideology. Authoritarianism, racial resentment, and moral traditionalism are positively associated with conservatism across all three samples (all $ps < .01$). While conservatives are nearly indistinguishable across samples, MTurk liberals score notably lower in each value than ANES liberals. These differences may be due, in part, to lower levels of religiosity (see Table A3 in the online Appendix). Egalitarianism is negatively correlated with ideology across all samples ($ps < .01$), though MTurk liberals score slightly higher than ANES liberals. Overall, we find the same relationships between values and political ideology, and conservatives are nearly indistinguishable across samples. However, we find a stronger relationship between values and political ideology in the MTurk sample, with liberals taking more characteristically liberal positions. The stronger correlations may reflect the fact that MTurk respondents tend to be more politically knowledgeable than the nationally representative samples (Berinsky et al., 2012), though we cannot directly test this claim.

Figure 3 displays social and economic ideology as a function of self-reported ideology. Ideological subgroups are highly similar across samples, though MTurk liberals have more liberal social preferences and slightly more liberal economic preferences, consistent with the results above.

Psychological models of political ideology

We now move away from descriptive statistics and ask whether researchers investigating the psychological antecedents of political ideology would reach the same conclusions if they relied on an MTurk sample rather than an ANES sample. Below we present six Ordinary Least Squares (OLS) models, each predicting one of three dependent variables (self-reported ideology, social ideology, and economic ideology), using one of the two sets of independent variables (personality traits or values) and demographic controls. For similar approaches, see Gerber et al. (2010), Mondak and Halperin (2008), and Feldman and Johnston (2014). For ease of comparison, we plot the coefficients from the models and exclude demographic variables (see Tables A4–7 in the online Appendix for full model results).

The left column of Figure 4 displays the coefficients for personality traits predicting self-reported ideology. Starting at the top, Openness is a significant predictor of liberal

