Lie for a Dime: When Most Prescreening Responses Are Honest but Most Study Participants Are Impostors

Jesse J. Chandler[1] and Gabriele Paolacci[2]

Abstract

The Internet has enabled recruitment of large samples with specific characteristics. However, when researchers rely on participant self-report to determine eligibility, data quality depends on participant honesty. Across four studies on Amazon Mechanical Turk, we show that a substantial number of participants misrepresent theoretically relevant characteristics (e.g., demographics, product ownership) to meet eligibility criteria explicit in the studies, inferred by a previous exclusion from the study or inferred in previous experiences with similar studies. When recruiting rare populations, a large proportion of responses can be impostors. We provide recommendations about how to ensure that ineligible participants are excluded that are applicable to a wide variety of data collection efforts, which rely on self-report.

Keywords

The last few years have witnessed increasing use of the Internet for conducting psychological research. Crowdsourcing platforms such as Amazon Mechanical Turk (MTurk) are an especially appealing source of participants because they allow research studies to be conducted at a fraction of the time and cost required with more traditional participants (Gosling & Mason, 2015; Paolacci & Chandler, 2014). In 2015, more than 500 papers using MTurk data were published in social science journals with impact factors greater than 2.5 (Chandler & Shapiro, 2016), including more than 40% of papers published in the *Journal of Personality and Social Psychology* and more than 20% of papers published in *Psychological Science* (Zhou & Fishbach, 2016). The number of papers that recruit participants online from all sources is yet larger.

Compared to samples such as students enrolled in psychology classes, online convenience samples are larger and more diverse in terms of age, education level, ethnicity, and so on (see Casey, Chandler, Levine, Proctor, & Strolovitch, 2017, for a recent large demographic survey of MTurk). This allows researchers to target subpopulations with specific characteristics, which can provide various benefits. Sometimes a group is of unique theoretical or social interest. In other cases, specific groups are recruited for methodological reasons such as the possibility to match manipulations or measurements to a population's specific experiences (e.g., Taylor, Lichtman, & Wood, 1984), improve the external validity of research (e.g., Gneezy & Imas, in press), or reduce theoretically irrelevant variance (e.g., by restricting handedness; Hamerman & Johar, 2013).

Researchers have begun to take advantage of the diversity of Internet users to target specific samples. Focusing only on MTurk, researchers have recruited participants of specific ages (Connell, Brucks, & Nielsen, 2014), races (Brown & Segrist, 2016), religions (Fergus & Rowatt, 2015), employment status (Konstam, Tomek, Celen-Demirtas, & Sweeny, 2015), immigrant status (Bernal, 2014), veteran status (Lynn, 2014), weight (Pearl, Puhl, & Dovidio, 2015), and sexual orientation (Zou, Anderson, & Blosnich, 2013). Other researchers have recruited people with specific life experiences such as pregnancy (Arch, 2014), fatherhood (Parent, McKee, Rough, & Forehand, 2016; Schleider & Weisz, 2015), bereavement (Papa, Lancaster, & Kahler, 2014), and prior tobacco use (Cougle, Hawkins, Macatee, Zvolensky, & Sarawgi, 2014; Johnson, Herrmann, & Johnson, 2015). Clinical researchers have recruited people with specific psychopathological symptoms (e.g., depression and anxiety; Reece & Danforth, 2016; Yang, Friedman-Wheeler, & Pronin, 2014) and medical conditions (e.g., cancer; Arch & Carr, 2016).

[1] Mathematica Policy Research, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA
[2] Rotterdam School of Management, Erasmus University Rotterdam, Rotterdam, the Netherlands

Corresponding Author:
Jesse J. Chandler, Mathematical Policy Research, Institute for Social Research, University of Michigan, 220 E Huron St, Suite 300, Ann Arbor, MI 48104, USA.
Email: jchandler@mathematica-mpr.com

A major challenge of recruiting specific subpopulations is that eligibility usually relies on participant self-report, and claiming eligibility is rewarded through compensation for completing the study. This is particularly true when recruiting samples online because the truthfulness of responses can be difficult to verify. Some survey platforms such as MTurk offer a limited set of prebuilt screening criteria. Researchers using other recruitment methods or selecting participants based on less widely used criteria must prescreen their own participants. Current prescreening practices vary widely, with some researchers simply asking ineligible people not to participate. Others use more sophisticated methods to limit participation but without validating the extent to which these efforts prevent workers from reattempting the survey. In this article, we investigate the extent to which researchers can rely on self-reported eligibility when they recruit specific samples.

