

Does the Sun Revolve Around the Earth?

A Comparison between the General Public and On-line Survey Respondents in Basic Scientific Knowledge

Emily A. Cooper^a, Hany Farid^b,

^a*Department of Psychology, Stanford University*

^b*Department of Computer Science, Dartmouth College*

Abstract

We conducted an on-line survey using a set of factual science questions that are commonly administered to assess scientific literacy. We report that the on-line population performed substantially better on this standard assessment than the traditional survey population. For example, it has been widely reported that 1 in 4 Americans do not know that the Earth revolves around the Sun; amongst the on-line population, this ratio is reduced to 1 in 25. While new on-line platforms provide researchers with unprecedented ease of access to a large sample population for studying trends in public knowledge and attitudes, it is unclear how representative the on-line population sample is of the US population at large. We discuss the potential reasons for this discrepancy, and the implications for conducting research on-line.

1. Introduction

Amazon Mechanical Turk (AMT: www.mturk.com) is an online labor market where anonymous workers are paid to perform short tasks. AMT affords fast, easy, and inexpensive access to a sample population for studying the behavior, psychology, knowledge, or attitudes of a large target population (Buhrmester, Kwang, & Gosling, 2011; Horton, Rand, & Zeckhauser, 2011; Paolacci, Chandler, & Ipeirotis, 2010). There is some concern, however, that biases in the AMT population may compromise the generalizability of results acquired using this platform (Goodman, Cryder, & Cheema, 2013; Horton et al., 2011; Paolacci et al., 2010). For example, compared to the general US population, AMT workers in the US are younger, more educated,

less racially diverse, and more likely to be female (Ipeirotis, 2010; Mason & Suri, 2012; Ross, Irani, Silberman, Zaldivar, & Tomlinson, 2010; Simons & Chabris, 2012). By its nature, the AMT population is also self-selecting and has frequent access to the Internet.

This population profile has not proven problematic for replicating several basic behavioral and psychological findings using AMT workers as participants (Crump, McDonnell, & Gureckis, 2013; Goodman et al., 2013; Horton et al., 2011). It should be noted, however, that these types of experiments often rely on the convenience sample of university research participant pools, and tap behaviors that are thought not to vary widely amongst the population at large. For research that aims to measure trends in knowledge and attitudes across a diverse geographic population, idiosyncrasies in the AMT population may be more problematic. Traditional survey methods take great pains to generate an unbiased sample population, for example, by using random digit phone dialing, in-person interviews, and non-response bias correction methods (Moser & Kalton, 1971).

Recently, researchers have begun using AMT to survey specific knowledge in the US population. For example, one survey assessed medical knowledge about cancers (Carter, Difeo, Bogie, Zhang, & Sun, 2014), one assessed knowledge related to water conservation (Attari, 2014), and two assessed pop-science knowledge about human memory (Patihis, Ho, Tingen, Lilienfeld, & Loftus, 2014; Simons & Chabris, 2012). Here, we report the results of an AMT survey assessing general scientific literacy that reveals a substantial deviation from previous traditional surveys. Scientific literacy is defined as the ability of adults to read and understand scientific findings as reported by the media. This understanding is thought to form the basis of the public's ability to engage in discussions of public policy related to scientific topics such as pollution, energy production, and public health (Miller, 1983, 1998). Scientific literacy is traditionally measured via survey questions that test factual scientific knowledge and understanding of the scientific process. As measured by traditional national surveys, scientific literacy has been relatively stable since the 1980's and is surprisingly low – a small percentage of the public exhibit high scores for basic textbook scientific knowledge (Allum, Sturgis, Tabourazi, & Brunton-Smith, 2008; Miller, 1983, 1998).

We surveyed 1014 AMT workers using a subset of factual science questions administered in the 2012 General Social Survey (GSS). The GSS is a national survey that uses probability sampling of the US population and is conducted largely by in-person interview (“GSS - General Social Survey,”

2014; Smith, 1978). Responses to science-related questions in this survey are used as indicators of scientific literacy for the annual Science and Engineering Indicators (SEI) published by the National Science Foundation (“National Science Board - Science and Engineering Indicators 2014,” 2014) and have been widely reported in the media (Henderson, 2014; Neuman, 2014; O’Neill, 2014).

Our survey of AMT workers found surprisingly large deviations from the trends of the 2012 GSS. Across the board, AMT respondents exhibited much higher scientific knowledge than the GSS respondents, even after accounting for demographic differences between the two surveys. Looking between demographic groups, we also found much smaller gender and age differences, meaning the AMT population is overall more homogeneous in their scientific literacy. We discuss how the scientific literacy of the AMT population may have implications for AMT research applications.

