

Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants

David J. Hauser · Norbert Schwarz

© Psychonomic Society, Inc. 2015

Abstract Participant attentiveness is a concern for many researchers using Amazon's Mechanical Turk (MTurk). Although studies comparing the attentiveness of participants on MTurk versus traditional subject pool samples have provided mixed support for this concern, attention check questions and other methods of ensuring participant attention have become prolific in MTurk studies. Because MTurk is a population that *learns*, we hypothesized that MTurkers would be more attentive to instructions than are traditional subject pool samples. In three online studies, participants from MTurk and collegiate populations participated in a task that included a measure of attentiveness to instructions (an instructional manipulation check: IMC). In all studies, MTurkers were more attentive to the instructions than were college students, even on novel IMCs (Studies 2 and 3), and MTurkers showed larger effects in response to a minute text manipulation. These results have implications for the sustainable use of MTurk samples for social science research and for the conclusions drawn from research with MTurk and college subject pool samples.

Keywords Instructional manipulation checks · Participant attentiveness · MTurk · College students

With the increasing use of Amazon's Mechanical Turk (MTurk) workers in social science research (for reviews, see Mason & Suri, 2012; Paolacci & Chandler, 2014), participant

attentiveness has received considerable attention. Although typical undergraduate subject populations are often motivated to participate in studies because of an interest in psychology, MTurk participants are unsupervised and anonymous, complete surveys in unknown locations, and are motivated by financial incentives. Because of these differences, researchers often worry that MTurk participants are inattentive to instructions and provide poor-quality data (Chandler, Mueller, & Paolacci, 2014).

Indeed, some research suggests that MTurk participants are less attentive to instructions than participants from traditional subject pools. MTurk participants had lower pass rates on instructional manipulation checks (IMCs—i.e., trick questions designed to assess participants' attention to instructions; Oppenheimer, Meyvis, & Davidenko, 2009) than did supervised college participants in one experiment (Goodman, Cryder, & Cheema, 2013, Study 2). Furthermore, MTurkers may have issues with fully reading instructions (Crump, McDonnell, & Gureckis, 2013, Studies 8, 9, and 10; Kapelner & Chandler, 2010), and MTurkers have self-reported engaging in distractions such as cell phones (Clifford & Jerit, 2014) and multitasking while completing surveys (Chandler et al., 2014). However, other studies have suggested that MTurkers are just as attentive to instructions as traditional subject pools. For instance, MTurk participants' performance on an IMC did not differ from that of a well-paid supervised community sample in another experiment (Goodman et al., 2013, Study 1). Furthermore, MTurk participants pass attention check questions at rates similar to those of college samples and Internet forum samples (Paolacci, Chandler, & Ipeirotis, 2010), and pass factual manipulation check questions at higher rates than do other Internet samples (Berinsky, Huber, & Lenz, 2012).

Finally, the data from a recent collaborative psychology study bear on this issue. Klein et al. (2014) used Oppenheimer et al.'s (2009) IMC in a controlled replication

D. J. Hauser (✉)
Department of Psychology, University of Michigan, 3233 East Hall,
530 Church Street, Ann Arbor, MI 48109-1043, USA
e-mail: djhauser@umich.edu

N. Schwarz
University of Southern California, Los Angeles, CA, USA

study that collected data from multiple sites, including several North American colleges and MTurk. Importantly, at some college sites, subject pool participants completed the study online. Figure 1 shows the IMC pass rates (and 95% confidence intervals) for each college site that completed the study online and for MTurkers. As is shown, MTurkers (94% pass) passed the IMC at higher rates than did those at any of the four college sites completing the same online survey.