Research on data collected from online samples provides suggestive evidence that people fraudulently gain access to research studies. Online research panels often include unusually large proportions of participants who claim membership in rare categories (Jones, House, & Gao, 2015; Miller, 2006). In perhaps the starkest illustration, one study found that 14% of survey participants claimed to own a Segway human transporter (Downes-Le Guin, Meechling, & Baker, 2006). Providing more direct evidence of fraud, a study of medical research participants recruited using newspapers and Craigslist found that 14% of survey participants admitted to fabricating a health condition to gain eligibility to a clinical trial, with "professional" research participants particularly likely to engage in fraudulent behavior (Devine et al., 2013).

It is hard to anticipate the degree to which MTurk studies conducted on particular subsamples may be threatened by impostors. All available evidence suggests that workers are no more dishonest than other people when completing experimental tasks (Beramendi, Dutch, & Matsuo, 2014; Cavanagh, 2014; Farrell, Grenier, & Leiby, in press). However, people (including workers) are not immune from temptations to cheat when doing so offers a monetary reward (e.g., Goodman, Cryder, & Cheema, 2013; Suri, Goldstein, & Mason, 2011). Because workers (and survey panelists more generally) wish there was more work available to them (Berg, 2016), they may be motivated to misrepresent their identity to gain access to more research studies.

Importantly, the proportion of impostors in a particular sample depends not only on the proportion of fraudulent participants in the population but also on the base rate of truly eligible sample members. As a simple example, if 5% of the population is willing to fraudulently gain access to a study that targets a group of only 5% of the population, about half of the final sample will consist of impostors (for a related discussion, see Casscells, Schoenberger, & Grayboys, 1978). Consequently, the degree to which fraud is a problem depends heavily on the prevalence of the population of interest to the researcher, and even seemingly negligible rates of fraudulent behavior can substantially increase the number of impostors and sampling error.

In sum, the extent and conditions under which researchers can rely on self-reported study eligibility is unknown and pressing to answer. Across fours studies, we show that MTurk workers lie to gain admission to studies when they become aware of prescreening requirements through reading explicit inclusion criteria (Studies 1–2) or through being excluded due to ineligibility (Studies 3–4a) or through prior exposure to a study with similar inclusion criteria (Study 4b). Fraudulent participants manage to complete surveys even when commonly used countermeasures are employed (Study 4a). Fraud rates may be particularly high when studies are more lucrative (Study 3) but seem to be independent of workers' experience completing research studies (Studies 1–2) and past quality of work (Study 4a). After reporting this evidence on the unreliability of self-reported eligibility, we discuss solutions for researchers to ensure their crowdsourced samples match the desired characteristics.

Study 1

Method

Unless specified otherwise, across all studies we recruited workers who had completed at least 100 MTurk tasks with a 95% or greater ratio of approved/submitted tasks (following Peer, Vosgerau, & Acquisti, 2014). Workers were compensated with US$0.10 per estimated minute of participation following recommended best practices (Chandler & Shapiro, 2016).

In Study 1, 2,397 workers who had completed a previous study answered three questions about state education testing. They were then asked whether they were parent or guardian of a child with autism. Crucially, participants were randomly assigned to either an explicit prescreening or a control condition. Participants in the explicit prescreening condition were first told that we were trying to determine participant eligibility for another study. Reports of a child with autism were treated as potentially fraudulent.

As an additional factor, the impact of workers' experience completing research studies on their propensity to engage in fraud was investigated. Worker experience was estimated by summing the total number of Human Intelligence Tasks (HITs) that each worker completed within a large sample of researcher-posted HITs collected in prior years (data taken from Stewart et al., 2015).

Results

Participants were more likely to indicate that they had a child with autism in the prescreening condition (7.8%; 93/1,196) than in the control condition (4.3%; 52/1,201), $B = 0.67$, 95% confidence interval (CI) = [.29, 1.04], Wald $\chi^2 = 12.74$, $p < .001$, $d = 0.15$. There were no other main effects or interactions, $ps > .21$.

Discussion

About 3.5% of the participants provided a potentially fraudulent response. However, fraudulent participants would have had a substantial impact had this been an effort to explicitly recruit parents and guardians of autistic children: Due to the rarity of autism, 45% of the self-identified eligible participants in the explicit prescreening condition are potentially fraudulent. There was no evidence that more experienced workers are more likely to engage in fraudulent behavior.

There are two limitations to this study. First, the deception is relatively mild and technically not fraudulent: Participants merely indicate interest in a future study and do not provide data that they believe will actually be used for research. Second, the truthfulness of our target self-report is not observable. It is possible that participants in the control condition underreport levels of autism, and the pretext for the question in the experimental condition induced participants to be more honest. We address these issues in Study 2.

Study 2

Study 2 compares the proportion of workers who change their self-reported sexual orientation when it is or is not explicitly required to fulfill study inclusion criteria. A third condition examines whether merely asking about sexual orientation at the beginning of the survey suggests to participants that researchers are recruiting members of a specific rare category, mimicking the use of prescreening questions at the beginning of a survey to identify and exclude ineligible participants.