2. Methods

2.1. *Scientific Literacy Survey*

Using a standard Mechanical Turk survey template, we published a HIT (Human Intelligence Task) titled “Thirteen Question Quiz” with the short description “Answer thirteen short multiple choice questions.” The HIT was limited to respondents in the United States over the age of 18, to those that have a HIT Approval Rate greater than or equal to 95%, and to those that have 50 or more previously approved HITs. Respondents were paid \$0.50 regardless of how they performed on the survey.

Shown in Table 1 are the thirteen questions asked of each AMT respondent. The questions numbered 1-7 relate to scientific literacy (Miller, 1983, 1998). The questions numbered 8-10 provide basic demographic information (gender, age, and education). Interlaced within these ten questions are three simple control questions, which are used to ensure that the respondent reads each question. We published a total of 1037 HITs each of which were completed. The total sample size was decided upon before publishing the HITs, and determined as a number large enough to warrant comparison with the GSS sample. Of the completed HITs, 23 (2.2%) were excluded because the respondent either failed to answer all of the questions, or incorrectly answered one or more of the simple control questions. In half of the HITs, Question 3 was accompanied by a simple illustration of each option (Earth around Sun and Sun around Earth), to ensure that any incorrect responses were not due

#	Statement/Question	Response Choices
1	The center of the Earth is very hot.	<i>True</i> False
-	One plus one is three.	True False
2	The continents on which we live have been moving their locations for millions of years and will continue to move in the future.	<i>True</i> False
3	Does the Earth go around the Sun, or does the Sun go around the Earth?	<i>Earth around Sun</i> Sun around Earth
-	Strawberries are red.	True False
4	All radioactivity is man-made.	True <i>False</i>
5	Electrons are smaller than atoms.	<i>True</i> False
-	There are five hours in a day.	True False
6	Lasers work by focusing sound waves.	True <i>False</i>
7	The universe began with a huge explosion.	<i>True</i> False
8	What is your gender?	Male Female
9	What is your age?	18-24 25-34 35-44 45-54 55-64 65 or older
10	What is your highest level of education?	Didn't finish high school Finished high school Some college Finished college Graduate/professional degree

Table 1: Survey Questions. The online AMT survey contained seven scientific questions from the GSS (1-7), three control questions (not numbered), and three demographic questions (8-10). The correct answers are shown in italics.

to confusion caused by the wording of the question, which was identical to the GSS ballot wordings. However, the illustration did not affect the response accuracy, so we report combined results for both survey versions throughout. A spreadsheet of the survey results are included as Supplemental Material.

The GSS data to which we compared our survey were taken from appendix summary tables 7-9 and 7-10 included in the 2012 National Science Foundation SEI (“National Science Board - Science and Engineering Indicators 2014,” 2014).

2.2. Matching Demographics

The demographics of the AMT respondents differed from the GSS respondents. We adjusted for these differences by weighting the AMT responses as follows. Each respondent is categorized by their gender (male (m), female (f)), age (under or equal to the age of 44 (y), over the age of 44 (o)), and education level (no college degree (n), college degree (c)). We then computed the ratio, α , between the percentage of the GSS respondents and the AMT respondents that fell into each of these categories:

$$\begin{aligned} \alpha_m &= 49/58 & \alpha_f &= 51/42 \\ \alpha_y &= 48/85 & \alpha_o &= 52/15 \\ \alpha_n &= 70/53 & \alpha_c &= 30/47. \end{aligned} \tag{1}$$

Based on their demographic, a weighting (w) was computed for each respondent. For example, the weighting for a female respondent, under the age of 44, and holding a college degree is: $w = \alpha_f \times \alpha_y \times \alpha_c$, or the weighting for a male respondent, over the age of 44, and not holding a college degree is: $w = \alpha_m \times \alpha_o \times \alpha_n$. With a weighting factor computed for each respondent, we re-computed the percent correct for each question as:

$$\frac{\sum_{i \in C} w_i}{\sum_{i=1}^N w_i}, \tag{2}$$

where the set C corresponds to all correct responses and N is the total number of responses. That is, each respondent contributed to the overall accuracy proportional to the size of their demographic group relative to the GSS demographic group.