Furthermore, Klein et al. (2014) also ran the study at multiple college sites where participants completed the same study in the lab under researcher supervision. Figure 2 displays the IMC pass rates (and 95% confidence intervals) for each North American college site that completed the study in the lab and for MTurkers, who completed it online. As is shown, MTurkers passed the IMC at significantly higher rates than did supervised undergraduates at 15 out of the 19 college sites. Only supervised undergraduates at Occidental College (Oxy: 89% pass), the University of Florida (UFL: 87% pass), the University of Virginia (UVA: 90% pass), and Washington and Lee University (WL: 88% pass) passed the IMC at rates comparable to the unsupervised MTurkers. Undergraduates under researcher supervision at the remaining 15 sites showed markedly lower attentiveness. Thus, although some studies have suggested that MTurkers are less attentive than traditional samples, other studies and newer data suggest the opposite. Given these mixed results, the present research was designed to bear upon this debate.

One potential reason for these discrepancies across studies may lie in the fact that MTurk is a nonreplenishing subject pool that *learns* over time (Chandler et al., 2014). As researchers have debated over MTurkers' attentiveness in the

pages of journal articles, researchers using MTurk samples have taken note and instituted measures to prevent potentially inattentive participants on MTurk from introducing error into their studies. The IMC was introduced as a measure of participant attentiveness and a ready-made attention filter (Oppenheimer et al., 2009; but also see Kittur, Chi, & Suh, 2008). Since then, IMCs have proliferated in MTurk studies and have been recommended in MTurk methods articles (Goodman et al., 2013; Paolacci et al., 2010). Researchers have even used performance on IMCs and other attention checks as a criterion for worker compensation (Chandler et al., 2014). With attention to instructions being so incentivized on MTurk in recent years, experienced MTurk participants may have *learned* to pay close attention to minor aspects of the instructions and may pass IMCs at higher rates than before. Recent research has suggested that this is indeed the case; MTurkers with more completed human intelligence tasks (HITs) and higher HIT acceptance ratios (high-reputation workers) are more likely to pass IMCs and are more attentive in various tasks (Peer, Vosgerau, & Acquisti, 2014).

Furthermore, over the years, research conducted on MTurk has begun to deliberately sample only high-reputation MTurkers. Recent articles have suggested restricting MTurk samples to include only high-reputation workers (e.g., Peer et al., 2014). Furthermore, highly influential MTurk-related methods articles have set a precedent for such restrictions as a way to ease concern over data quality (Berinsky et al., 2012; Goodman et al., 2013; Paolacci et al., 2010). Thus, whereas some MTurkers with low reputations may be less attentive to instructions, studies run in the modern MTurk paradigm would likely restrict participation to high-reputation samples and would not include low-reputation workers.

On the other hand, traditional samples for psychological research (undergraduates) have an entirely different experience. Most participate in tasks as a requirement for passing an introductory psychology course and have little time to *learn* about the norms of tasks and surveys. There are few incentives to pay attention to the task and instructions, which is why researchers often bring participants into a laboratory setting to attempt to guarantee their attention. For these reasons, we may begin to see MTurk populations overtaking traditional subject pool populations in attentiveness to instructions.

Furthermore, whereas much of the prior research has compared MTurk participants to supervised subject pool participants, no one has compared MTurk participants to *unsupervised* subject pool participants. With MTurkers being unsupervised and subject to distractions and subject poolers being supervised and bereft of distractions, differences in attentiveness between these populations may be attributable to differences in their situations. With modern software, researchers can allow subject pool participants to participate in online studies in much the same way as MTurkers—unsupervised and on