Method

MTurk workers who identified as heterosexual ($N = 324$) in an earlier survey (Casey et al., 2017) completed a "Personality Study." The blatant prescreening condition was displayed to a further 24 participants who exited the survey without providing data, perhaps because they believed that they were ineligible to complete it. Sample size is discussed in the preregistration of this study (osf.io/nprxs). Participants were randomly assigned to one of three conditions. In the control condition, participants reported their sexual orientation at the end of the survey. In the blatant prescreening condition, participants reported their sexual orientation at the beginning of the survey after being told that only lesbian, gay, or bisexual (LGB) people were eligible to participate. In the subtle prescreening condition, participants reported their sexual orientation at the beginning of the survey. Reports of LGB sexual orientation that were inconsistent with previously reported sexual orientation were treated as potentially fraudulent.

After completing a filler questionnaire, participants were asked to check a box if they felt that their data should not be used for any reason. They were told that checking the box would not affect their payment. The impact of worker

experience on fraud was investigated in the same manner as in Study 1.

Results

Using binary logistic regression, fraudulent behavior was regressed on condition (Block 1: dummy coded), worker experience (Block 2: mean centered), and the interactions between each of the prescreening conditions and worker experience (Block 3). Dummy codes were assigned to the subtle and blatant prescreening conditions to compare the effect of worker experience in the two experimental conditions to the effect of worker experience in the control condition.

In the control condition, 3.8% (4 of 104) of participants identified as LGB. In the subtle prescreening condition, 3.5% (4 of 114) of participants identified as LGB. In the blatant prescreening condition, 45.3% (48 of 106) of participants identified as LGB, significantly more than the other 2 conditions, odds ratio = 0.08, 95% CI [0.03, 0.19], Wald $\chi^2 = 31/42$, $p < .001$. There were no other main effects or interactions, all ps >.27. Only one participant (in the subtle prescreening condition) indicated that their data should be excluded from the analysis.

Discussion

When prescreening criteria were explicit, almost half of the heterosexual participants who completed the survey misrepresented their sexual orientation in an attempt to meet qualifications. Including those (presumably honest participants) who exited the survey, 36.9% of participants misrepresented themselves. We found no evidence of fraudulent participants in the subtle prescreening condition, suggesting that participants do not assume that questions at the beginning of a survey will affect survey eligibility. Again, worker experience did not predict fraudulent responses. Notably, virtually none of the fraudulent participants indicated that their data should be discarded, despite being told their response would not affect payment.

It is possible that participants did not deliberately lie but rather were motivated to select a particular definition of sexual orientation, which allowed them to meet study criteria. Studies 1 and 2 are also limited in that they focus on explicit prescreening. While some researchers explicitly list prescreening criteria, others use a strategy closer to that in the subtle prescreening condition and terminate ineligible responses. Study 3 addresses these limitations by using a different prescreening question and exclusion method.

Study 3

Study 3 examines whether participants who are screened out of a study due to ineligibility will reattempt it. As a secondary question, this study examines whether higher payments induce more fraud.

Table 1. Number of Honest and Fraudulent Men and Women in Low- and High-Paying Surveys (Study 3).

	Low Pay		High Pay	
	Honest	Fraudulent	Honest	Fraudulent
Men	116	13	206	53
Women	147	3	256	34
Total	263	16	462	87

Note. Low-pay participants were paid US$0.05 per minute and high-pay participants were paid US$0.20 per minute. Fraudulent participants are those who initially reported a biological sex consistent with an earlier survey and inconsistent with study eligibility criteria and then reattempted the survey and reported a different biological sex.

Participants

MTurk workers ($N = 828$) who previously reported their biological sex were recruited for a study described as lasting 5 min and were paid either US$0.25 or US$1.00 for their time. Sample size was smaller than specified in the preregistration (osf.io/vwmza) due to difficulties recruiting participants in the low-pay condition (see also Buhrmester, Kwang, & Gosling, 2011; Mason & Watts, 2010).

Method

Participants were assigned to one of the four surveys in a 2 (worker sex: male vs. female) × 2 (pay: low, i.e., US$0.25 vs. high, i.e., US$1.00) design. Workers were assigned a qualification so they could only see the survey assigned to them (for technical details, see Chandler, Mueller, & Paolacci, 2014). Worker sex was determined by responses to the survey by Casey, Chandler, Levine, Proctor, and Strolovitch (2017). Pay was randomly assigned.

Participants read a consent form and indicated whether they agreed to participate in a study about "personality." Agreeing to participate generated an observation that included the participant's MTurk WorkerID (for technical details, see Peer, Paolacci, Chandler, & Mueller, 2012). When participants reported their true sex, they received a message telling them that no more participants of their sex were required and were terminated from the survey. Participants who reattempted the survey produced a second observation, allowing multiple submissions from the same participant to be identified and linked together. After the survey, participants reported whether their data should be discarded as in Study 2. Fraud was defined as reporting a sex consistent with that provided in the previous survey and then returning to the survey and reporting a different sex.

