International Journal of Communication 15(2021), 4383–4398 1932–8036/20210005

Do the Online Activities of Amazon Mechanical Turk Workers Mirror Those of the General Population? A Comparison of Two Survey Samples

AARON SHAW[1]
Northwestern University, USA

ESZTER HARGITTAI
University of Zurich, Switzerland

Amazon Mechanical Turk (AMT) offers a relatively low-cost alternative to traditional expensive survey samples, which likely explains its popularity among survey researchers. An important question about using such samples is whether they are representative of the larger Internet user population. Though prior research has addressed this question about demographic characteristics, little work has examined how AMT workers compare with others regarding their online activities—namely, social media experiences and online active engagement. This article analyzes survey data administered concurrently on an AMT and a national sample of U.S. adults to show that AMT workers are significantly more likely to use numerous social media, from Twitter to Pinterest and Reddit, as well as have significantly more experiences contributing their own online content, from posting videos to participating in various online forums and signing online petitions. The article discusses the implications of these findings for research that uses AMT as a sampling frame when examining questions related to social media use and active online engagement.

Keywords: Amazon Mechanical Turk, survey methods, data bias, social media use, online participation

Administering surveys on the general population using traditional methods such as postal mail and phone can be prohibitively expensive. Online samples offer a helpful alternative with considerably lower costs (Couper & Miller, 2008). Amazon's Mechanical Turk (AMT) is one example of a platform that offers cheaper alternatives to traditional data collection for academics (Mason & Suri, 2012). While this is a

---

Aaron Shaw: aaronshaw@northwestern.edu
Eszter Hargittai: pubs@webuse.org
Date submitted: 2020-11-24

[1] The authors contributed equally to this work and are grateful to the anonymous reviewers for helpful feedback on the article. We also acknowledge the support of Merck (known as MSD outside the United States and Canada) and the Robert and Kaye Hiatt Fund at Northwestern University. Thanks to Sanja Stojanovic, who contributed to formatting and proofreading the final article.

welcomed addition to possible data sources, biases may be present in relying on such a platform for sampling. Depending on one's research questions, sampling on AMT may present serious limitations. For example, if research questions have to do with social media uses—an increasingly large focus of scholarly attention—then knowing how AMT respondents may differ from the general population in their social media use experiences is essential. If research questions are related to differences in active online engagement, then knowing how the AMT respondent pool differs from the general Internet user population in online participation is significant. Little prior work has evaluated AMT samples on these factors.

This article analyzes responses to identical survey questions administered to both AMT participants and a U.S. national sample at the same time. We compare the two samples on their use of social media sites, focusing on what platforms participants use and also look at how the samples differ in their active online engagement. We show wide variation in these online experiences of the two samples. We discuss the implications of these findings for data collected on AMT for studying people's online behaviors especially those concerning social media experiences and active online engagement.

Prior Work Evaluating AMT Samples

When compared with studies that mainly rely on undergraduate samples, as is often the case in psychology and some communication research, AMT offers more diversity on age and education than those more traditional sampling frames (e.g., Horton, Rand, & Zeckhauser, 2011; Paolacci & Chandler, 2014). Work in that domain has also shown that AMT samples result in high data quality when it comes to psychometric scales, test–retest outcomes, and the implications of varying compensation rates. Considerable scholarship has also evaluated AMT survey samples to determine how they compare on demographic characteristics, psychological attitudes, as well as religious and political beliefs (Berinsky, Huber, & Lenz, 2012; Clifford, Jewell, & Waggoner, 2015; Goel, Obeng, & Rothschild, 2017; Hargittai & Shaw, 2020; Horton et al., 2011; Levay, Freese, & Druckman, 2016; Weinberg, Freese, & McElhattan, 2014). While such evaluations tend to find that demographically AMT samples are not representative of the general population, they also observe that AMT samples replicate previous findings about many beliefs and personality traits.

Closest to our analysis are two studies that compared responses from AMT and population-representative samples. Redmiles, Kross, Pradhan, and Mazurek (2017) administered the same survey on AMT, a Census-representative Web panel, and on a probabilistic telephone sample. In addition to comparing demographic characteristics, they asked about several online behaviors, especially focusing on those concerning privacy and security. Of particular interest here is that they found that among AMT respondents, 96.7% reported using social media compared with 73.7% of the general population telephone survey (the figure was 90.7% for the probabilistic Web panel) showing that AMT participants are indeed not representative when it comes to social media uses. In addition to replicating that question about social media use in general on different survey samples, we build on it by disaggregating social media platforms to see whether AMT workers are similar in their rate of social media adoption for some platforms versus others. Hargittai and Shaw (2020) compared the Internet skills of an AMT and national sample showing that the AMT sample had higher Internet skills, a factor that considerable research has shown to correlate positively with various online activities (e.g., Büchi, Just, & Latzer, 2016; Correa, 2010; Hargittai & Hinnant, 2008). We build on that work by examining whether specific types of active online engagement also differ by sample type. In the next section, we review work showing differential rates of social media adoption and active online engagement across the population to motivate our focus on these Internet experiences.