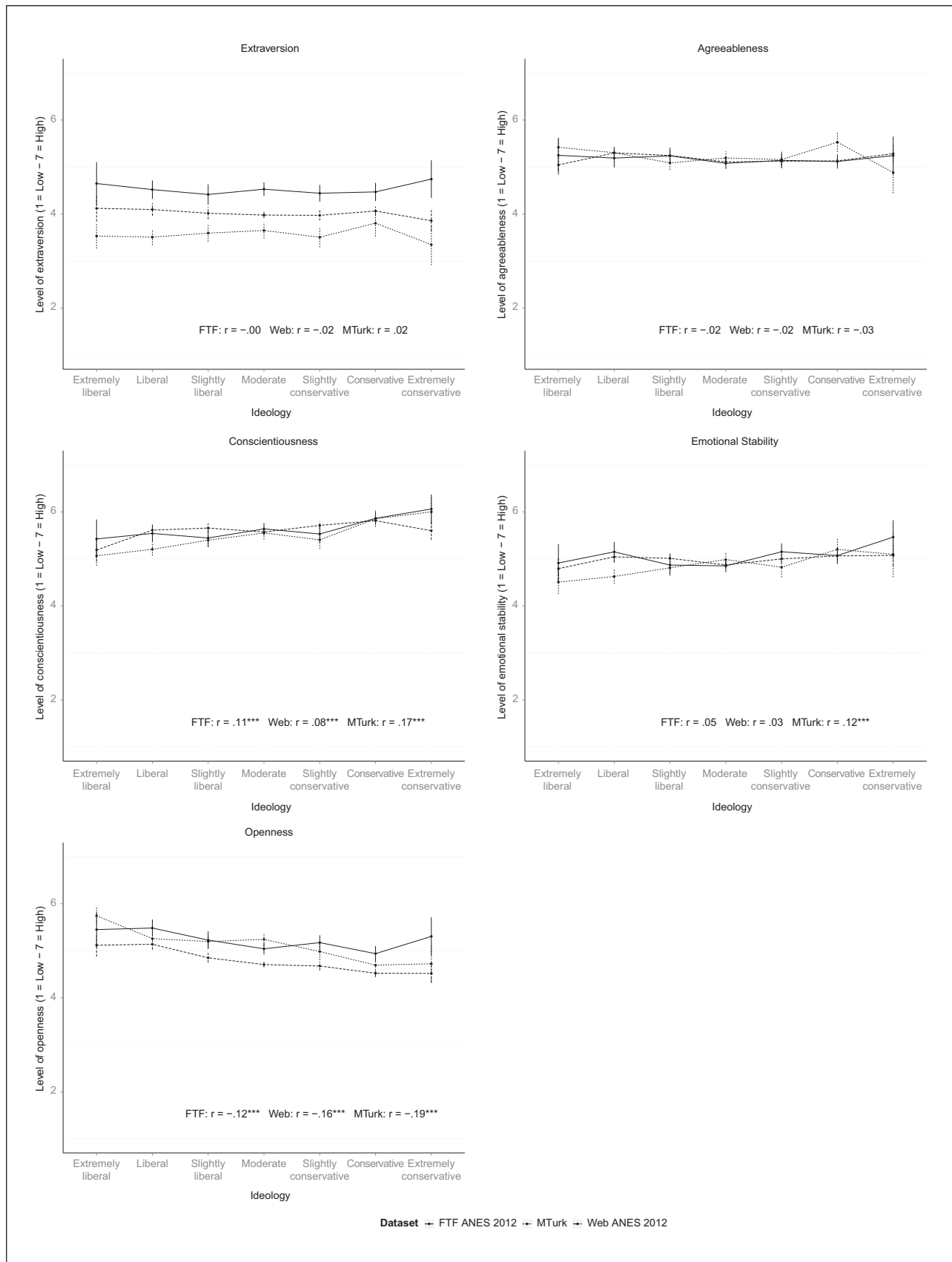


Figure 1. Sample differences in personality traits by ideology.

* $p < .05$; ** $p < .01$; *** $p < .001$.