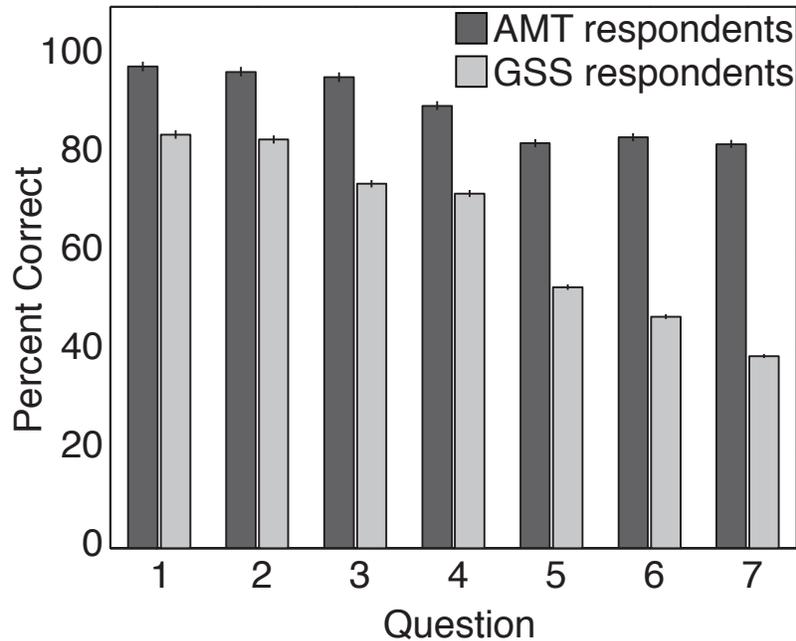


Figure 1: Percent correct responses for scientific literacy questions. The dark bars correspond to our AMT respondents and the light bars correspond to the GSS respondents. The error bars correspond to 95% confidence intervals calculated using the Clopper-Pearson method. See Table 1 for the list of questions.

3. Results

Shown in Figure 1 (dark bars) is the percentage of responses that are correct for each of the seven scientific literacy questions, Table 1. These percentages are 98%, 97%, 96%, 90%, 82%, 83%, and 82%. The average across all questions is 90%. Also shown in Figure 1 (light bars) is the same result for the GSS respondents to these same seven questions. The average across all questions for these respondents is 65%, considerably lower than for the AMT respondents. For example, based on the GSS study, it was widely reported that 1 in 4 Americans does not know that the earth revolves around the sun (Henderson, 2014; Neuman, 2014; O’Neill, 2014), whereas this statistic drops to 1 in 25 for our AMT respondents.

At least one explanation for the differences between the AMT and GSS respondents may be a difference in demographics. Shown in Figure 2 is a comparison of gender, age, and education of these two populations. Over-

all, the AMT respondents are more likely to be male, younger, and have a higher level of education. To correct for these demographic differences, we re-weighted our responses to match the GSS demographics (see *Methods*). The adjusted AMT accuracy for each of the seven questions is 98%, 94%, 94%, 90%, 82%, 82%, and 80%, for an overall accuracy of 89%. These accuracies are only slightly lower than the non-adjusted levels and still considerably higher than levels for the GSS respondents. Demographics alone cannot explain the differences in the responses between these populations.

We also observed considerably less variability between individual demographic groups. For example, one of the poorest performing demographics in the GSS survey is females, with an average accuracy of 57% as compared to 72% for males. In our survey, however, the females have an average accuracy of 87% as compared to 91% for the males. The GSS respondents under the age of 44 years have an accuracy of 67% as compared to 63% for those over the age of 44. Our respondents have an accuracy of 90% regardless of age. Similarly, The GSS respondents without a college degree have an accuracy of 57% as compared to 78% for those with a college degree. Our respondents have an accuracy of 88% and 92%, respectively. Overall, the AMT respondents are considerably more homogeneous across demographic groups.

The high rate of correct responses in our survey could be interpreted as evidence that AMT respondents may have used external sources to answer the questions. This, however, seems unlikely given that the survey was completed in an average of just over two minutes. In addition, since payment is not contingent on correct responses, respondents have little incentive to fact-check themselves, and similar surveys have reported wide-spread incorrect responses to more specialized science-based questionnaires, with rates similar to those collected via phone-survey (Carter et al., 2014; Simons & Chabris, 2012).