Variations in Social Media Adoption and Active Online Engagement

Online activities are neither universally nor equally distributed throughout the general population. Our focus on social media adoption and active online engagement builds from two extensive bodies of prior research, both of which we synthesize briefly here. Earlier research indicates that both sets of behaviors vary widely depending on sociodemographic attributes as well as other factors.

Since the early days of social media's rise, scholars have documented its unequal diffusion across the population (boyd, 2012; Hargittai, 2007). Though popular media accounts may give the impression that the whole world is on social media, this has never been the case. Research has shown considerable variations over time across platforms by population groups, meaning that not only is the use of social media not universal, people's sociodemographic characteristics relate to their likelihood of social network site adoption. When MySpace and Facebook were the most popular such sites, boyd (2012) and Hargittai (2007) both documented socioeconomic differences in their adoption, finding that those from less privileged backgrounds were more likely to be on MySpace, whereas those from higher SES adopted Facebook at higher rates. As Twitter gained traction, Hargittai and Litt (2011) showed that African American young adults were more likely to start using the site than people from other racial and ethnic backgrounds.

Blank and Lutz (2017) relied on the 2013 Oxford Internet Surveys to examine how various factors relate to the adoption of Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram. They found that age, gender, and income explain the use of various services. In their analyses of how sociodemographics relate to platform adoption, they control for factors such as autonomy of use, Internet skills, self-efficacy, privacy concerns, and more, so it is difficult to say whether their data may show more differences across population groups without those controls. Analyzing data from the 2015 British Election Study, Mellon and Prosser (2017) found age, gender, and education differences when comparing Facebook and Twitter with the general population. Analysis of survey data about Belgian adults from 2017, 2018, and 2019 also showed that age, gender and other factors were related to who was using Facebook, Twitter, and Instagram (Hellemans, Willems, & Brengman, 2020).

A study of undergraduate students at a U.S. university in 2014–15 found that women were considerably more likely to use Instagram than men (Sheldon & Bryant, 2016). Gazit, Aharony, and Amichai-Hamburger (2019) looked at gender differences in the use of Facebook, Instagram, Twitter, and WhatsApp among Israeli college students in 2017–18. They found (a) large variation in the popularity of these platforms (WhatsApp was most popular, Twitter least), and that (b) popularity varied by gender, whereby women used WhatsApp and Instagram more while men used Twitter more.

In sum, research spanning well over a decade across several countries both on younger populations and the general population has demonstrated that sociodemographic factors such as age, gender, race/ethnicity, and education are related to who adopts which social network sites. Given these differences, it is important to examine whether being an AMT user is itself related to being on various social media platforms.

Active online engagement is another area where research has identified significant variations across the population (Hargittai & Jennrich, 2016). It is beyond the scope of this article to review that literature fully, but we cite a few pieces to show the general patterns that have emerged. By active online participation, we refer to actions whereby the user shares their own voice and content online whether in the form of posting pictures or videos, voting in a poll, engaging in question-and-answer discussions, and so on. Researchers have pointed out that "listening" itself and nonactive participation can be an important contribution to online communities (Crawford, 2009; Lutz & Hoffmann, 2017), and it is not our intention to negate those contributions. Nevertheless, we focus part of our analysis on people who actively put their content online because doing so indicates a distinct kind of behavior with important implications for equity, representation, and inclusion.

In one of the first studies to look at such active online participation, Hargittai and Walejko (2008) found that while college women were more likely to create certain content, their male peers were more likely to share it. Also looking at university undergraduates, Correa (2010) observed variations by gender and race/ethnicity in online participatory behaviors. Analyzing data from 17 Pew Research Center surveys about American adults' online production activities between 2000 and 2008, Schradie (2011) showed variations in active engagement by age, gender, race/ethnicity, and education. Survey results from UK Internet users collected in 2011 also showed differences by age and education (Blank, 2013). Relying on 2014 data about Finnish adults' Internet uses, Ertiö, Kukkonen, and Räsänen (2018) found differences by age, gender and education when it comes to sharing content on social media and publishing one's own material.

While the cited studies examined a multitude of online participatory activities across a diverse set of respondents resulting in different specific findings, one common thread in the literature is that sociodemographics relate to who engages actively online. Given that such activities are not randomly distributed across the population, it may be that AMT workers themselves are more or less prone to engaging in them. Knowing where they fall on this activity spectrum is important depending on what research questions people are asking when relying on AMT as a sampling frame for their studies.