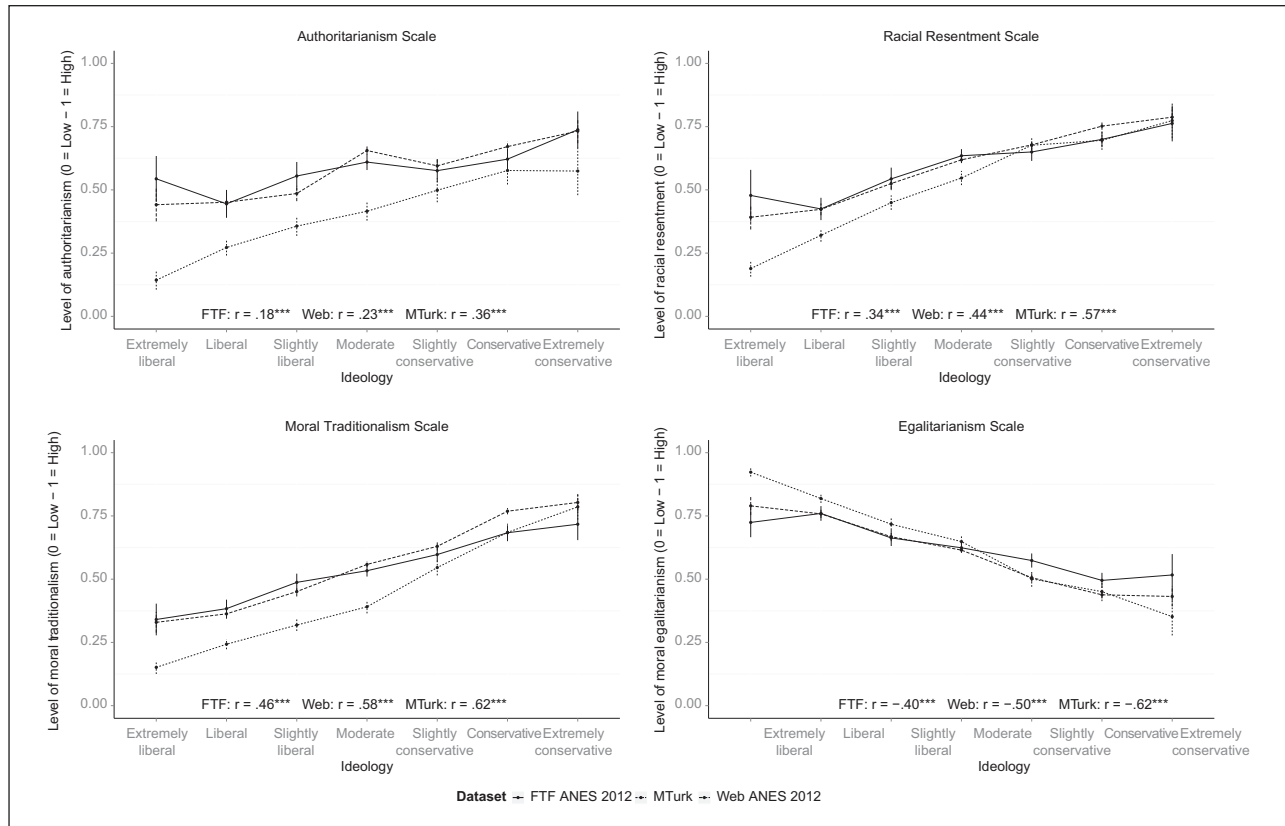


Figure 2. Sample differences in values by ideology.

* $p < .05$; ** $p < .01$; *** $p < .001$.

ideology across all samples ($ps < .001$). The coefficients on Emotional Stability are positive in all samples and similar in magnitude, but statistically significant only in the MTurk sample ($p < .05$). Conscientiousness is a strong predictor of conservatism across all samples ($ps < .001$). Agreeableness is a modest predictor of liberalism in only the Web and MTurk samples ($ps < .001$). Finally, Extraversion does not significantly predict ideology in any of the samples. Overall, the results are nearly identical across samples.

The middle-left panel of Figure 4 displays the results of a similar analysis predicting economic ideology. Openness significantly predicts liberalism in all samples ($ps < .01$). Emotional Stability is positive across all samples but only statistically significant in the Web sample ($p < .001$). Conscientiousness predicts conservatism in all samples ($ps < .05$). Agreeableness predicts liberalism in all samples ($ps < .001$), though this effect falls just short of statistical significance in the Web sample. Extraversion is positive in all samples, but only statistically significant in the Web sample. Overall, the results are again highly similar across samples, though the Web sample showed two minor deviations from the other samples.

The bottom-left panel of Figure 4 shows the results for social ideology. Openness significantly predicts liberalism for all samples ($ps < .001$). Emotional Stability has a null

effect in all samples. Conscientiousness is positive for all samples, but is statistically significant only in the ANES samples ($ps < .01$). Agreeableness significantly predicts liberalism in the Web and MTurk samples ($ps < .05$), but is null in the FTF sample. Finally, Extraversion is null for all samples. Overall, there was some variation in the results across samples, though the ANES samples seemed to disagree with each other about as often as with the MTurk sample.

We now conduct the same three analyses using values as our independent variables. The top-right panel of Figure 4 displays the results for self-reported ideology. Authoritarianism does not significantly predict ideology in any of the samples. Racial resentment and moral traditionalism both significantly predict conservatism across all samples ($ps < .001$). Lastly, egalitarianism is a significant predictor of liberalism across all samples ($ps < .001$). Overall, the results are substantively identical across samples.

The middle-right panel of Figure 4 displays the results predicting economic ideology. Authoritarianism predicts more liberal economic ideology across all samples ($ps < .05$). Racial resentment and moral traditionalism both predict conservatism across all samples ($ps < .001$). Lastly, egalitarianism is a strong predictor of liberalism for all samples ($ps < .001$). Again, we find no substantive differences between samples.

