

Do studies of political ideology that use Mechanical Turk samples have any validity?

*POLAMURI RAMA MOHAN REDDY 1, AAKITI PRAVEEN 2,
ASSOCIATE PROFESSOR 1, ASSISTANT PROFESSOR 2,
Mail ID:polamuriramohanreddy@gmail.com, Mail ID:praveenakiti@gmail.com
Dept.: Mechanical
Nagole Institute of Technology and Science,
Kuntloor(V),Hayathnagar(M),Hyderabad,R.R.Dist.-501505.*

Abstract

Researchers are increasingly turning to Amazon's Mechanical Turk (MTurk) to find participants for their studies. We know very little about whether liberals and conservatives recruited through MTurk have the same psychological dispositions as their counterparts in the mass public, despite the fact that great attention has been paid to the demographic distinctions between MTurk samples and the national population. Some have suggested that MTurk's selection mechanism renders irrelevant the subject pool for investigating important topics in political science, however this is not supported by data. This paper's goal is to assess the veracity of this claim by contrasting a large MTurk sample with two nationally representative samples (one done online and the other face-to-face). We analyze the three samples to see if there are any consistent personality or value-based explanations for political ideology. The findings from the three samples are almost comparable, with just slight differences in the effect sizes. For the most part, the ideological and political differences between our MTurk sample's liberals and conservatives resemble those seen in the general population, but MTurk liberals seem to have more stereotypically liberal beliefs and attitudes than liberals in representative samples. Overall, our findings support the idea that MTurk may be used as a reliable recruiting method for studies of political ideology's psychological effects.

Keywords

Sample comparison, external validity, and Mechanical Turk.

Introduction

Amazon's MTurk (Mechanical Turk) is being used more and more often by academics to find participants in surveys of the general population (e.g., Ahler, 2014; Arceneaux, 2012; Clifford, 2014; Grimmer et al., 2012; Huber and Paris, 2013; Johnston et al., 2015). Using MTurk, you can quickly recruit a wide range of participants at a fraction of the cost of paid online panels (Berinsky et al., 2012). There are many different ways that researchers have tried to verify MTurk's reliability as a sample recruitment method. Framing effects (Berinsky et al., 2012; Weinberg et al., 2014), decisionmaking biases (Goodman et al., 2013; Paolacci et al., 2010), economic games (Horton et al., 2011), and cognitive psychology tasks are only some of the areas where experimental findings have been repeated on MTurk (Crump et al., 2013).

According to the findings of others, MTurk data conform to industry-standard psychometric criteria (Buhrmester et al., 2011; Shapiro et al., 2013). Respondents using MTurk are just as attentive, if not more so, than those using other methods (Hauser and Schwarz, forthcoming; Paolacci et al., 2010; Weinberg et al., 2014).

Still, MTurk's usage raises a number of questions (e.g., Chandler et al., 2014; Krupnikov and Levine, 2014). Much of the argument over whether or not MTurk can be used as a reliable recruiting tool has followed the field's "near preoccupation" with the importance of a sample's external validity (McDermott, 2002: 334). There is a persistent pattern of findings suggesting MTurk samples differ from the general U.S. population in being more politically liberal, younger, less religious, and less ethnically diverse (Berinsky et al. 2012; Huff and Tingley, 2015).

Distinctive Liberal and Conservative Mindsets

There is a wide range of psychological differences between liberals and conservatives, but the most prominent are differences in personality and values. Over the course of many decades of study, the Big Five personality qualities—Openness to experience, conscientiousness, extroversion, agreeableness, and emotional stability—have emerged as a valuable framework for analyzing the consistent features that underpin people's actions (Costa and McCrae, 1992). Ideology can be reliably predicted by two of the Big Five qualities in particular. Scores on measures of Openness are often higher among liberals, who show a stronger propensity to seek out and consider novel perspectives (Gerber et al., 2013). When it comes to reliability and self-control, conservatives tend to have higher Conscientiousness scores (Gerber et al., 2013).