Data and Methods

We draw on two data sets collected at overlapping times to compare respondents from an Amazon Mechanical Turk sample consisting of U.S.-based participants with a national U.S. sample. Both surveys were administered online. For the national sample, we contracted with the independent research organization NORC (formerly the National Opinion Research Center) at the University of Chicago to administer questions to their AmeriSpeak panel online. The AmeriSpeak panel is designed to produce a representative sample of U.S. adults 18 years of age and over. Sampling procedure details are available from NORC (http://www.norc.org/Research/Capabilities/pages/amerispeak.aspx). In brief, NORC uses strata based on age, race/ethnicity, education, and gender as well as multiple factors related to differential response rates to approximate the target population. Supplementary documentation provided by NORC related to the specific attributes of the AmeriSpeak sample surveyed in this study is available from the authors on request. We ran the AmeriSpeak survey May 25–July 5, 2016, and the AMT survey June 24–28, 2016.

Independent Variables: Sociodemographics

Background variables about respondents such as their age, gender, education, income, and race/ethnicity were supplied by NORC based on previous data collection about the AmeriSpeak panel. We asked AMT respondents similarly about their sociodemographic characteristics so that the data could be comparable. We asked year of birth, which we subtracted from the year of data collection to get respondents' ages. The survey asked whether the respondent was male or female. We asked about highest level of school completed, with five answer options, which we recoded into three dummies indicating high school or less, some college, and college or more. The household income question had ranges as responses, which we recoded to midpoint values for a continuous measure. In the regression analyses, we use the square root of income as that best approximates a normal distribution. We have data on work status, which we use to compare those employed with others (not working, retired, disabled, other). The survey asked what best describes where the respondent lives and from this we created a dummy for rural residents. To measure race and ethnicity, following U.S. Census conventions, we first inquired about whether the person is of Spanish, Hispanic, or Latino descent, followed by a question about race with these categories: White/Anglo/Caucasian/Middle Eastern, Black/African American, American Indian or Alaska Native, Asian, and some other race. We created mutually exclusive dummies for race and ethnicity with White as base in the regression models.

Measures: General Internet Experiences

We include measures for how much autonomy participants have in freely accessing the Internet when and where they want to, how much time they spend online, and their Internet skills. Prior literature has found these variables important in understanding people's online experiences (Ahn, 2011; Haight, Quan-Haase, & Corbett, 2014; Howard, Rainie, & Jones, 2001).

To measure autonomy of use, we asked, "At which of these locations do you have access to the Internet—that is, if you wanted to, you could use the Internet at which of these locations?"—followed by nine options such as home, workplace, and friend's home. To assess frequency of use, we asked, "On an average weekday, not counting time spent on email, chat, and phone calls, about how many hours do you spend visiting websites?"; we then asked the same question about "average Saturday or Sunday." The answer options ranged from "None" to "6 hours or more," with six additional options in between. We calculated weekly hours spent on the Web by multiplying the answer to the first question by five, the second question by two, and adding these two figures together.

To assess Internet skills, we use a validated, established measure (Hargittai & Hsieh, 2012). Respondents were asked to rate their level of understanding of six Internet-related terms (such as cache, spyware, phishing) on a 5-point scale ranging from *no understanding* to *full understanding*. The Internet skills measure is the mean of the six items (Cronbach's alpha = .94).

Dependent Variables: Social Media Use and Online Participatory Activities

Social Media Use

We asked participants whether they use various social media by first asking whether they have heard of certain sites, and if they responded yes, then following up with:

> Have you ever visited the following sites and services? For each site or service, indicate if no, you have never visited it; yes, you have visited it in the past, but do not visit it nowadays; yes, you currently visit it sometimes; yes, you currently visit it often.

The sites we include here are either the most popular social media platforms or have been the subject of considerable academic work: Facebook, Instagram, LinkedIn, Pinterest, Reddit, Snapchat, and Twitter.

Online Participatory Activities

To get a sense of how engaged people are online, we asked about several online participatory activities with the following wording: contributed to a citizen science project online, contributed to a crowdfunding campaign, made a loan on a microfinance site, signed a petition on an online petition site, added a coupon code to a site with coupon codes, submitted a product review on a specific brand retailer's site, asked or answered a question in an online forum, asked or answered a question in a social Q&A site, posted a video privately, posted a video publicly. These were dichotomous yes (1) and no (0) questions.

Analysis

We compare the two samples in multiple ways. First, we calculate descriptive statistics for all of our measures. Second, to test whether these relationships are robust to controlling for other variables, we estimate multiple regression models where we analyze the various online activities as outcomes to see whether being in the NORC versus AMT sample makes a difference.