Results

Results (Table 1) were analyzed using a generalized linear model with pay, gender, and their interaction as predictors. Participants were more likely to reattempt the survey in the high-pay condition (15.8%) than in the low-pay condition (5.7%), $B = 0.83$, 95% CI [0.18, 1.48], Wald $\chi^2 = 6.32$, $p = .02$,

$d = 0.18$. There was also a main effect of sex, $B = 0.66$, 95% CI [0.19, 1.13], Wald $\chi^2 = 7.67$, $p = .01$, $d = 0.19$, reflecting that 8.4% of women and 17.0% of men were fraudulent. The interaction between pay and sex was not significant, $B = 1.04$, 95% CI [−0.32, 2.40], Wald $\chi^2 = 1.92$, $p = .17$. Four participants (all of whom provided fraudulent responses) indicated that their data should be excluded from the analysis.

Discussion

Most participants honestly abandoned the study after being told that they were ineligible. However, a small proportion of participants reattempted the survey and modified their responses to meet inclusion criteria. Fraud was more prevalent when compensation was higher.

The impact of fraudulent participants on data quality varied as a function of both pay and the distribution of gender in the workforce. In the best case, when paying "males" US$0.05 per minute, 263 workers would have attempted the survey, 129 of which would be true men. Of the 147 women, 3 would have lied about their gender, leading to a 2.3% (3 of 132) fraud rate. In the worst case, when paying "females" US$0.20 cents per minute, 462 workers would have attempted the survey, 290 of which would be true women. Of the 259 men, 53 would have lied about their gender, leading to a 15.5% (53 of 343) fraud rate.

Study 4a

Study 4a examines whether participants can defeat a common method of preventing duplicate responses (a cookie placed in the web browser cache). Cookies will prevent some people from reattempting the survey but do not work on web browsers configured to block them and can be thwarted by deleting them or by retaking the survey using a different browser or device. As a secondary question, this study examines whether workers with a history of lower quality work are more or less likely to engage in fraud than the high-quality samples typically recommended to researchers (Peer et al., 2014).

Method

In this study, 645 MTurk participants were randomly assigned to one of the four conditions in a 2 (Worker quality: high vs. low) × 2 (Design: allows vs. prevents duplicates) design. Sample size is discussed in the preregistration of this study (osf.io/bekpj). MTurk measures worker quality by calculating the proportion of tasks a worker has submitted that have been approved by the people who requested them (the HIT Approval Ratio). Individual worker scores are not disclosed by MTurk, but workers who fall within specific percentage ranges can be selectively recruited. For this reason, worker quality was dichotomized as above and below 95%, which is a recommended quality threshold among requesters (Peer et al., 2014). Cookies were either disabled (the default setting in Qualtrics) or enabled using the "Prevent Ballot Box Stuffing"

Table 2. Number of Honest and Fraudulent Low- and High-Quality Workers When Reattempting the Survey Is Allowed or Prevented (Study 4a).

	Allow Reattempt		Prevent Reattempt	
	Honest	Fraudulent	Honest	Fraudulent
Low quality	25 (1)	4	53 (5)	0
High quality	149 (4)	34	367 (13)	13
Total	174 (5)	38	420 (18)	13

Note. Low (high)-quality participants had less than (at least) 95% of previously submitted work approved. Honest participants are those who only completed the survey once. Honest participants who said they owned an Oculus Rift headset (OR) are included in this total and indicated in parenthesis above. Fraudulent participants are those who initially reported that they did not own an OR and then reattempted the survey and reported owning one. Reattempts were prevented using Qualtrics' "Prevent Ballot Box Stuffing" feature.

Table 3. Number of Honest and Fraudulent Workers Across Repeated Surveys (Study 4b).

	A Week Later		A Month Later	
	Honest	Fraudulent	Honest	Fraudulent
Does not meet criteria	167	0	149	0
Met criteria	4	32	3	10

Note. Participants met criteria if they claimed to own an Oculus Rift headset in Study 4b. Honest respondents are those whose first response was consistent with their first response in Study 4a and fraudulent respondents are those whose first response was inconsistent with their first response in Study 4a.

option. To prevent workers from seeing both designs, the design conditions were posted sequentially, and workers who attempted to complete a HIT were excluded from subsequently posted HITs.

As in Study 3, participants read a consent form and indicated whether they agreed to participate in the study. They then indicated the items they owned in a list of personal electronics. Critically, the list included an Oculus Rift VR headset (OR), a niche product that had been on the market for only 4 months at the time the study was conducted. Participants who did not indicate that they owned an OR were told "Thanks for your willingness to participate. Unfortunately, you do not meet the screening criteria for this study." Fraud was defined as first claiming not to own an OR and then reattempting the survey and claiming to own one. After completing a filler questionnaire, participants reported whether their data should be discarded as in Study 2.