Less conclusive results have been found when looking at associations between other qualities and

political leanings; nonetheless, there is some evidence that Extraversion is connected with conservatism (e.g., Gerber et al., 2010; Mondak and Halperin, 2008) and that Agreeableness predicts liberalism (Mondak, 2010). The ideals of liberals and conservatives are even more diametrically opposed to one another. We zero in on egalitarianism, moral traditionalism, authoritarianism, and racial resentment—four values often used in studies of political attitudes and identities. Respect for authority, loyalty to one's group, and loyalty to one's family are the first three values. The fourth, racial resentment, is a mixture of racist bigotry with the American principles of independence and self-sufficiency (Henry and Sears, 2002).

Personality traits of Amazon Mechanical Turk participants

Personality does seem to have a role in self-selection into online panels (Brüggen and Dholakia, 2010; Dollinger and Leong, 1993; Rogelberg et al., 2003), but there is no evidence to suggest that MTurk is any different from other panels in terms of selection. No known research has examined what motivates people to work on MTurk, however there is evidence to suggest that MTurk samples are distinct from other samples in terms of personality and other inclinations. MTurk employees exhibit lower levels of Extraversion, Emotional Stability, and Openness, as well as poorer self-esteem, as compared to an adult population sample (Goodman et al., 2013). Need for Cognition and Need to Evaluate are likewise more strongly elevated in MTurk individuals than in national samples (Berinsky et al., 2012). To sum up, there is some data suggesting that MTurk samples vary from other groups with respect to personality characteristics and other psychological dispositions.

However, there is little data on whether or whether MTurk's self-selection mechanism contributes to partisan gaps. When comparing an MTurk sample to a nationally representative sample, Scherer et al. (2014) found no significant variations in partisanship regarding System Justification. According to Grimmer et al. (2012), there is little to no difference between MTurk and nationally representative samples when it comes to the correlations between partisanship, ideology, and opinions of Barack Obama. However, no studies has systematically looked into whether or not

liberals and conservatives on MTurk have the same personality characteristics and values as their counterparts in the general population. Specifically, do liberal and conservative individuals on MTurk share the same beliefs and personality factors that motivate ideological disparities in other sample types?

Research Strategies and Information Collected

Similar to previous research (Clifford and Jerit, 2014; Hainmueller et al., 2015; Jerit et al., 2013), we compare a large MTurk-sourced sample to two national norms. Our standard data comes from the ANES 2012 Time Series Study, which was done before and after the 2012 US presidential election and included interviews with a total of 3860 people (1413 FTF) (Web). The FTF sample used an address-based sampling frame and was gathered using computer-assisted self-interviewing. To compile their online sample, researchers at GfK Knowledge Networks used their own address-based sampling frame. In June of 2015, we released an MTurk survey titled "Personality and Values Survey." When it comes to the Human Intelligence Test (HIT).

Brief Survey - Please Respond" Time estimate: 8-10 min. The Committee for the Protection of Human Subjects at the University of Houston has given its support to this research. Survey, demography, politics, personality, psychology, and values were some of the HIT's associated keywords. At least 95% of the 1500 US citizens who were offered it said they would take it. Test subjects received \$0.40 USD for their time and effort. Table A1 in the downloadable Appendix provides demographic data.

Measures Because we rely only on the ANES for all of our metrics, we will simply quickly go through them here (see online Appendix for full question wording). 3 There are generally accepted seven-point measures for assessing political ideology and partisanship. Feldman and Johnston's method is used to quantify social and economic ideology as indicators of policy opinions (2014). All independent variables have been encoded so that more cautious responses result in higher values. The ten-item Personality Inventory is used to assess the Big Five characteristics (TIPI; Gosling et al., 2003). Consensus on a set of statements designed to gauge moral conservatism, anti-racism

sentiment, and egalitarianism. The Child Traits Battery is used to assess authoritarianism (Feldman and Stenner, 1997). Age, education, gender, wealth, and religiosity are some of the demographic variables we use as controls (after Feldman and Johnston, 2014).