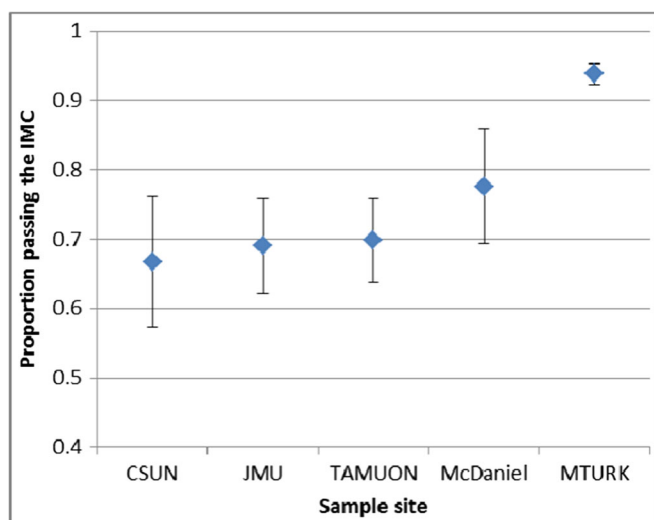


Fig. 1 Proportions passing the instructional manipulation check (IMC) for each Klein et al. (2014) North American college site that completed the study online and for MTurkers. Bars represent upper and lower bounds of the 95% confidence intervals. All data are derived from the openly available data of “Investigating Variation in Replicability,” by R. A. Klein et al., 2014, *Social Psychology*, 45, pages 142–152

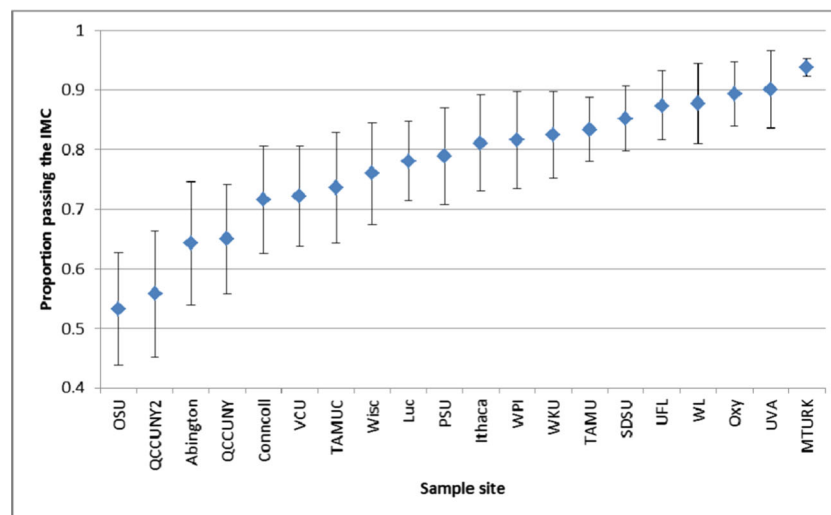


Fig. 2 Proportion passing the instructional manipulation check (IMC) for each Klein et al. (2014) North American college site that completed the study under lab supervision and for MTurkers. Bars represent upper and

lower bounds of the 95% confidence intervals. All data are derived from the openly available data of “Investigating Variation in Replicability,” by R. A. Klein et al., 2014, *Social Psychology*, 45, pages 142–152

a personal computer. In this uncontrolled environment, the same distractions that presumably diminish the attention of MTurk participants may also adversely affect subject pool participants. Therefore, a comparison of *online* subject pool participants to MTurkers bears on questions of theoretical and applied interest: First, do attention differences persist when situations are allowed to freely vary for both populations? Second, how attentive are online, unsupervised subject pool participants?

Study 1

We hypothesized that because MTurk studies strongly incentivize attention to instructions, MTurkers would be more likely to read instructions than would subject poolers in an unsupervised online survey. MTurkers and subject poolers were directed to an unsupervised online survey (containing an IMC) through a study participation website. If MTurkers are more likely to pay close attention to instructions, then they should pass the IMC at higher rates than subject pool participants.

Method

Participants A total of 396 workers (254 male, 142 female) from MTurk completed the online survey in exchange for 30 US cents. The HIT was posted on November 11, 2013, and was restricted to US workers who had not participated in any of our prior tasks containing IMCs (nonrepeating) with at least a 95% approval rating and 100 or more approved HITs. We made these restrictions in the present and following studies in order to accurately represent the modern MTurk study

paradigm; most recent MTurk research makes such sample restrictions, following the precedent and suggestions of influential MTurk methods articles (Berinsky et al., 2012; Goodman et al., 2013; Paolacci et al., 2010; Peer et al., 2014). Although the HIT did appear on MTurk-related forums, no posts mentioned the presence of attention check questions.¹

Eighty-five participants (32 male, 53 female) from the Fall 2013 undergraduate subject pool of a large Midwestern university completed the online survey in exchange for introductory psychology course credit. As is shown in Fig. 1, prior research has suggested that a large effect ($\phi = .50$) is a conservative estimate for the difference in IMC pass rates between MTurkers and online college students (Klein et al., 2014). A power analysis showed that 80% power for a large effect would require a total sample size of 32 participants (Faul, Erdfelder, Lang, & Buchner, 2007). However, we deliberately oversampled from both MTurk and the online subject pool for convenience and in order to examine unrelated questions (details available in the supplemental materials²).