Results

Results (Table 2) were analyzed using GZLM with worker quality and design as predictors. The interaction was dropped from the model because it created a quasi-complete separation in the data. Participants were more likely to reattempt the survey when duplicates were allowed (17.9%, 38 of 212) than when duplicates were prevented (3.0%, 13 of 433), $B = 1.97$,

95% CI [1.31, 2.62], Wald $\chi^2 = 34.61$, $p = .001$, $d = 0.48$. Worker quality did not influence fraud rates, but if anything, low-quality workers were less fraudulent (4.9%, 4 of 82) than high-quality workers (8.3%, 47 of 563), $B = 0.67$, 95% CI [−0.40, 1.74], Wald $\chi^2 = 1.50$, $p = .22$. Only four participants (two of whom were fraudulent) indicated that their data should be excluded from the analysis.

Discussion

Most participants were honest, but 17.9% of participants reattempted the survey when no effort was made to prevent them from doing so. This was reduced to 3.0% when multiple responses were prevented by placing a cookie on participants' computers. Again, data quality is a function of the frequency of both honest and fraudulent responses. In the high-quality sample that prevented participants from reattempting the survey, 13 participants reported owning an OR on their first attempt and 13 participants were fraudulent, resulting in a fraud rate of 50% (Table 2).

Study 4b

Study 4b examines whether participants apply lessons learned about specific prescreening criteria to a subsequent study. This study was not preregistered.

Method

A week after Study 4a (referred to as T1 in the present study), a second identical survey (referred to as T2) was made available for 12 hr to all participants who had attempted it. Two hundred three unique participants (referred to as the week later group) were recorded. Five weeks later, the T2 survey was made available for 2 weeks to participants who were not in the week later group. During this period, 162 unique participants (referred to for convenience as the month later group) were recorded. Fraud was defined as claiming not to own an OR in the first attempt of the survey at T1 and then claiming to own one in the first attempt of the survey at T2.

Results

Seven participants (four in the week later group and three in the month later group) reported that they owned an OR in their first attempt of T1 and again in T2, and these are presumed to be honest responses. No participants claimed to own an OR in their first attempt of T1 and then denied owning an OR in their first attempt of T2. At T2, in the week later group, an additional 32 participants reported owning an OR on their very first survey attempt, a significant increase, McNemar Test $\chi^2(1, n = 203) = 30.03$, $p < .001$, $d = 0.83$. In the month later group, an additional 10 participants indicated that they owned an OR on their first survey attempt, a significant increase, McNemar Test $\chi^2(1, n = 162) = 8.10$, $p < .01$, $d = 0.46$. There were significantly fewer fraudulent responses in the month later condition than in the week later condition, $\chi^2(1, N = 365) = 7.22$,

$p < .01$, $d = 0.28$, suggesting that people not only learn but also forget about screening criteria over time (Table 3).

When conducting a series of related studies, a researcher might prevent those who completed the first study from attempting subsequent studies (Chandler et al., 2014). However, doing so does not prevent those who merely attempted the first study from reattempting subsequent studies, and these individuals might have also gained important knowledge about how to defeat prescreening measures. To illustrate, among participants who attempted but did not complete the survey at T1, 11% (20 of 178) provided a fraudulent response on their first survey attempt in the week later group, and 3.4% (5 of 145) provided a fraudulent on their first survey attempt in the month later group. Again, the proportion of participants who provided a fraudulent response on their first attempt of the survey was higher in the week later group than in the month later group, $\chi^2(1, n = 323) = 6.79$, $p < .01$, $d = 0.28$.

Discussion

Although earlier studies found that more experienced workers are no more likely to defeat prescreening measures, two pieces of evidence suggest that repeated exposure to the same prescreening question increases the likelihood of fraudulent responses to it. First, the proportion of fraudulent responses was higher at T2 than at T1, even when excluding participants who successfully completed the survey at T1. Second, a substantial proportion of these participants indicated that they owned an OR on their first attempt at T2. The fact that more participants reported owning an OR in the week later group than the month later group is consistent with the hypothesis that these responses reflect a memory of being excluded at T1 (and their subsequent forgetting, see also Chandler, Paolacci, Peer, Mueller, & Ratliff, 2015) rather than large gains in OR market penetration over the data collection period.

General Discussion

Prior research suggested that participants might misrepresent themselves to gain access to research studies, based on the increased incidence rates of self-reported rare events (Downes-Le Guin et al., 2006; Jones et al., 2015; Miller, 2006). This article extends this work by providing direct evidence of fraudulent participants based on the consistency of participant self-reports with previously collected information. Our studies used MTurk participants as an illustrative example, but there is every reason to believe that these issues extend to all studies that selectively recruit paid research participants on the basis of self-report.