Results

Table 1 presents the descriptive statistics for both the NORC and AMT samples showing that there are considerable differences by age (younger on AMT), income (lower on AMT), education (higher on AMT), race (fewer Blacks, more Asians on AMT), ethnicity (fewer people of Hispanic descent on AMT), and rural residence (more common on AMT). The two samples also diverge in their online experiences whereby the average AMT respondent can access the Internet in more locations, spends more time online, and has higher Internet skills than the average NORC participant. Given these variations in sociodemographics and basic online experiences, it will be important to control for these factors when seeing whether they explain any differences we observe across the two samples in social media experiences. Next, we turn to discussing those variations.

Table 1. Descriptive Statistics About Both Samples, With Asterisks Indicating Statistically Significant Differences Between the Two Samples.

	NORC				AMT			
	Percentage	Mean	*SD*	*N*	Percentage	Mean	*SD*	*N*
Background								
Age (18–94)***		48.7	16.9	1,512		33.8	11.3	1,202
Income in U.S. $1,000s (2.5–225)***		71.5	54.4	1,512		51.5	38.1	1,159
Female	51			1,512	48			1,203
Employed	62			1,512	62			1,204
Rural resident**	13			1,512	18			1,204
Education								
High school or less***	26			1,512	11			1,204
Some college	32			1,512	33			1,204
Bachelor's or higher***	43			1,512	56			1,204
Race & Ethnicity								
White	71			1,511	73			1,200
Hispanic**	12			1,511	8			1,204
Black*	11			1,511	9			1,200
Asian***	3			1,511	9			1,200
Native American	2			1,511	2			1,200
Internet Experiences								
Internet autonomy (0–9)***		4.8	2.3	1,512		5.8	2	1,204
Internet use frequency (0–42)***		14.7	10.8	1,491		24	12.1	1,198
Internet skills (1–5)***		3.4	1.1	1,512		4	0.8	1,203

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

Use of Social Media Platforms by Survey Sample

The top part of Table 2 shows what portion of the two samples uses various social media platforms. AMT respondents are statistically significantly more likely to use all of the social media included on the survey. The smallest difference is in Facebook use, which 80% of the NORC sample compared with 84% of the AMT sample reports visiting. The next most popular platform among NORC respondents is Pinterest, at 42%, which is more popular among AMT respondents, at 50%. A third (33%) of NORC participants reported LinkedIn use, compared with 39% of AMT. Instagram is similarly popular among NORC respondents, at 34%, while considerably more popular in the AMT sample, at 54%. We see vast differences in Twitter use, which just over a quarter (27%) of NORC respondents visit compared with almost two-thirds (65%) of AMT participants. The difference is even more pronounced for Reddit, which is the least popular social media platform in the NORC sample, at 12%, compared with almost three-quarters (73%) of the AMT sample visiting the site. There is also considerable difference in Snapchat use, with AMT respondents 50% more likely to report using it (20% vs. 31%). Given such wide divergences in social media experiences, it is essential that studies looking at social media use be careful about where they draw their samples depending on their research questions. For example, if the point of a research project is to learn more about Reddit users, then AMT may offer a helpful sampling frame. On the other hand, if the point is to examine the role Twitter plays in how Internet users in general are exposed to information, then relying on an AMT sample may be problematic.

Table 2. Experiences With Online Activities of Both Samples.

	NORC		AMT		
	Percentage	*N*	Percentage	*N*	% Δ
Social Media Use					
Facebook**	80	1,506	84	1,203	4
Pinterest***	42	1,495	50	1,199	8
LinkedIn**	33	1,498	39	1,202	6
Instagram***	34	1,501	54	1,200	20
Twitter***	27	1,504	65	1,200	48
Snapchat***	20	1,491	31	1,194	11
Reddit***	12	1,490	73	1,193	61
Online Participatory Activities					
Contributed to citizen science project online***	2	1,509	7	1,203	5
Contributed to a crowdfunding campaign***	13	1,505	27	1,200	14
Made a loan on a microfinance site***	2	1,508	5	1,199	3
Signed a petition on an online petition site***	35	1,508	50	1,200	15
Added a coupon code to a site with coupon codes	36	1,505	36	1,199	0
Submitted a product review on a specific brand retailer's site	36	1,508	37	1,202	1
Asked or answered a question in an online forum***	47	1,509	64	1,202	17
Asked or answered a question in a social Q&A site***	15	1,508	30	1,200	15
Posted a video publicly***	39	1,508	55	1,202	16
Participated in a political poll***	40	1,506	54	1,202	14

Note. % Δ = percentage point difference between the AMT and NORC samples.
** $p < .01$, *** $< .001$.