4. Discussion

The results of our survey reinforce the well-known discrepancies between rigorous probability samples and internet-based non-probability samples in population surveys (Baker et al., 2013; Yeager et al., 2011). When large differences such as those reported here result from probability and non-probability samples, it suggests that factors related to survey participation are correlated with the research question of interest. Statistical correction techniques, such as the demographics matching described in this paper, have

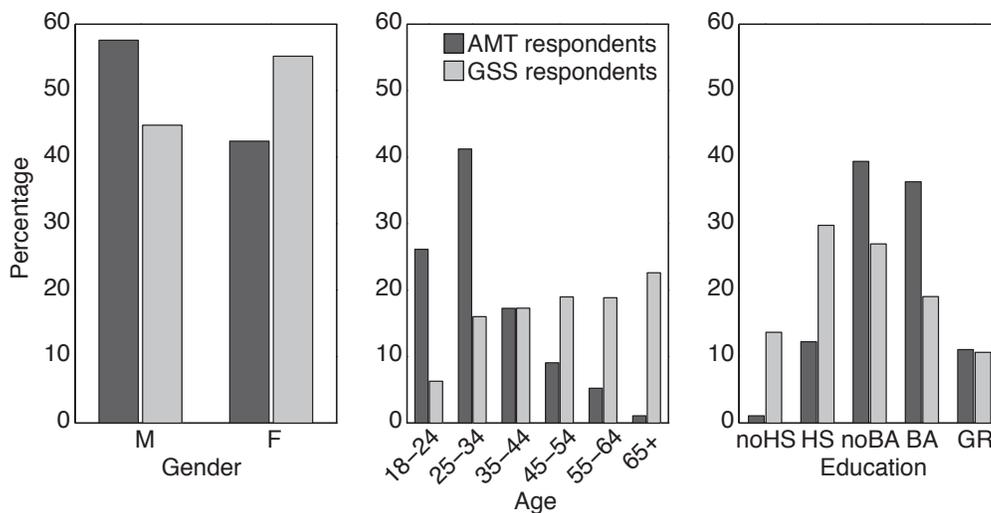


Figure 2: Respondent demographics. The dark bars correspond to the demographics (gender, age, education) of our AMT respondents and the light bars correspond to the demographics of the GSS respondents. Abbreviations: Didn't finish high school (noHS), finished high school (HS), some college (noBA), finished college (BA), graduate/professional degree (GR).

been shown to reduce such coverage bias, but cannot eliminate it completely (Couper, 2000; Dever, Rafferty, & Valliant, 2008; Schonlau, van Soest, Kapteyn, & Couper, 2009). AMT workers, therefore, and perhaps the subset of the US population that is highly Internet-engaged, may have a higher level of scientific knowledge than the US population as a whole. It would be reasonable to assume that the difference in knowledge reported here is not isolated to only science-related knowledge, and that the AMT workers may be generally more literate and knowledgeable than the average population. Any studies asking new questions about population knowledge or attitudes using AMT should take care to consider this potential difference. A good strategy might be to include some related standard survey questions in order to assess this knowledge difference.

On the other hand, there are several discrepancies in how our survey and the GSS survey were administered that may contribute to the better performance apart from population sample differences. Our survey consisted of 13 written questions, which participants selected to respond to while actively seeking tasks on AMT, and completed in an average of two minutes. In comparison, the participants of the GSS survey were selected by population

sampling, and responded to a lengthy in-person verbal interview that lasted 90 minutes on average “General Social Survey - FAQs,” 2014). Given the importance of surveys as a barometer of public engagement and ability related to the sciences, it is worthwhile to consider that these surveys face significant challenges associated with developing a pure assessment of scientific knowledge and literacy.

