Article

P | U | S

Public Understanding of Science
2016, Vol. 25(2) 146–153

DOI: 10.1177/0963662514554354
pus.sagepub.com
SAGE

Does the Sun revolve around the Earth? A comparison between the general public and online survey respondents in basic scientific knowledge

Emily A. Cooper
Stanford University, USA

Hany Farid
Dartmouth College, USA

Abstract
We conducted an online survey using a set of factual science questions that are commonly administered to assess fact-based scientific literacy. We report that the online 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 does not know that the Earth revolves around the Sun, whereas among the online population, this ratio is reduced to 1 in 25. While new online platforms provide researchers with unprecedented ease of access to a large sample population for studying trends in public knowledge and attitudes, generalizing from online population samples to the US population at large poses a considerable challenge. We discuss the potential reasons for this discrepancy and the implications for conducting research online.

Keywords
Amazon Mechanical Turk, online surveys, scientific literacy, survey methods

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 et al., 2011; Horton et al., 2011; Paolacci et al., 2010). Depending on the type of research question being asked, however, biases in the AMT population may compromise the

Corresponding author:
Emily A. Cooper, Department of Psychology, Stanford University, 450 Serra Mall, Stanford, CA 94305, USA.
Email: eacooper@stanford.edu

generalizability of results acquired using this platform (Goodman et al., 2013; Horton et al., 2011; Paolacci et al., 2010). For example, compared to the general US population, AMT workers in the United States are younger, more educated, less racially diverse, and more likely to be female (Ipeirotis, 2010; Mason and Suri, 2012; Ross et al., 2010; Simons and 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 et al., 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 among 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 are an immediate concern. 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 and 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 et al., 2014), one assessed knowledge related to water conservation (Attari, 2014), and two assessed pop-science knowledge about human memory (Patihis et al., 2014; Simons and Chabris, 2012). Here, we report the results of an AMT survey assessing general factual scientific knowledge aspects of 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 through 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 1980s and is surprisingly low—a small percentage of the public exhibit high scores for basic "textbook" scientific knowledge (Allum et al., 2008; Miller, 1983, 1998).

We surveyed 1014 AMT workers using a subset of factual physical 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, 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—SEI, 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 knowledge. We discuss how the scientific knowledge of the AMT population may have implications for AMT research applications.

2. Methods

Scientific knowledge survey

Using a standard Mechanical Turk survey template, we published a Human Intelligence Task (HIT) titled "Thirteen Question Quiz" with the short description "Answer thirteen short multiple choice

Table 1. Survey questions.

No.	Statement/Question	Response choices
1	The center of the Earth is very hot.	*True*/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.	*True*/False
3	Does the Earth go around the Sun, or does the Sun go around the Earth?	*Earth around Sun*/Sun around Earth
–	Strawberries are red.	*True*/False
4	All radioactivity is man-made.	True/*False*
5	Electrons are smaller than atoms.	*True*/False
–	There are 5 hours in a day.	True/*False*
6	Lasers work by focusing sound waves.	True/*False*
7	The universe began with a huge explosion.	*True*/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

AMT: Amazon Mechanical Turk; GSS: General Social Survey.
The online AMT survey contained seven scientific questions from the GSS (1–7), three control questions (not numbered), and three demographic questions (8–10). Correct answers to the scientific and control questions are indicated with italics.

questions." The HIT was limited to respondents in the United States over the age of 18 years, to those who have a HIT Approval Rate greater than or equal to 95%, and to those who have 50 or more previously approved HITs. Respondents were paid US$0.50 regardless of how they performed on the survey.

Shown in Table 1 are the 13 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 10 questions are 3 simple control questions, which are used to ensure that the respondent reads each question. We published a total of 1037 HITs, each of which was 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 to confusion caused by the wording of the question, which was identical to the GSS ballot wording. 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 is 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—SEI, 2014).

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

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

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.