As noted earlier, it is important to examine whether these differences are simply a reflection of sociodemographic variations and frequency or autonomy of use and Internet skills across the two samples or, whether controlling for such factors, we find independent associations between being on AMT versus in the general Internet user population. To answer this question, we fit logistic regression models with each social media platform use as the outcome. As the first row of results (AMT sample) in Table 3 shows, for Pinterest, LinkedIn, and Instagram, once we control for other factors, AMT workers are no more likely to visit these platforms than NORC respondents. In the case of Twitter and especially Reddit, they are much more likely to be users of the platform even when we control for other factors. That is, independent of the differences in age, education, race/ethnicity, Internet use frequency, autonomy, and skills between the AMT versus national sample, AMT respondents are more likely to use these two platforms. For Snapchat and marginally for Facebook ($p < .085$), they are less likely.

Active Online Engagement by Survey Sample

Next, we turn to comparing the two samples on online participatory activities. Here, again, AMT respondents are significantly more active than NORC participants. In all, but two of the 10 cases we observe statistically significant variations (see bottom half of Table 2). The exceptions are having "added a coupon code to a site with coupon codes" at 36% in both samples, and having "submitted a product review on a specific brand retailer's site" at 37% (AMT) and 36% (general population). With no activity are NORC respondents more engaged than AMT participants. Whether it concerns making a loan on a microfinance site or publicly posting a video, AMT respondents have considerably more experiences. The rightmost column in Table 2 shows the percentage point differences between the two samples.

As with social media platform use, next we turn to logistic regression analyses to see whether these findings hold once controlling for sociodemographic characteristics, as well as Internet use frequency, autonomy, and skills. Table 4 shows that even while accounting for all of those sample differences, AMT workers are still more likely to have contributed to a crowdfunding campaign, to have made a loan on a microfinance site, to have asked or answered a question on an online forum as well as on a social Q&A site, to have submitted a vote to an online political poll, and marginally ($p < .055$) to have contributed to a citizen science project. They are, however, not more likely to have signed a petition, to have contributed a coupon to a coupon code site, to have submitted a product review on a specific brand's site, or to have posted a video publicly. In sum, even once controlling for numerous ways in which the samples differ, AMT workers are still more likely to engage in several online activities than NORC respondents.

Table 3. Logistic Regression Analyses on Social Media Platform Uses.

	Facebook		Pinterest		LinkedIn		Instagram		Twitter		Snapchat		Reddit	
AMT sample	−.24#	.14	.13	.11	.01	.11	.06	.10	1.09***	.10	−.30*	.12	2.38***	.12
Age	−.01	.00	−.01	.00	.01*	.00	−.05***	.00	−.01***	.00	−.07***	.00	−.05***	.00
Female	1.06***	.11	1.69***	.09	.12	.09	.59***	.09	−.02	.09	.38***	.10	−.61***	.12
Hispanic	−.02	.19	−.04	.15	.07	.15	.51**	.15	.01	.15	.19	.16	−.38	.20
Black	−.11	.18	−.23	.15	.05	.15	.54***	.15	.23	.15	.04	.16	−.22	.19
Asian	−.21	.23	−.20	.19	.29	.19	.29	.19	.19	.20	−.35	.21	.82**	.25
Native American	−.08	.41	−.34	.33	−.87*	.39	−.53	.35	−.75*	.36	−1.03*	.48	.25	.44
Some college	.30*	.15	.01	.13	.55***	.15	−.00	.13	.21	.14	−.16	.15	.54**	.18
College or more	.28	.15	−.03	.13	1.00***	.14	.10	.13	.21	.13	−.32*	.15	.48**	.18
Income	−.00	.00	.00	.00	.00***	.00	.00**	.00	.00	.00	.00	.00	−.00	.00
Rural resident	.29	.16	.27*	.12	−.25	.13	−.23	.13	−.01	.13	−.25	.15	−.17	.16
Employed	.38**	.12	.13	.10	.40***	.10	.14	.10	.25*	.10	.28*	.11	−.40**	.13
Autonomy of use	.10***	.03	.12***	.02	.08***	.02	.10***	.02	.06**	.02	.16***	.03	.06	.03
Frequency of use	.02**	.01	.01	.00	.01	.00	.01*	.00	.02***	.00	.01**	.00	.02***	.01
Internet skills AMT	.14*	.06	.19***	.05	.49***	.05	.30***	.05	.37***	.05	.13*	.06	.80***	.07
N	2,630		2,616		2,622		2,624		2,626		2,611		2,606	

< .10, * *p* < .05, ** *p* < .01, *** < .001.

Table 4. Logistic Regression Analyses on Active Online Engagement.