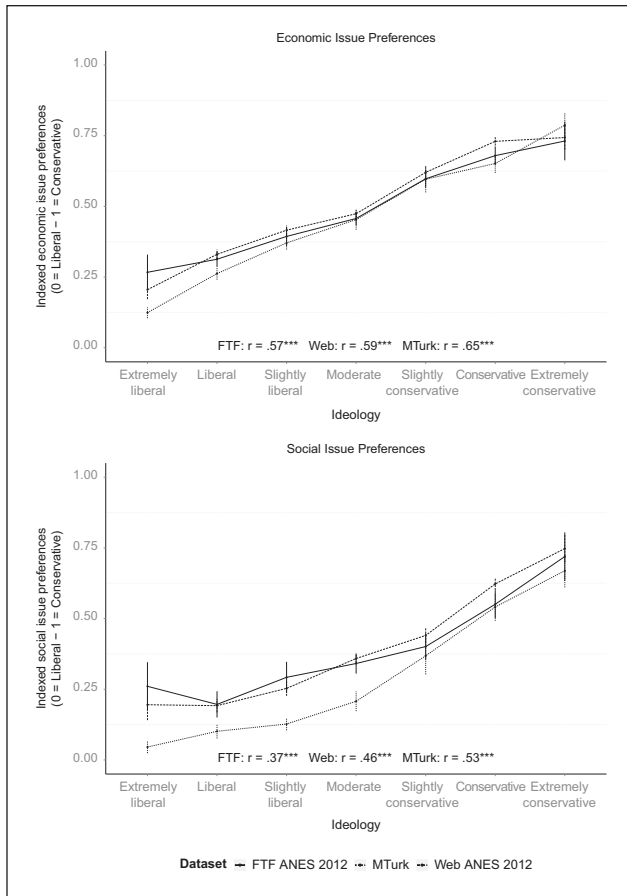


Figure 3. Sample differences in economic and social issue preferences by ideology.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Finally, the bottom-right panel of Figure 4 shows the results predicting social ideology. Authoritarianism significantly predicts conservatism in both ANES samples ($ps < .001$), but is null in the MTurk sample. Racial resentment is positive in all samples, but only statistically significant in the Web sample ($p < .05$). Moral traditionalism strongly predicts social conservatism for all samples and egalitarianism predicts liberalism across all samples ($ps < .001$). Overall, the results are highly similar, with two exceptions, in which the Web sample and MTurk sample disagreed with the other samples.

Do coefficient estimates significantly differ across samples?

While the results above are highly similar across samples, there were some apparent differences in effect sizes. To test whether these effects are significantly different from each other, we estimated a series of models pooling two of the three samples together at a time (e.g., Web vs. MTurk). Each model includes a dummy variable indicating the sample (e.g., Web = 0, MTurk = 1) and interactions between the

dummy variable and each of our independent variables (excluding demographic variables). The interaction terms in each model indicate whether the coefficients are significantly different in magnitude across samples. We repeated this approach for each of our four dependent variables (self-reported ideology, partisanship, economic ideology, and social ideology). Full model results are shown in Tables A8–14. For each of the three sample comparisons, this amounts to 36 tests (for a total of 108), raising concerns about multiple comparisons. We address this concern by controlling the false discovery rate (FDR) (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001), which we set to .05 (e.g., Battaglini et al., 2007).

Beginning with the comparison between the ANES FTF and Web samples as a benchmark, we find one significant difference in effect size out of 36 tests. However, after controlling for the FDR, we find no significant differences. Comparing MTurk to the FTF sample, we find five significant differences, though only one remains after controlling the FDR (authoritarianism predicting social ideology). Comparing MTurk to the Web, we find six significant differences, but after controlling the FDR, only three out of 36 remain. These three differences are in moral traditionalism predicting economic ideology, racial resentment predicting partisanship, and authoritarianism predicting social ideology. The only apparent pattern in these few differences is that all three independent variables are values endorsed more by conservatives. Overall, however, our MTurk sample generated results that are substantively nearly identical to the ANES and effect sizes that are indistinguishable in the vast majority of cases.

Conclusion

While much research has documented the characteristics of MTurk workers, there has been little research on whether the selection process of joining MTurk results in subjects who are psychologically different from the general population in terms of the relationship between personality traits, values, and ideology. Our results suggest that MTurk conservatives share the same personality traits and values as conservatives drawn from high-quality national samples. Indeed, our conservatives were largely indistinguishable across samples. The differences that we did find were largely among the liberal subjects. MTurk liberals appear to hold more characteristically liberal values and political attitudes. As for personality traits, we found few differences across samples with the exception of Extraversion. Here, the differences were primarily in the average level of the variable (rather than its relationship with ideology). Whereas the FTF sample was the most extraverted, the MTurk sample was the least.

Following the approach of recent research in political psychology, we also examined the psychological predictors of political ideology. The results were substantively nearly identical across samples, indicating that a researcher would draw largely the same conclusions if they chose to sample