3. Results

Shown in Figure 1 (dark bars) is the percentage of responses that are correct for each of the seven scientific knowledge 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. Overall, 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 section "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

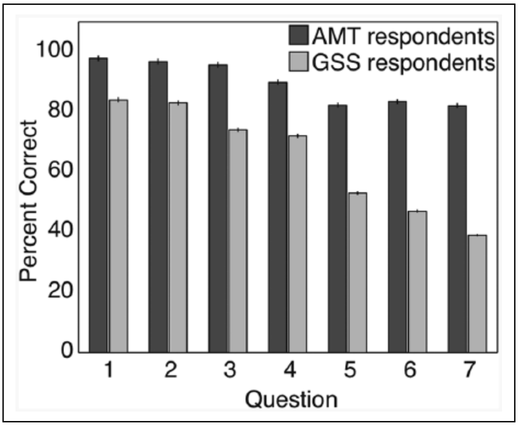

Figure 1. Percent correct responses for scientific knowledge questions.
AMT: Amazon Mechanical Turk; GSS: General Social Survey.
The dark bars correspond to our AMT respondents, and the light bars correspond to the GSS respondents. The error bars correspond to 95% confidence intervals calculated using the Clopper–Pearson method. See Table 1 for the list of questions.

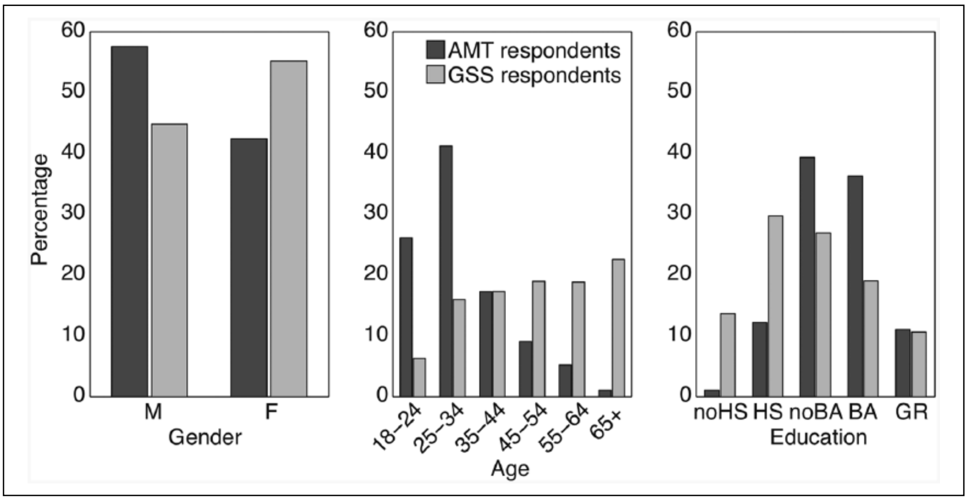

Figure 2. Respondent demographics.
AMT: Amazon Mechanical Turk; GSS: General Social Survey; noHS: didn't finish high school; HS: finished high school; noBA: some college; BA: finished college; GR: graduate/professional degree.
The dark bars correspond to the demographics (gender, age, education) of our AMT respondents, and the light bars correspond to the demographics of the GSS respondents. See Table 1 for unabbreviated response choices.

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 2 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 through phone survey (Carter et al., 2014; Simons and 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 article, have been shown to reduce such coverage bias, but cannot eliminate it completely (Couper, 2000; Dever et al., 2008; Schonlau et al., 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 2 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 (GSS—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.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a grant from the National Science Foundation (CNS-0708209).

References

Allum N, Sturgis P, Tabourazi D and 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.

Attari SZ (2014) Perceptions of water use. *Proceedings of the National Academy of Sciences of the United States of America* 111(14): 5129–5136. DOI:10.1073/pnas.1316402111.