	Citizen science		Crowd-funding		Loan		Petition		Coupon code		Product review		Online forum Q&A		Social Q&A		Video post		Political poll	
AMT sample	.48#	.25	.35**	.12	.56*	.26	.14	.10	−.06	.10	−.03	.10	.21*	.10	.33**	.12	−.13	.10	.42***	.10
Age	−.02	.01	−.02**	.00	−.01	.01	−.00	.00	.00	.00	.01***	.00	−.01*	.00	−.01**	.00	−.03***	.00	.02***	.00
Female	−.18	.21	.23*	.11	.02	.22	.57***	.09	.31***	.08	.47***	.09	.45***	.08	.27**	.10	.14	.09	−.15	.08
Hispanic	−.09	.37	−.23	.19	−.52	.48	−.20	.14	−.20	.15	.04	.15	−.24	.14	.06	.17	−.03	.14	−.25	.15
Black	−.39	.41	−.81***	.21	.18	.33	−.16	.14	.07	.14	.44**	.14	−.02	.14	.46**	.16	.04	.14	−.45**	.14
Asian	.62	.32	−.20	.21	.14	.38	−.14	.18	.10	.18	−.38	.20	−.75***	.18	−.36	.23	−.60**	.19	−.83***	.19
Native American	−.52	1.03	−.03	.41	1.05	.56	−.35	.33	−.00	.31	.23	.32	.76*	.36	.27	.36	−.02	.32	−.08	.32
Some college	.52	.41	.65**	.20	.41	.45	.46***	.13	.19	.12	.13	.13	.29*	.12	.28	.16	.00	.13	.64***	.13
College or more	.94*	.40	1.00***	.19	1.09*	.42	.44**	.13	.01	.12	−.05	.13	.08	.12	.08	.16	−.31*	.13	.78***	.13
Income	−.00	.00	.00	.00	−.00	.00	−.00	.00	.00	.00	.00*	.00	−.00	.00	−.00	.00	−.00	.00	−.00	.00
Rural resident	.19	.28	−.15	.15	−.69	.39	−.07	.12	−.02	.12	−.04	.12	.17	.12	.34*	.14	.11	.12	−.01	.12
Employed	.31	.24	.24	.12	.54*	.27	−.09	.09	−.03	.09	.09	.09	.12	.09	.08	.11	.12	.10	−.16	.09
Autonomy of use	−.04	.05	.06*	.02	−.14**	.05	.06**	.02	.05**	.02	.12***	.02	.12***	.02	.08**	.03	.12***	.02	.11***	.02
Frequency of use	.02	.01	.01*	.00	.01	.01	.01**	.00	−.00	.00	.01	.00	.01**	.00	.02***	.00	.02***	.00	.01**	.00
Internet skills AMT	.66***	.14	.44***	.07	.38**	.14	.42***	.05	.14**	.05	.35***	.05	.29***	.05	.45***	.06	.33***	.05	.37***	.05
N	2,633		2,628		2,630		2,631		2,627		2,632		2,632		2,630		2,632		2,630	

# *p* < .10, * *p* < .05, ** *p* <.01, *** < .001.

Discussion

We observe several notable similarities and differences between the AMT and general population samples compared in this study. Overall, AMT workers were more likely to have visited most of the social media sites and engaged in most of the online activities we include in our survey questions, although adjustment for sociodemographic factors and Internet use experiences and skills accounts for some of these differences. More active forms of online participation appear more differentiated across the two samples, with AMT workers more likely to engage in other online activities even after adjusting for several factors.

Some of the prior work we considered at the outset of this article suggests that differences between AMT survey samples and general population samples need not prevent the use of AMT for research purposes. We agree that AMT and other nonrepresentative online data sources offer important advantages over population samples (most critically in terms of speed and cost). However, the findings we report here underscore the challenges of using AMT to understand behaviors, attitudes, and experiences that may be associated with participation in social media sites and related online activities. Overall, the AMT workers were more likely to engage in nearly every online activity we considered in this study, even after adjusting for background attributes. In terms of specific sites and activities, these relationships varied widely. Depending on the research domain, these associations could threaten the validity of generalizing findings from a sample of AMT workers to a broader population of Internet users. While techniques such as multilevel regression with poststratification may offer a path forward in such circumstances (Goel et al., 2017; Park, Gelman, & Bafumi, 2004), these techniques remain outside the mainstream in communication research and should undergo additional assessment in these research domains.