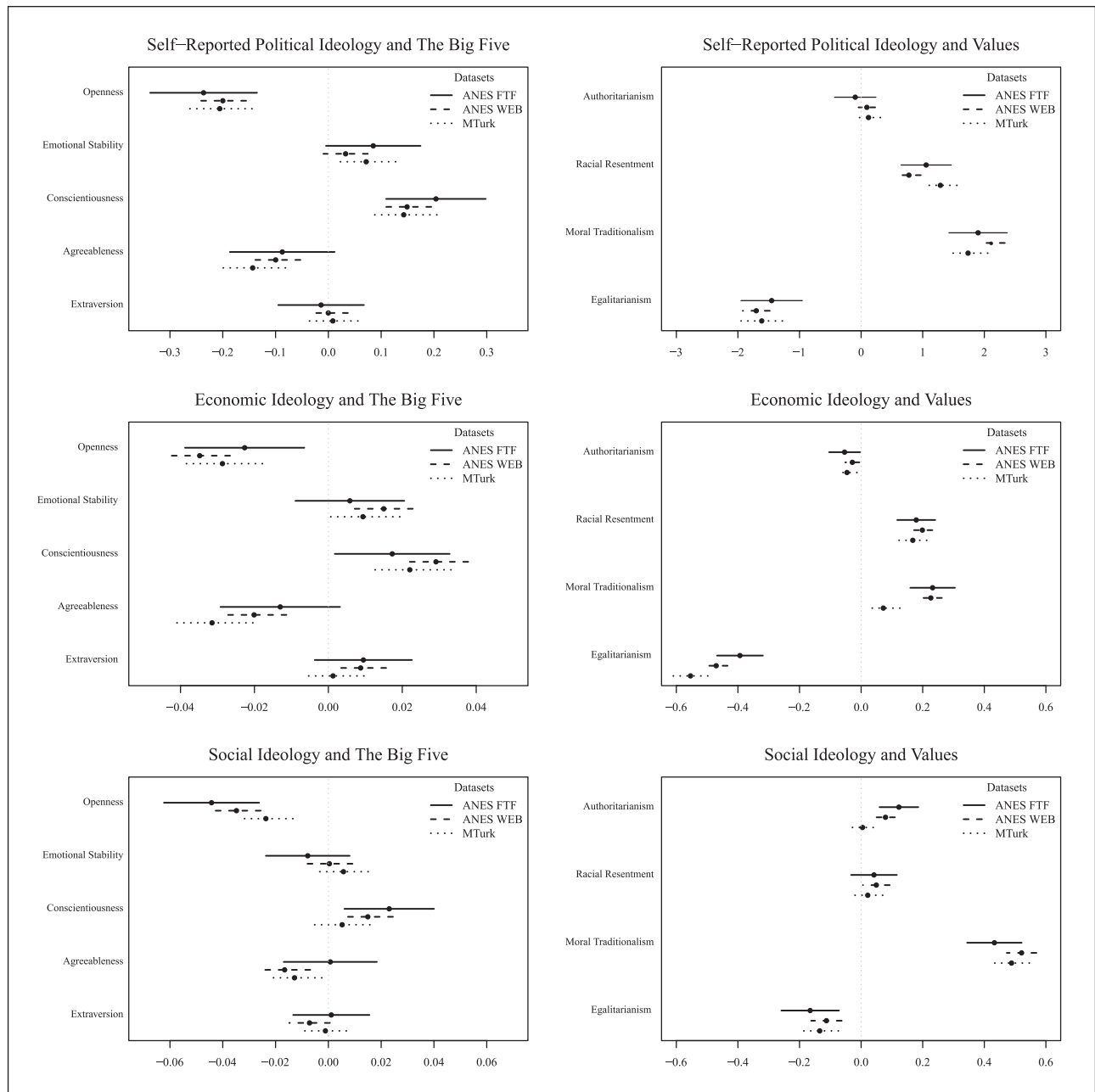


Figure 4. The effects of personality and values on political ideology.

from MTurk, rather than rely on the ANES. Indeed, MTurk effect sizes significantly differed from the ANES in only four out of 72 tests. Most of these cases involved the predictive power of conservative values, such as moral traditionalism. It is unclear why these variables might have weaker predictive power among the MTurk sample, but it might indicate shifting political divisions among a younger and less religious population.

Overall, our results suggest that the same values and personality traits that motivate ideological differences in the mass public also divide liberals and conservatives on

MTurk. Our study thus provides evidence for the validity of samples drawn from MTurk for psychological research on ideology. While conservatives in our sample closely matched nationally representative samples in personality and values, our liberals took on more characteristically liberal attitudes and dispositions. As a result, it may be that experimental treatments, such as persuasive frames, that are designed to target liberal values are more effective among MTurk liberals than among liberals drawn from a representative sample. Researchers should be aware of this possibility and should consider directly measuring the

dispositions their treatments are intended to target, rather than relying on ideology as a proxy variable.