Results

To begin, we create bar charts showing how each independent variable correlates with each ideological scale. There was little to no discernible difference in the partisanship results, which are shown in the accompanying Appendix (Figures A1–A3). Average levels of character traits by ideology of self-placement are shown in Figure 1. (ANES FTF, ANES Web, and MTurk). The online Appendix includes Table A2, which displays the means and sample differences. The top left panel shows that extraversion has a consistent correlation with ideology across all datasets. In contrast, the ANES FTF sample is considerably more extroverted than the ANES Web and MTurk samples, whereas MTurk individuals are less extroverted than the FTF sample (Cohen's $d = .59$) and the Web sample ($d = .30$). Cohen's d is a measure of effect magnitude relative to the variance in the dependent variable and is calculated by dividing the difference in averages between two groups by their combined standard deviation (Cohen, 1988). As can be seen in the panel on the top right, there is no statistically significant correlation between agreeableness and political ideology across any of the samples, nor are there any major variations in mean agreeableness scores between the samples. All three samples show a positive relationship between conscientiousness and conservatism (PS .01). Although a positive correlation exists between emotional stability and ideology across all samples, it is only statistically significant in the MTurk sample ($p .01$). Last but not least, there is a negative correlation between Openness and ideology across all samples ($p .01$). When compared to the Web sample ($d = .39$) and the Free and Open-Source Software (FTF) sample ($d = .09$), the MTurk sample scores much better in Openness.

Other than Emotional Stability, connections between the Big Five and political ideology are statistically consistent across samples. When compared to both ANES groups, the MTurk sample likewise scores lower in Extraversion. The median values of each ideology are shown in Figure 2. In

all three samples, conservatism is positively associated with authoritarianism, racial animosity, and moral traditionalism (all $ps .01$). While there is little to no difference in the groups when it comes to conservatism, liberals on MTurk do much worse than their ANES counterparts across the board. One possible explanation for these variations is less widespread religious observance (see Table A3 in the online Appendix). Although MTurk liberals had a somewhat higher score than ANES liberals, there is a negative correlation between egalitarianism and ideology in all samples ($ps .01$). In sum, we find consistent associations between values and political ideology, and conservatives are essentially identical across samples. In the MTurk sample, however, we discover a more robust correlation between values and political ideology, with liberals more consistently adopting stereotypically liberal positions. Although we cannot directly evaluate the claim, the larger correlations may be attributable to the fact that MTurk respondents tend to be more politically knowledgeable than the nationally representative samples (Belinsky et al., 2012). Social and economic ideology as stated by the individual is shown in Figure 3. Consistent with these findings, there is a significant degree of homogeneity across ideological groupings across samples; nonetheless, liberals on MTurk tend to have more liberal social preferences and somewhat more liberal economic preferences.

Ideology and politics as a function of personality types

Now that we've gotten past the descriptive statistics, we want to know if studies examining the psychological precursors of political ideology that used MTurk samples instead of ANES samples came to the same conclusions. We offer six Ordinary Least Squares (OLS) models below, each of which predicts one of three dependent variables (self-reported ideology, social ideology, and economic ideology) using one of two sets of independent variables (personality characteristics or values, and demographic controls). See Gerber et al. (2010), Mondak and Halperin (2008), and Feldman and Johnston (2009) for examples of related methodologies (2014). We exclude demographic factors from our coefficient plots for simplicity of comparison (see Tables A4–7 in the online Appendix for full model results). The coefficients for personality factors that influence self-reported ideology are shown in the left column

of Figure 4. First and foremost, openness is a strong indicator of liberal

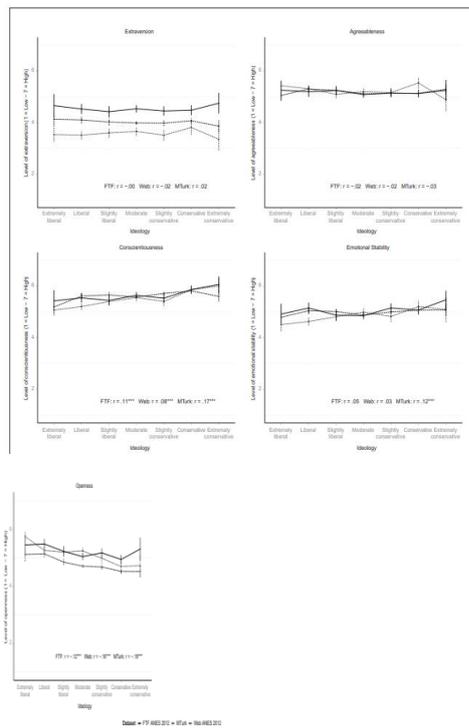


Figure 1. Sample differences in personality traits by ideology. * $p < .05$; ** $p < .01$; *** $p < .001$