The results also illustrate an important point about the relationships between different kinds of online activities. These relationships are rarely observed because data are often collected about subsets of activities or from a single website. But relationships between online activities can impact the accuracy and validity of research findings when any website or activity is used as a sampling frame for studies that intend to draw inferences about other online behaviors or Internet users as a whole (Hargittai, 2020; Tufekci, 2014). Generalizing across sites or activities in this way is not an insoluble issue, but so long as it remains unresolved, the results may be biased in unknown ways. This remains an open area of investigation.

Conclusion

As a quick search on academic databases shows, tens of thousands of scientific articles have been written based on responses from workers on the microtask platform Amazon Mechanical Turk. The ease, speed, and low cost of access to study participants have clearly made AMT a popular platform for researchers. The quality of research partly depends on the suitability of a sample to the research questions at hand. Working with the appropriate sampling frame can be crucial for avoiding biases in one’s data. Given the severity of these issues, an entire literature has developed examining the quality of AMT data, whether regarding response quality or sample biases. This article contributes to this body of work by examining how AMT respondents compare with a national sample of participants regarding their social media platform adoption and their active online engagement. Findings show that, even when controlling for the variations in sociodemographics of the two samples, the AMT sample has more online experiences. AMT workers are more likely to be on five of the seven social media sites we examined and are more likely to engage in eight of the 10 online activities we studied. While readers might suggest reasons why some sites like Twitter or Reddit attract AMT workers more than the general population, we caution against such post-hoc explanations without a deeper understanding of the mechanisms of self-selection driving participation across any given site. At the outset of this study, we were not aware of a compelling theory or empirical basis on which to anticipate which specific sites AMT workers would use more heavily than the general Internet-using population. While the vast majority of U.S. Internet users have visited some sites (e.g., Facebook), others like Twitter or LinkedIn do not enjoy such widespread adoption. The fact that some of these variations are associated with the data sources used in this study and others are not underscores that this remains a poorly understood aspect of online behavior.

In the absence of deeper understanding of the mechanisms of selection involved in engaging with specific websites or rigorous evaluation of more sophisticated methods of statistical weighting and adjustment, we conclude that caution provides the safest route to unbiased results. When research questions concern topics, behaviors, or attitudes related to social media engagement and online participation, it may be best to avoid AMT as a sampling frame to make sure that the biases of the data set do not jeopardize the findings of the study.

References

Ahn, J. (2011). Digital divides and social network sites: Which students participate in social media? *Journal of Educational Computing Research*, *45*(2), 147–163. doi:10.2190/EC.45.2.b

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*(3), 351–368. doi:10.1093/pan/mpr057

Blank, G. (2013). Who creates content? *Information, Communication & Society*, *16*(4), 590–612. doi:10.1080/1369118X.2013.777758

Blank, G., & Lutz, C. (2017). Representativeness of social media in Great Britain: Investigating Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram. *American Behavioral Scientist*, *61*(7), 741–756. doi:10.1177/0002764217717559

boyd, d. (2012). White flight in networked publics: How race and class shaped American teen engagement with MySpace and Facebook. In L. Nakamura & P. A. Chow-White (Eds.), *Race after the Internet* (pp. 203–222). New York, NY: Routledge.

Büchi, M., Just, N., & Latzer, M. (2016). Caring is not enough: The importance of Internet skills for online privacy protection. *Information, Communication & Society*, *20*(8), 1261–1278. doi:10.1080/1369118X.2016.1229001

Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are samples drawn from Mechanical Turk valid for research on political ideology? *Research & Politics*, *2*(4), 1–9. doi:10.1177/2053168015622072

Correa, T. (2010). The participation divide among “online experts”: Experience, skills and psychological factors as predictors of college students’ Web content creation. *Journal of Computer-Mediated Communication*, *16*(1), 71–92. doi:10.1111/j.1083-6101.2010.01532.x

Couper, M. P., & Miller, P. V. (2008). Web survey methods introduction. *Public Opinion Quarterly*, *72*(5), 831–835. doi:10.1093/poq/nfn066

Crawford, K. (2009). Following you: Disciplines of listening in social media. *Continuum: Journal of Media & Cultural Studies*, *23*(4), 525–535. doi:10.1080/10304310903003270

Ertiö, T., Kukkonen, I., & Räsänen, P. (2018). Social media activities in Finland: A population-level comparison. *Convergence*, *26*(1) 193–209. doi:10.1177/1354856518780463

Gazit, T., Aharony, N., & Amichai-Hamburger, Y. (2019). Tell me who you are and I will tell you which SNS you use: SNSs participation. *Online Information Review*, *44*(1), 139–161. doi:10.1108/OIR-03-2019-0076

Goel, S., Obeng, A., & Rothschild, D. (2017). *Online, opt-in surveys: Fast and cheap, but are they accurate?* (Working Paper). Stanford, CA: Stanford University. Retrieved from https://perma.cc/G7QE-UFWA