Participants will deceive researchers if they believe it is necessary to gain access to a study (Studies 1 and 2) and particularly when the reward is high (Study 3), but the necessity of fraud can become apparent in a variety of ways, including explicitly soliciting specific participants (Study 1), terminating ineligible responses (Studies 3 and 4a), and through exposure to previous studies with similar prescreening criteria (Study

4b). Fortunately, participants do not seem to have lay theories that make fraudulent responses difficult to detect: They are not predisposed to falsely identify as belonging to rare categories in early demographic questions (Study 2) and do not select all possible responses in early "select all that apply" questions to the extent sometimes observed in commercial market research panels (Study 4a; see Downes-Le Guin et al., 2006).

Future research would benefit from understanding which kinds of questions are particularly vulnerable to fraudulent responses. For example, participants might be more willing to deceive researchers about more subjective or less observable attributes or more willing to deceive in ways that are socially or personally desirable than in ways that are undesirable. Also, it is possible that participants will also draw on other sources of information (e.g., consent forms or author affiliation) to make inferences about eligibility requirements. In one particularly concerning example, Devine et al. (2013) found that people look up eligibility criteria for lucrative clinical trials using study preregistration documents and belonged to informal networks that shared this information with others.

The inclusion of participants who do not truly meet prescreening criteria is problematic. Fraudulent responses create an obvious validity problem and may lead to erroneous conclusions about the population of interest (for illustrations, see Siegel, Navarro, & Thomson, 2015; Wessling, Huber, & Netzer, in press). Fraudulent responses can conceal true relationships between variables if participants who provide fraudulent prescreening responses are truthful when reporting other information. In other cases, fraudulent participants could create artifactual relationships between variables if their responses systematically differ from those of truthful participants—a possibility that future research should explore. For instance, participants who lie about their identity (e.g., gender) to get access to a study might complete the survey according to their assumptions of how someone truly eligible to complete the study might respond. The potential to introduce systematic bias makes fraudulent participants a potentially greater threat to the integrity of data than other kinds of problem participants that have been more exhaustively studied (e.g., inattentive participants; Peer et al., 2014) because systematic bias cannot be overcome by simply increasing sample size.

Researchers can use several techniques to minimize participant fraud. Ideally, screening data would be collected as a part of a stand-alone prescreening study collected either by a sample provider or by the researcher. Eligible participants can be immediately routed to a second survey or be contacted at a future date to complete studies that they qualify for (Springer, Martini, Lindsey, & Vezich, 2016). Alternatively, data can be collected in a single study, with participants who do not possess the desired prescreening characteristics routed to other studies without these restrictions. If prescreening or sample routing are not practical, fraudulent participants can at least be minimized using existing methods to discourage duplicate participants (e.g., Qualtrics' "Prevent Ballot Box Stuffing" option). In Study 4a, this approach reduced the rate of fraudulent responding by 80%. Impostors can also be detected by examining the

Internet Protocol (IP) addresses of people who attempt the survey or, on MTurk, through associating responses to a unique participant identifier. Asking participants to self-identify as providing poor-quality data was consistently ineffective in our studies.

Our findings also suggest that when posting studies that include prescreening criteria, researchers should carefully consider whether participants should be naive not only to study contents but also to eligibility criteria. In some cases, it may be best for researchers to prevent not only participants who have completed related studies but also participants who have attempted to complete related studies from attempting later studies that use the same prescreening criteria. Doing so requires collecting the WorkerIDs of people who attempt the survey (not just those who complete it) and then excluding these workers using TurkPrime (Litman, Robinson, & Abberbock, 2016). Moreover, as has occasionally been noted with popular research paradigms (Chandler et al., 2014, 2015; Rand et al., 2014), it is possible that in some contexts, certain prescreening criteria could be so common that participants will learn to habitually lie (for an example of this in consumer research panels, see Downes-Le Guin et al., 2006).

Importantly, the impact of fraudulent responses on a data quality is determined as much by the rarity of the group of interest, as the overall prevalence of fraud and fraud rates can be intolerably high even when recruiting moderately rare groups. To illustrate, about 18% of participants in Study 4a responded fraudulently when no steps were taken to prevent them from doing so. When recruiting a moderately rare population (e.g., the 11% of participants who identify as LGB; Casey et al., 2017), this will produce a fraud rate of about 62% (18 of 29). Even when countermeasures are added, 3% of participants responded fraudulently, producing a fraud rate of about 21% (3 of 14), which is better but in some cases but may still be unacceptable. Thus, when rare populations must be recruited, researchers need to be especially vigilant for fraud and make every effort to identify impostors. Fortunately, the tools with which to do so exist and are compatible with commonly used survey software and sample recruitment platforms.