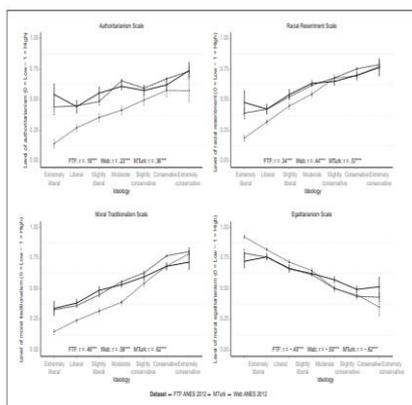
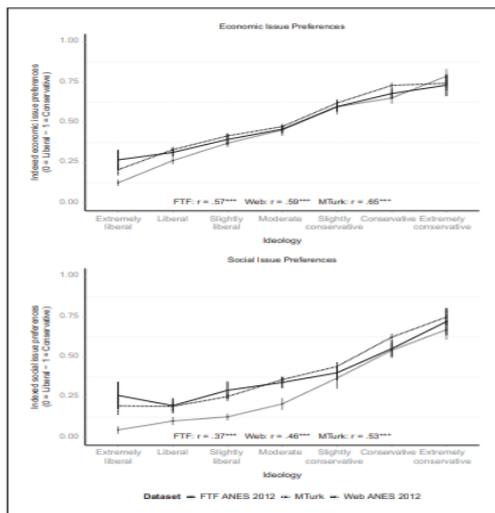


Figure 2. Sample differences in values by ideology. * $p < .05$; ** $p < .01$; *** $p < .001$

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



As seen in Figure 3, there is a wide range of opinions on economic and social issues among the sample population. * $p < .05$; ** $p < .01$; *** $p < .001$.

Finally, the findings in predicting social ideology are shown in the bottom right panel of Figure 4. Both ANES samples show a significant ($p .001$) relationship between authoritarianism and conservatism, but the Murk sample shows no such relationship. All samples show a positive reading for racial animosity, but the Web sample is the only one with statistical significance ($p .05$). All samples show a significant correlation between moral traditionalism and social conservatism, and between egalitarianism and liberalism ($p .001$). The findings are quite consistent with one another with just two notable outliers: the Web sample and the Murk sample.

If so, do sample-to-sample variations in estimated coefficients constitute a significant phenomenon?

While there is a lot of consistency in the findings across samples, the impact sizes vary somewhat. We built a series of models pooling two of the three data at a time to test for statistical significance between these effects (e.g., Web vs. MTurk). There are interactions between the dummy variable and each of our independent variables, and the models also include a dummy variable indicating the same (e.g., Web = 0, MTurk = 1). (Excluding demographic variables). Interaction terms in each model reveal whether or not coefficients vary significantly between samples. We followed this method for each of our four dependent variables

(self-reported ideology, partisanship, economic ideology, and social ideology). The whole model output is shown in Tables A8–A14. This equates to 36 tests (for a total of 108), with the three sample comparisons combined. To deal with this issue, we regulate the FDR (Benjamin and Hochberg, 1995; Benjamin and Yakutia, 2001), which we determine to be .05 (Benjamin and Hochberg, 1995; Benjamin and Yakutia, 2001). (e.g., Battalion et al., 2007). Using the difference in effect size between the ANES FTF and Web samples as a starting point, we discover that just one of the 36 tests is statistically significant. However, when we account for the FDR, we find no discernible variations. We identify five statistically significant differences between MTurk and the FTF sample; however, after correcting for FDR, only one of these differences survives (authoritarianism predicting social ideology). We detect six significant differences when comparing MTurk to the Web; however, after controlling for the FDR, only three of the original 36 differences remain. These three distinctions are the extent to which one's moral traditionalism foretells one's economic ideology, one's racial resentment foretells one's partisanship, and one's authoritarianism foretells one's social ideology. The only common thread among these minor deviations is that conservatives tend to place more emphasis on the three independent factors. As a whole, however, our MTurk sample produced findings that are almost indistinguishable from the ANES, including effect sizes.