Haight, M., Quan-Haase, A., & Corbett, B. A. (2014). Revisiting the digital divide in Canada: The impact of demographic factors on access to the Internet, level of online activity, and social networking site usage. *Information, Communication & Society*, *17*(4), 503–519. doi:10.1080/1369118X.2014.891633

Hargittai, E. (2007). Whose space? Differences among users and non-users of social network sites. *Journal of Computer-Mediated Communication*, *13*(1), 276–297. doi:10.1111/j.1083-6101.2007.00396.x

Hargittai, E. (2020). Potential biases in big data: Omitted voices on social media. *Social Science Computer Review*, *38*(1), 10–24. https://doi.org/10.1177/0894439318788322

Hargittai, E., & Hinnant, A. (2008). Digital inequality: Differences in young adults’ use of the Internet. *Communication Research*, *35*(5), 602–621. doi:10.1177/0093650208321782

Hargittai, E., & Hsieh, Y. P. (2012). Succinct survey measures of Web-use skills. *Social Science Computer Review*, *30*(1), 95–107. doi:10.1177/0894439310397146

Hargittai, E., & Jennrich, K. (2016). The online participation divide. In M. Lloyd & L. A. Friedland (Eds.), *The communication crisis in America, and how to fix it* (pp. 199–213). New York, NY: Palgrave Macmillan. https://doi.org/info:doi/10.1057/978-1-349-94925-0_13

Hargittai, E., & Litt, E. (2011). The tweet smell of celebrity success: Explaining variation in Twitter adoption among a diverse group of young adults. *New Media & Society*, *13*(5), 824–842. doi:10.1177/1461444811405805

Hargittai, E., & Shaw, A. (2020). Comparing Internet experiences and prosociality in Amazon Mechanical Turk and population-based survey samples. *Socius*, *6*, 1–11. doi:10.1177/2378023119889834

Hargittai, E., & Walejko, G. (2008). The participation divide: Content creation and sharing in the digital age. *Information, Communication & Society*, *11*(2), 239–256. doi:10.1080/13691180801946150

Hellemans, J., Willems, K., & Brengman, M. (2020). Daily active users of social network sites: Facebook, Twitter, and Instagram-use compared to general social network site use. In F. J. Martínez-López & S. D'Alessandro (Eds.), *Springer Proceedings in Business and Economics Series: Advances in digital marketing and eCommerce* (pp. 194–202). Cham, Switzerland: Springer. doi:10.1007/978-3-030-47595-6_24

Horton, J. J., Rand, D. G., & Zeckhauser, R. 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

Howard, P. E. N., Rainie, L., & Jones, S. (2001). Days and nights on the Internet: The impact of a diffusing technology. *American Behavioral Scientist*, *45*(3), 383–404. doi:10.1177/0002764201045003003

Levay, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of Mechanical Turk samples. *SAGE Open*, *6*(1), 1–17. doi:10.1177/2158244016636433

Lutz, C., & Hoffmann, C. P. (2017). The dark side of online participation: Exploring non-passive and negative participation. *Information, Communication & Society*, *20*(6), 876–897. doi:10.1080/1369118X.2017.1293129

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23. doi:10.3758/s13428-011-0124-6

Mellon, J., & Prosser, C. (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, *4*(3), 1–9. doi:10.1177/2053168017720008

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, *23*(3), 184–188. doi:10.1177/0963721414531598

Park, D. K., Gelman, A., & Bafumi, J. (2004). Bayesian multilevel estimation with poststratification: State-level estimates from national polls. *Political Analysis*, *12*(4), 375–385. doi:10.1093/pan/mph024

Redmiles, E. M., Kross, S., Pradhan, A., & Mazurek, M. L. (2017). *How well do my results generalize? Comparing security and privacy survey results from Mturk and Web panels to the U.S.* (Technical report). Retrieved from http://drum.lib.umd.edu/handle/1903/19164

Schradie, J. (2011). The digital production gap: The digital divide and Web 2.0 collide. *Poetics*, *39*(2), 145–168. doi:16/j.poetic.2011.02.003

Sheldon, P., & Bryant, K. (2016). Instagram: Motives for its use and relationship to narcissism and contextual age. *Computers in Human Behavior*, *58*, 89–97. doi:10.1016/j.chb.2015.12.059

Tufekci, Z. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. *Eighth International AAAI Conference on Weblogs and Social Media (ICWSM '14)*, 505–514. Retrieved from http://arxiv.org/abs/1403.7400

Weinberg, J., Freese, J., & McElhattan, D. (2014). Comparing data characteristics and results of an online factorial survey between a population-based and a crowdsource-recruited sample. *Sociological Science*, *1*, 292–310. doi:10.15195/v1.a19