Declaration of Conflicting Interests

Funding

References

Arch, J. J. (2014). Cognitive behavioral therapy and pharmacotherapy for anxiety: Treatment preferences and credibility among pregnant and non-pregnant women. *Behaviour Research and Therapy*, *52*, 53–60.

Arch, J. J., & Carr, A. L. (2016). Using mechanical Turk for research on cancer survivors. *Psycho-Oncology*. doi:10.1002/pon.4173

Beramendi, P., Duch, R. M., & Matsuo, A. (2014). *When lab subjects meet real people: Comparing different modes of experiments*. Paper presented at Asian Political Methodology Conference, Taipai, Taiwan. Retrieved from https://cess-web.nuff.ox.ac.uk/files/pdfs/working_papers/CESS_DP2014_003.pdf

Berg, J. (2016). Income security in the on-demand economy: Findings and policy lessons from a survey of crowdworkers. *Comparative Labor Law & Policy Journal*, *37*. Retrieved from SSRN: https://ssrn.com/abstract=2740940

Bernal, D. R. (2014). *Acculturation, acculturative stress, social status, and well-being among English language proficient immigrants* (Doctoral dissertation). University of Miami. Retrieved from http://scholarlyrepository.miami.edu/oa_dissertations/1328/

Brown, D. L., & Segrist, D. (2016). African American career aspirations examining the relative influence of internalized racism. *Journal of Career Development*, *43*, 177–189.

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*, 3–5.

Casey, L., Chandler, J., Levine, A. S., Proctor, A., & Strolovitch, D. (2017). *Demographic characteristics of a large sample of US workers*. Retrieved from https://osf.io/preprints/psyarxiv/8352x/

Casscells, W., Schoenberger, A., & Graboys, T. B. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, *299*, 999–1001.

Cavanagh, T. M. (2014). *Cheating on online assessment tests: Prevalence and impact on validity* (Unpublished doctoral dissertation). Colorado State University. Retrieved from https://dspace.library.colostate.edu/handle/10217/83718

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*, 112–130.

Chandler, J., Paolacci, G., Peer, E., Mueller, P., & Ratliff, K. A. (2015). Using nonnaive participants can reduce effect sizes. *Psychological Science*, *26*, 1131–1139.

Chandler, J., & Shapiro, D. (2016). Conducting clinical research using crowdsourced convenience samples. *Clinical Psychology*, *12*, 53–81.

Connell, P. M., Brucks, M., & Nielsen, J. H. (2014). How childhood advertising exposure can create biased product evaluations that persist into adulthood. *Journal of Consumer Research*, *41*, 119–134.

Cougle, J. R., Hawkins, K. A., Macatee, R. J., Zvolensky, M. J., & Sarawgi, S. (2014). Multiple facets of problematic anger in regular smokers: Exploring associations with smoking motives and cessation difficulties. *Nicotine & Tobacco Research*, *16*, 881–885.

Devine, E. G., Waters, M. E., Putnam, M., Surprise, C., O'Malley, K., Richambault, C., . . . Ciraulo, D. A. (2013). Concealment and

fabrication by experienced research subjects. *Clinical Trials*, *10*, 935–948.

Downes-Le Guin, T., Mechling, J., & Baker, R. (2006). Great results from ambiguous sources: Cleaning Internet panel data. In *ESOMAR World Research Gonference: Panel Research 2006*. Amsterdam, The Netherlands: ESOMAR.

Farrell, A. M., Grenier, J. H., & Leiby, J. (2017). Scoundrels or stars? Theory and evidence on the quality of workers in online labor markets. *The Accounting Review*, *92*, 93–114.

Fergus, T. A., & Rowatt, W. C. (2015). Uncertainty, god, and scrupulosity: Uncertainty salience and priming God concepts interact to cause greater fears of sin. *Journal of Behavior Therapy and Experimental Psychiatry*, *46*, 93–98.

Gneezy, U., & Imas, A. (in press). Lab in the field: Measuring preferences in the wild. In A. Banerjee & E. Duflo (Eds.), *Handbook of field experiments*. doi:10.1016/bs.hefe.2016.08.003

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, *26*, 213–224.

Gosling, S. D., & Mason, W. (2015). Internet research in psychology. *Annual Review of Psychology*, *66*, 877–902.

Hamerman, E. J., & Johar, G. V. (2013). Conditioned superstition: Desire for control and consumer brand preferences. *Journal of Consumer Research*, *40*, 428–443.

Johnson, P. S., Herrmann, E. S., & Johnson, M. W. (2015). Opportunity costs of reward delays and the discounting of hypothetical money and cigarettes. *Journal of the Experimental Analysis of Behavior*, *103*, 87–107.

Jones, M. S., House, L. A., & Gao, Z. (2015). Respondent screening and revealed preference axioms: Testing quarantining methods for enhanced data quality in Web panel surveys. *Public Opinion Quarterly*, *79*, 687–709.