Instructional manipulation check The IMC (adapted from Oppenheimer et al., 2009) presented the lure question “Which of these activities do you engage in regularly? (click on all that apply),” along with sports response options. However, contained within a large block of instructions was

¹ Forums searched included <http://mturkforum.com>, <http://mturkwiki.net>, <http://mturkgrind.com>, <http://mturklist.com>, <http://reddit.com/r/HITsWorthTurkingFor>, <http://turkernation.com>, and others. Google searching your requester name and HIT title often yields any available posts regarding your HIT.

² The supplemental materials are available here: <http://goo.gl/C8xcsJ>.

a sentence that informed participants to ignore the sports options and instead write “I read the instructions” in the box marked “other.” Those who did so were considered to have passed the IMC.

Although the content of our adapted IMC was similar to Oppenheimer et al.’s (2009) IMC, the instruction was different. In the classic version, participants were instructed to click on the title of the question in order to demonstrate participation. Our adapted version borrowed from a pilot study in Oppenheimer et al. (2009) by asking participants to input a special phrase in response to the question.

Procedure All participants were directed from their participant recruitment portals (MTurk for MTurkers, SONA [www.sona-systems.com] for undergraduates) to a Qualtrics survey. The survey contained six questions designed to measure adherence to Gricean norms (Schwarz, 1994) and the IMC. The Gricean norm questions were standard survey questions asking participants to judge their behavioral frequency of watching television and engaging in infrequent behaviors (such as getting haircuts and attending poetry readings) and to rate their success in life (details available in the supplemental materials).

The IMC appeared as either the first or the last question in the survey. Interestingly, MTurk participants were marginally more accurate on the IMC when it appeared last in the survey (97% correct) versus first (93% correct): $\chi^2(1, N = 396) = 3.3$, $p = .070$, $\phi = .09$. The effect of IMC order was not significant for subject pool participants: $\chi^2(1, N = 87) = 1.0$, $p = .32$.

Results

We collapsed the IMC pass rates across IMC orders to compare the attentiveness of the two populations. As predicted, the IMC pass rate for the subject pool was substantially lower than the pass rate for MTurk; 95% of MTurkers passed the IMC, as compared to only 39% of subject pool participants, $\chi^2(1, N = 481) = 168.8$, $p < .0001$, $\phi = .60$.

Study 2

Although Study 1 suggested that subject pool participants are less likely to fully read the instructions in an unsupervised online task, it also utilized an IMC that has been used by others in many prior studies. Many MTurkers participate in numerous studies, so nonnaiveté might account for the high IMC pass rate on MTurk. Indeed, if MTurk participants had seen the IMC before, they would easily be able to identify the question as a “trick” question and to answer it correctly without having to read the instructions. To address this alternative explanation, we presented a novel IMC to MTurk and subject pool populations in another unsupervised online survey. If

MTurkers are more likely to fully read the instructions, they should pass the novel IMC at a higher rate than subject pool participants.

Method

Participants A total of 185 workers (111 male, 71 female, 3 unspecified) from MTurk completed the survey in exchange for 30 US cents. The HIT was posted on February 17, 2014, and was restricted to nonrepeating US workers with a 95% approval rating and 100 or more approved HITs. Although the HIT did appear on MTurk-related forums, no posts mentioned the presence of attention check questions.

A total of 245 participants (142 male, 103 female) from the Winter 2014 undergraduate subject pool of a large Midwestern university completed the online survey in exchange for psychology course credit. We again deliberately oversampled (relative to an expected large effect size) from both MTurk and the online subject pool for convenience and in order to examine unrelated questions (details available in the supplemental materials).