Conclusion

Much study has been done on the demographics of MTurk employees, but less on whether or whether the selection procedure for joining MTurk results in individuals that are psychologically distinct from the general population in terms of the link between personality traits, values, and ideology. Our findings indicate that conservatives on MTurk have similar beliefs and character characteristics to those seen in large-scale national samples of conservatives. Indeed, our conservatives showed

little to no variation across batches. Where we did discover disparities, it was mostly among those who identified as liberal. It seems that MTurk liberals are more typical in their ideals and political outlooks. While we did find some variation in Extraversion across groups, other personality characteristics were quite consistent. Average values of the variable were where the biggest gaps appeared here (rather than its relationship with ideology). As expected, the FTF sample showed the highest levels of extroversion, while the MTurk group showed the lowest. We followed the lead of previous studies in political psychology and looked at the psychological determinants of political ideology. Substantially, the findings were consistent across samples, suggesting that a researcher might reach the same conclusions regardless of which samples were used.

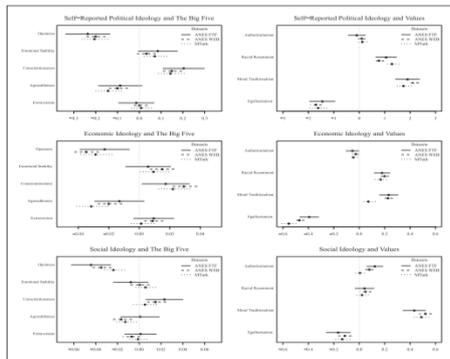


Figure 4: The influence of character traits on one's political stance.

rather than relying on the ANES, from MTurk. In fact, only four out of 72 experiments had MTurk effect sizes that substantially deviated from the ANES. Conservative ideals, such as moral traditionalism, had a significant predictive role in the majority of these instances. Although the reasons for the MTurk sample's lower predictive value for these factors remain elusive, it may reflect growing political tensions among a younger, less religious demographic. Collectively, our findings imply that liberals and conservatives on MTurk are separated by the same values and personality factors that drive ideological divisions in the general population. In this way, our research demonstrates the reliability of MTurk samples for ideological research in the field of psychology. While the conservatives in our sample mirrored those of nationally representative samples, the liberals in our sample leaned more toward liberal norms and perspectives. Therefore, it's possible that

experimental treatments like persuasive frames that are meant to address liberal ideals are more successful among MTurk liberals than among liberals recruited from a representative population. Researchers need to be alert to this potential and go beyond using ideology as a proxy variable to measure the dispositions their therapies are meant to change.

References

[1] Ahler DJ (2014) *Self-fulfilling misperceptions of public polarization. The Journal of Politics* 76(03): 607–620.

[2] Arceneaux K (2012) *Cognitive biases and the strength of political arguments. American Journal of Political Science* 56(2): 271–285.

[3] Battalio M, Morton RB and Palfrey T (2007) *Efficiency, equity, and timing of voting mechanisms. American Political Science Review* 101(03): 409–424.

[4] Benjamin 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.

[5] 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.

[6] 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.

[7] Brügggen E and Dholakia UM (2010) *Determinants of Participation and response effort in web panel surveys. Journal of Interactive Marketing* 24(3): 239–250.

[8] 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.

[9] Bullock JG (2011) *Elite influence on public opinion in an informed electorate. American Political Science Review* 105(03): 496–515.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] Dollinger SJ and Leong FTL (1993) *Volunteer bias and the fivefactor model. The Journal of Psychology* 127(1): 29–36.

[16] Druckman JN, Peterson E and Slothuus R (2013) *How elite partisan polarization affects public opinion formation. American Political Science Review* 107(1): 57–79.

[17] Feldman S and Johnston C (2014) *Understanding the determinants of political ideology: Implications of structural complexity. Political Psychology* 35(3): 337–358.

[18] Feldman S and Stenner K (1997) *Perceived Threat and authoritarianism. Political Psychology* 18(4): 741–770.