Even despite the observed similarities between MTurk and other sample types, a number of important questions remain. MTurk subjects tend to be more politically knowledgeable than national samples (Berinsky et al., 2012) and as a result may be better sorted and more polarized. MTurk subjects also tend to be highly attentive, reducing concerns about satisficing. However, higher levels of attention may correspond with larger treatment effects or different information processing styles (e.g., Chong and Druckman, 2012), which may be due to the common use of attention checks (Hauser and Schwarz, 2015). The validity of MTurk as a recruitment tool will be an ongoing and topic-specific question.

Finally, we must note that our research does not provide blanket support for the use of MTurk samples for several reasons. Most obviously, our work focuses only on the relationship between political ideology, values, and personality. Our MTurk sample was also not a random sample of the MTurk workforce. As a result, we cannot be confident that our results apply to any sample drawn from MTurk. However, samples drawn from MTurk consistently diverge from the national population on several key characteristics, such as age, ideology, and religiosity. The persistence of these findings suggests that the demographics of MTurk are fairly stable (Berinsky et al., 2012; Buhrmester et al., 2011; Huff and Tingley, 2015; Krupnikov and Levine, 2014; Shapiro et al., 2013). Lastly, concerns remain about the effects of subject “savviness” or “non-naïveté” (Krupnikov and Levine, 2014; Chandler et al., 2014), cross-talk (Chandler et al., 2014), and other issues (for discussion, see Goodman et al., 2013; Shapiro et al., 2013). However, our research does suggest that MTurk workers largely share the psychological dispositions of their ideological counterparts in the mass public.

Declaration of conflicting interest

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) received no financial support for the research, authorship, and/or publication of this article.

Supplementary material

<https://journals.sagepub.com/doi/suppl/10.1177/2053168015622072>

Notes

1. For example, see recent discussion at the Cultural Cognition Project and Washington Post’s *Monkey Cage* (Kahan, 2013; Searles and Ryan, 2015).
2. This study was approved by the University of Houston Committee for the Protection of Human Subjects (Application #15560-EX – (6556)).

3. We edited wording of some questions to be appropriate for the survey mode.

References

- Ahler DJ (2014) Self-fulfilling misperceptions of public polarization. *The Journal of Politics* 76(03): 607–620.
- Arceneaux K (2012) Cognitive biases and the strength of political arguments. *American Journal of Political Science* 56(2): 271–285.
- Battaglini M, Morton RB and Palfrey T (2007) Efficiency, equity, and timing of voting mechanisms. *American Political Science Review* 101(03): 409–424.
- Benjamini Y and Hochberg Y (1995) Controlling the False discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 57(1): 289–300.
- Benjamini Y and Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* 29(4): 1165–1188.
- Berinsky AJ, Huber GA and Lenz GS (2012) Evaluating online labor markets for experimental research: Amazon.com’s Mechanical Turk. *Political Analysis* 20(3): 351–368.
- Brüggen E and Dholakia UM (2010) Determinants of Participation and response effort in web panel surveys. *Journal of Interactive Marketing* 24(3): 239–250.
- Buhrmester M, Kwang T and Gosling SD (2011) Amazon’s Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science* 6(1): 3–5.
- Bullock JG (2011) Elite influence on public opinion in an informed electorate. *American Political Science Review* 105(03): 496–515.
- Chandler J, Mueller P and Paolacci G (2014) Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods* 46(1): 112–130.
- Chong D and Druckman JN (2012) Counterframing effects. *The Journal of Politics* 75(1): 1–16.
- Clifford S (2014) Linking Issue stances and trait inferences: A theory of moral exemplification. *The Journal of Politics* 76(3): 698–710.
- Clifford S and Jerit J (2014) Is There a cost to convenience? An experimental comparison of data quality in laboratory and online studies. *Journal of Experimental Political Science* 1(2): 1–12.
- Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences*. New Jersey: Lawrence Erlbaum Associates.
- Costa PT and McCrae RR (1992) Four ways five factors are basic. *Personality and Individual Differences* 13(6): 653–665.
- Crump MJC, McDonnell JV and Gureckis TM (2013) Evaluating Amazon’s Mechanical Turk as a tool for experimental behavioral research. *PloS one* 8(3): e57410.
- Dollinger SJ and Leong FTL (1993) Volunteer bias and the five-factor model. *The Journal of Psychology* 127(1): 29–36.
- Druckman JN, Peterson E and Slothuus R (2013) How elite partisan polarization affects public opinion formation. *American Political Science Review* 107(1): 57–79.
- Feldman S and Johnston C (2014) Understanding the determinants of political ideology: Implications of structural complexity. *Political Psychology* 35(3): 337–358.
- Feldman S and Stenner K (1997) Perceived Threat and authoritarianism. *Political Psychology* 18(4): 741–770.