Novel IMC At the end of each task, participants were given a novel IMC, modeled on Oppenheimer et al. (2009). The IMC contained the lure question “Which of these personality traits best describe you and your personality? (click on all that apply)” followed by a list of 12 personality trait options. However, within a large block of instructions for the question, a sentence specified that to demonstrate attention to the instructions, participants should ignore the personality items and instead mark the “other” box and type “I read the instructions” into the accompanying text box. Participants who followed these instructions were scored as passing the novel IMC.

Procedure All participants were directed from their participant recruitment portals (MTurk for MTurkers, SONA for undergraduates) to a Qualtrics survey. The survey contained an unrelated sentence completion task, followed by the novel IMC. The task asked participants to generate the end to six sentence fragments, then to rate the valence of the ending and the intentionality of the subject (for more details, see the supplemental materials).

Results

As predicted, the MTurkers passed the novel IMC at a much higher rate (96% pass) than did online subject pool participants (26% pass), $\chi^2(1, N = 430) = 212.7$, $p < .0001$, $\phi = .70$. Therefore, even with a novel, unfamiliar IMC, MTurkers are more attentive to instructions and pass at higher rates than do subject pool participants in an unsupervised survey.

Study 3

The prior studies demonstrated that MTurkers pass both established (Study 1) and novel (Study 2) IMCs at higher rates than unsupervised subject pool participants. However, in both studies, the IMC required participants to sift through a large instructional block of text in order to ascertain the true purpose of the question. Additionally, both studies required participants to input a response in a free text format to complete the IMC. Thus, participants heuristically searching for such structural characteristics of questions (large blocks of text and text entry response boxes) might have been able to easily identify and pass IMCs without necessarily being more attentive to the instructions.

To rule out this alternative explanation, we created a novel IMC for Study 3 that was structurally dissimilar to the IMCs used in the prior studies. The last sentence in a short three-sentence introduction to the demographic questions instructed participants to mark the first two response options to the next question in order to demonstrate attention. Then, the next question asked participants to mark with which political parties they strongly identified, and contained two unpopular political parties as the first two response options. Thus, the IMC in Study 3 embedded the crucial information in a much smaller introductory text and contained a different correct response for passing the IMC, which did not require free text entry and was not associated with a text response box. If MTurkers pass IMCs at high rates because of heuristics that look for certain structural characteristics of IMCs, then they should pass this novel IMC at a similar rate to online subject pool participants. However, if MTurkers are truly more attentive than online subject pool participants, then MTurkers should pass this IMC at higher rates.

Prior research has also shown that attentive participants show larger effect sizes on well-established psychological tasks than do inattentive participants (Oppenheimer et al., 2009; Peer et al., 2014). If MTurkers are more attentive than online subject pool participants, then MTurkers should have a larger effect size on a well-established task than would online subject pool participants. However, if MTurkers pass IMCs at higher rates because of IMC-catching heuristics, we should expect to find no effect size differences between the two populations. Thus, Study 3 also contained Thaler's (1985) beer/soda-pricing task, in which minor wording variations in a scenario affect the amount that participants are willing to pay for an item; this task has appeared in prior research demonstrating how IMCs gauge attentiveness (Oppenheimer et al., 2009). If MTurkers are more attentive than online subject pool participants, then the effect of minute wording variations in the task should be larger for MTurkers than for online subject pool participants.

Method

Participants A total of 149 workers (103 male, 46 female) completed an online survey in exchange for 20 US cents. The HIT was restricted to US workers who had not participated in any of our prior tasks containing IMCs (nonrepeating) with at least a 95% approval rating and 100 or more approved HITs.

Ninety participants (46 male, 44 female) from the Fall 2014 undergraduate subject pool of a large Midwestern university completed the online survey in exchange for introductory psychology course credit. We again deliberately oversampled (relative to an expected large effect size) from both MTurk and the online subject pool.