5. References

1. Allum, N., Sturgis, P., Tabourazi, D., & Brunton-Smith, I. (2008). Science knowledge and attitudes across cultures: a meta-analysis. *Public Understanding of Science*, 17(1), 35-54. doi: 10.1177/0963662506070159
2. Attari, S. Z. (2014). Perceptions of water use. *Proceedings of the National Academy of Science*, 111(14), 5129-5136. doi: 10.1073/pnas.1316402111
3. Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., Tourangeau, R. (2013). Summary Report of the AAPOR Task Force on Non-probability Sampling. *Journal of Survey Statistics and Methodology*, 1(2), 90-143. doi: 10.1093/jssam/smt008.
4. 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(1), 3-5. doi: 10.1177/1745691610393980
5. Carter, R. R., Difeo, A., Bogie, K., Zhang, G. Q., & Sun, J. (2014). Crowdsourcing Awareness: Exploration of the Ovarian Cancer Knowledge Gap through Amazon Mechanical Turk. *PLoS One*, 9(1), e85508. doi: 10.1371/journal.pone.0085508
6. Couper, Mick P. (2000). Web Surveys. A Review of Issues and Approaches. *Public Opinion Quarterly*, 64, 464-494.
7. Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon’s Mechanical Turk as a tool for experimental behavioral research. *PLoS One*, 8(3), e57410. doi: 10.1371/journal.pone.0057410
8. Dever, Jill A., Rafferty, Ann, & Valliant, Richard. (2008). Internet surveys: Can statistical adjustments eliminate coverage bias? *Survey Research Methods*, 2(2), 47-62.
9. General Social Survey - FAQs. (2014). Retrieved 02-March, 2014, from <http://www3.norc.org/GSS+Website/FAQs>
10. Goodman, Joseph K., Cryder, Cynthia E., & Cheema, Amar. (2013). Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. *Journal of Behavioral Decision Making*, 26(3), 213-224. doi: 10.1002/bdm.1753
11. GSS - General Social Survey. (2014). Retrieved 02-March, 2014, from <http://www3.norc.org/GSS+Website>
12. Henderson, B. (2014). One in four Americans do not know the Earth circles the Sun. Retrieved 02-March, 2014, from <http://www.telegraph.co.uk/news/worldnews/northamerica/usa/10640518/one-in-four-Americans-do-not-know-the-Earth-circles-the-Sun.html>.

13. Horton, John J., Rand, David G., & Zeckhauser, Richard J. (2011). The online laboratory: conducting experiments in a real labor market. *Experimental Economics*, 14(3), 399-425. doi: 10.1007/s10683-011-9273-9
14. Ipeirotis, Panos. (2010). Demographics of Mechanical Turk. New York University Working Paper.
15. Mason, Winter, & Suri, Saddharth. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavioral Research Methods*, 44(1), 1-23.
16. Miller, Jon D. (1983). Scientific Literacy: A Conceptual and Empirical Review. *Daedalus*, 112(2), 29-48.
17. Miller, Jon D. (1998). The measurement of civic scientific literacy. *Public Understanding of Science*, 7(3), 203-223. doi: 10.1088/0963-6625/7/3/001
18. Moser, C. A., & Kalton, G. (1971). *Survey Methods in Social Investigation*. London: Heinemann Educational Books.
19. National Science Board - Science and Engineering Indicators 2014. (2014). Retrieved 02-March, 2014, from <http://www.nsf.gov/statistics/seind14>
20. Neuman, S. (2014). 1 in 4 Americans thinks the Sun goes around the Earth, survey says. Retrieved 02-March, 2014, from <http://www.npr.org/blogs/thetwo-way/2014/02/14/277058739>
21. O'Neill, I. (2014). 1 in 4 Americans don't know Earth orbits the Sun. Yes, really. Retrieved 02-March, 2014, from <http://news.discovery.com/space/astronomy/1-in-4-americans-dont-know-earth-orbits-the-sun-yes-really-140214.htm>
22. Paolacci, Gabriele, Chandler, Jesse, & Ipeirotis, Panagiotis G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5), 411.
23. Patihis, L., Ho, L. Y., Tingen, I. W., Lilienfeld, S. O., & Loftus, E. F. (2014). Are the "memory wars" over? A scientist-practitioner gap in beliefs about repressed memory. *Psychol Sci*, 25(2), 519-530. doi: 10.1177/0956797613510718
24. Ross, Joel, Irani, Lilly, Silberman, M., Zaldivar, Andrew, & Tomlinson, Bill. (2010). Who are the Crowdworkers?: Shifting Demographics in Amazon Mechanical Turk. *Proceedings of ACM CHI Conference*.
25. Schonlau, M., van Soest, A., Kapteyn, A., & Couper, M. (2009). Selection Bias in Web Surveys and the Use of Propensity Scores. *Sociological Methods & Research*, 37(3), 291-318. doi: 10.1177/0049124108327128.
26. Simons, D. J., & Chabris, C. F. (2012). Common (mis)beliefs about memory: a replication and comparison of telephone and Mechanical Turk survey methods. *PLoS One*, 7(12), e51876. doi: 10.1371/journal.pone.0051876
27. Smith, Tom W. (1978). Response rates on the 1975-1978 General Social Surveys with comparisons to the omnibus surveys of the survey research center, 1972-1976. *GSS Methodological Report*, 5, 1-6.
28. Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. *Public Opinion Quarterly*, 75(4), 709-747. doi: 10.1093/poq/nfr020.