- Gerber AS, et al. (2013) Big Five personality traits and responses to persuasive appeals: Results from voter turnout experiments. *Political Behavior* 35(4): 687–728.
- Gerber AS, et al. (2010) Personality and political attitudes: Relationships across issue domains and political contexts. *American Political Science Review* 104(1): 111–133.
- Goodman JK, Cryder CE and Cheema A (2013) Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making* 26(3): 213–224.
- Gosling SD, Rentfrow PJ and Swann WB (2003) A very brief measure of the Big-Five personality domains. *Journal of Research in Personality* 37(6): 504–528.
- 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* 106(4): 703–719.
- Hainmueller J, Hangartner D and Yamamoto T (2015) Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences of the United States of America* 112(8): 2395–2400.
- Hauser DJ and Schwarz N (forthcoming) Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior research methods*.
- Hauser DJ and Schwarz N (2015) It's a trap! Instructional Manipulation checks prompt systematic thinking on "tricky" tasks. *SAGE Open* 5(2): 1–6.
- Henry PJ and Sears DO (2002) The Symbolic Racism 2000 Scale. *Political Psychology* 23(2): 253–283.
- Horton JJ, Rand DG and Zeckhauser RJ (2011) The online laboratory: Conducting experiments in a real labor market. *Experimental Economics* 14(3): 399–425.
- Huber GA and 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(1): 385–397.
- Huff C and Tingley D (2015) "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics* 2(3): 1–12.
- Jerit J, Barabas J and Clifford S (2013) Comparing contemporaneous laboratory and field experiments on media effects. *Public Opinion Quarterly* 77(1): 256–282.
- Johnston CD, Newman BJ and Velez Y (2015) Ethnic change, personality, and polarization over immigration in the American public. *Public Opinion Quarterly* 79(3): 662–686.
- Kahan DM (2013) What's a "valid" sample? Problems with Mechanical Turk study samples, part 1. *Cultural Cognition Project Blog*. Available at: <http://www.culturalcognition.net/blog/2013/7/8/whats-a-valid-sample-problems-with-mechanical-turk-study-sam.html> (accessed 3 May 2015).
- Krupnikov Y and Levine AS (2014) Cross-sample comparisons and external validity. *Journal of Experimental Political Science* 1(1): 59–80.
- McDermott R (2002) Experimental methodology in political science. *Political Analysis* 10(4): 325–342.
- Mondak JJ (2010) *Personality and the Foundations of Political Behavior*. New York: Cambridge University Press.
- Mondak JJ and Halperin KD (2008) A framework for the study of personality and political behaviour. *British Journal of Political Science* 38(2): 335–362.
- Nyhan B and Reifler J (2010) When corrections fail: The persistence of political misperceptions. *Political Behavior* 32(2): 303–330.
- Paolacci G, Chandler J and Ipeirotis PG (2010) Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5(5): 411–419.
- Rogelberg SG, et al. (2003) Profiling active and passive nonrespondents to an organizational survey. *Journal of Applied Psychology* 88(6): 1104–1114.
- Scherer AM, Windschitl PD and Graham J (2014). An ideological house of mirrors political stereotypes as exaggerations of motivated social cognition differences. *Social Psychological and Personality Science* 6(2): 201–209.
- Searles K and Ryan JB (2015) Researchers are rushing to Amazon's Mechanical Turk. Should they? *Washington Post*. Available at: <http://www.washingtonpost.com/blogs/monkey-cage/wp/2015/05/04/researchers-are-rushing-to-amazons-mechanical-turk-should-they/> (accessed 4 May 2015).
- Shapiro DN, Chandler J and Mueller PA (2013) Using Mechanical Turk to study clinical populations. *Clinical Psychological Science* 1(2): 213–220.
- Weinberg JD, Freese J and McElhattan D (2014) Comparing data characteristics and results of an online factorial survey between a population-based and Crowdsourced-recruited sample. *Sociological Science* 1: 292–310.