Konstam, V., Tomek, S., Celen-Demirtas, S., & Sweeney, K. (2015). Volunteering and reemployment status in unemployed emerging adults a time-worthy investment? *Journal of Career Assessment*, *23*, 152–165.

Litman, L., Robinson, J., & Abberbock, T. (2016). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*. doi:10.3758/s13428-016-0727-z.

Lynn, B. M. D. (2014). *Shared sense of purpose and well-being among veterans and non-veterans* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3690322)

Mason, W., & Watts, D. J. (2010). Financial incentives and the performance of crowds. *ACM SigKDD Explorations Newsletter*, *11*, 100–108.

Miller, J. (2006). *Research reveals alarming incidence of 'undesirable' online panelists* (Research Conference Report synthesized from Presentation at Research Industry Summit: Improving Respondent Cooperation, Chicago, IL). Skokie, IL: RFL Communications. Retrieved from http://www.burke.com/Library/Articles/Jeff%20Miller%20RCR%20PDF.pdf

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

Papa, A., Lancaster, N. G., & Kahler, J. (2014). Commonalities in grief responding across bereavement and non-bereavement losses. *Journal of Affective Disorders*, *161*, 136–143.

Parent, J., McKee, L. G., Rough, J. N., & Forehand, R. (2016). The association of parent mindfulness with parenting and youth psychopathology across three developmental stages. *Journal of Abnormal Child Psychology*, *44*, 191–202.

Pearl, R. L., Puhl, R. M., & Dovidio, J. F. (2015). Differential effects of weight bias experiences and internalization on exercise among women with overweight and obesity. *Journal of Health Psychology*, *20*, 1626–1632.

Peer, E., Paolacci, G., Chandler, J., & Mueller, P. (2012). *Selectively recruiting participants from Amazon Mechanical Turk using Qualtrics*. Retrieved from SSRN 21006.

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, *46*, 1023–1031.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature communications*, *5*, Article 3677.

Reece, A. G., & Danforth, C. M. (2016, August 10). Instagram photos reveal predictive markers of depression. Retrieved March 3, 2017, from the arXiv database. arXiv:1608.03282v2 [cs.SI].

Schleider, J. L., & Weisz, J. R. (2015). Using Mechanical Turk to study family processes and youth mental health: A test of feasibility. *Journal of Child and Family Studies*, *24*, 3235–3246.

Siegel, J. T., Navarro, M. A., & Thomson, A. L. (2015). The impact of overtly listing eligibility requirements on MTurk: An investigation involving organ donation, recruitment scripts, and feelings of elevation. *Social Science & Medicine*, *142*, 256–260.

Springer, V., Martini, P., Lindsey, S., & Vezich, I. (2016). Practice-based considerations for using multi-stage survey design to reach special populations on Amazon's Mechanical Turk. *Survey Practice*, *9*(5), 1–8.

Stewart, N., Ungemach, C., Harris, A. J., Bartels, D. M., Newell, B. R., Paolacci, G., & Chandler, J. (2015). The average laboratory samples a population of 7,300 Amazon Mechanical Turk workers. *Judgment and Decision Making*, *10*, 479–491.

Suri, S., Goldstein, D. G., & Mason, W. A. (2011). *Honesty in an online labor market*. Menlo Park, CA: Association for the advancement of Artificial Intelligence.

Taylor, S. E., Lichtman, R. R., & Wood, J. V. (1984). Attributions, beliefs about control, and adjustment to breast cancer. *Journal of Personality and Social Psychology*, *46*, 489–502.

Wessling, K. S., Huber, J., & Netzer, O. (in press). Character misrepresentation by Amazon Turk workers: Assessment and solutions. *Journal of Consumer Research*.

Yang, K., Friedman-Wheeler, D. G., & Pronin, E. (2014). Thought acceleration boosts positive mood among individuals with minimal to moderate depressive symptoms. *Cognitive Therapy and Research*, *38*, 261–269.

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology*. Advance online publication. doi:10.1037/pspa0000056

Zou, C., Andersen, J. P., & Blosnich, J. R. (2013). The association between bullying and physical health among gay, lesbian, and bisexual individuals. *Journal of the American Psychiatric Nurses Association*, *19*, 356–365.

Author Biographies

Jesse J. Chandler received his PhD from the University of Michigan. He is a researcher at Mathematica Policy Research and adjunct faculty at the Institute for Social Research. He is interested in survey methodology, online research studies, decision-making, and human computation.

Gabriele Paolacci received is PhD from Ca' Foscari University of Venice, Italy. He is an assistant professor of marketing at the Rotterdam School of Management, Erasmus University. He studies consumer decision-making and the methodology of data collection in behavioral research.

Handling Editor: Dominique Muller