Baker R, Brick JM, Bates NA, Battaglia M, Couper MP, Dever JA, et al. (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.

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. DOI:10.1177/1745691610393980.

Carter RR, Difeo A, Bogie K, Zhang GQ and 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.

Couper MP (2000) Web surveys: A review of issues and approaches. *Public Opinion Quarterly* 64: 464–494.

Crump MJ, McDonnell JV and Gureckis TM (2013) Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE* 8(3): e57410. DOI:10.1371/journal.pone.0057410.

Dever JA, Rafferty A and Valliant R (2008) Internet surveys: Can statistical adjustments eliminate coverage bias? *Survey Research Methods* 2(2): 47–62.

General Social Survey (GSS) (2014) Available at: http://www3.norc.org/GSS+Website (accessed 02 March 2014)

General Social Survey (GSS)—FAQs (2014) Available at: http://www3.norc.org/GSS+Website/FAQs (accessed 02 March 2014)

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. DOI:10.1002/bdm.1753.

Henderson B (2014) One in four Americans do not know the Earth circles the Sun. Available at: http://www.telegraph.co.uk/news/worldnews/northamerica/usa/10640518/one-in-four-Americans-do-not-know-the-Earth-circles-the-Sun.html (accessed 02 March 2014)

Horton JJ, Rand DG and Zeckhauser RJ (2011) The online laboratory: Conducting experiments in a real labor market. *Experimental Economics* 14(3): 399–425. DOI:10.1007/s10683-011-9273-9.

Ipeirotis P (2010) Demographics of Mechanical Turk. *New York University* Working Paper, March.

Mason W and Suri S (2012) Conducting behavioral research on Amazon's Mechanical Turk. *Behavioral Research Methods* 44(1): 1–23.

Miller JD (1983) Scientific literacy: A conceptual and empirical review. *Daedalus* 112(2): 29–48.

Miller JD (1998) The measurement of civic scientific literacy. *Public Understanding of Science* 7(3): 203–223. DOI:10.1088/0963-6625/7/3/001.

Moser CA and Kalton G (1971) *Survey Methods in Social Investigation*. London: Heinemann Educational Books.

National Science Board—Science and Engineering Indicators 2014 (2014) Available at: http://www.nsf.gov/statistics/seind14 (accessed 02 March 2014)

Neuman S (2014) 1 in 4 Americans thinks the Sun goes around the Earth, survey says. Available at: http://www.npr.org/blogs/thetwo-way/2014/02/14/277058739 (accessed 02 March 2014)

O'Neill I (2014) 1 in 4 Americans don't know Earth orbits the Sun. Yes, really. Available at: http://news.discovery.com/space/astronomy/1-in-4-americans-dont-know-earth-orbits-the-sun-yes-really-140214.htm (accessed 02 March 2014)

Paolacci G, Chandler J and Ipeirotis PG (2010) Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5(5): 411–419.

Patihis L, Ho LY, Tingen IW, Lilienfeld SO and Loftus EF (2014) Are the "memory wars" over? A scientist-practitioner gap in beliefs about repressed memory. *Psychological Science* 25(2): 519–530. DOI:10.1177/0956797613510718.

Ross J, Irani L, Silberman M, Zaldivar A and Tomlinson B (2010) Who are the crowdworkers? Shifting demographics in Amazon Mechanical Turk. In: *Proceedings of ACM CHI conference*, Atlanta, GA, 10–15 April 2010.

Schonlau M, van Soest A, Kapteyn A and 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.

Simons DJ and Chabris CF (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.

Smith TW (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.

Yeager DS, Krosnick JA, Chang L, Javitz HS, Levendusky MS, Simpser A, et al. (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.

Author biographies

Emily A. Cooper is a Postdoctoral Research Scholar at Stanford University. Her research interests are in human visual perception and neuroscience.

Hany Farid is a Professor of Computer Science at Dartmouth College. His research interests are in image processing, computer vision, and human perception.