Novel IMC At the end of the survey, participants were given a novel IMC within the demographic block of questions. The question block introduction read "Finally, we have a few demographic questions for you. Please answer the questions below. For the next question, mark the first two response options to demonstrate attention." The first question (the IMC) contained the lure question "Which political parties do you strongly affiliate with? Mark all that apply." followed by a list of eight American political parties: Citizens party, Socialist Action party, Constitution party, Libertarian party, Green party, Democratic party, Republican party, Independent. Participants selecting both the Citizens party and the Socialist Action party were scored as passing the IMC.

Soda-pricing task Participants completed a task modeled on the soda-pricing task adapted from Oppenheimer et al. (2009) and originally found in Thaler (1985). Participants were asked to imagine the following scenario to the best of their ability and to answer the following question (between-subjects manipulation in parentheses):

Imagine that you are on the beach on a hot day. For the last hour, you have been thinking about how much you would enjoy an ice cold can of soda. Your companion needs to go to the bathroom and offers to bring back a soda from the only nearby place where drinks are sold, which happens to be a run-down grocery store (fancy resort). Your companion asks how much you are willing to pay for the soda and will only buy it if it is below the price you state. How much are you willing to pay?

The question was followed by an open text response box. Thaler (1985) found that participants typically are willing to pay more for the can of soda when it is sold by the fancy resort (rather than the run-down grocery store). Furthermore, because the manipulation involves a subtle variation in wording between the scenarios, attentive participants show stronger effects (Oppenheimer et al., 2009).

Procedure All participants were directed from their participant recruitment portals (MTurk for MTurkers, SONA for undergraduates) to a Qualtrics survey. Participants first completed an unrelated semantic judgment task in which they judged the similarity of five word pairs. Participants then completed the soda-pricing task, followed by an unrelated valence inference task in which participants judged the likelihood of two events, given a sentence that varied in one word. Finally, participants completed the demographic questions (containing the IMC). Importantly, the word manipulation in the valence inference task did not affect IMC pass rates for either MTurkers or subject pool participants ($ps > .14$). For more details on the unrelated tasks, see the supplemental materials.

Results and discussion

Novel IMC pass rates As predicted, the MTurkers passed the novel IMC at a much higher rate (25.5% pass) than did online subject pool participants (2.2% pass), $\chi^2(1, N = 239) = 21.8, p < .001, \phi = .30$. Noticeably, this IMC was more difficult than those used in prior studies (25.5% pass rate for the MTurkers in Study 3 vs. 96% and 95% pass rates for MTurkers in Studies 1 and 2, and 2.2% pass rate for subject pool participants vs. 29% and 36%). However, even with the increased difficulty, MTurkers still demonstrated more attentiveness to instructions, passing at higher rates than the unsupervised subject pool participants. This was the case even though simple heuristics (looking for a “text box” or a “large instruction block”) cannot account for MTurkers’ superior performance in this study.

Soda-pricing task effect sizes To reduce the impacts of outliers and unequal variances across conditions, we first rank-transformed willingness to pay (WTP; 1 = lowest WTP, 239 = highest WTP). In order to examine whether the minor wording variation of expectation differentially affected MTurkers versus subject poolers, we conducted a 2 (sample: MTurk, subject pool) \times 2 (expectation: fancy resort, run-down grocery store) between-subjects analysis of variance on ranked WTP. Replicating prior research, expectations affected the WTP, as was evident in a significant main effect of expectation, $F(1, 235) = 21.56, p < .001, \eta_p^2 = .08, 95\% \text{ CI } [10.8, 26.7]$: Participants were willing to pay more for the soda when it was sold by a fancy resort ($M = 147.2, SE = 5.9$) than when it was sold by a run-down grocery store ($M = 111.7, SE = 5.8$).

As we predicted, the strength of the expectation effect depended on the sample, as could be seen in a significant two-way interaction between expectation and sample, $F(1, 235) = 4.35, p = .038, \eta_p^2 = .02, 95\% \text{ CI } [0.5, 16.4]$. We diagnosed this interaction with simple effect tests of expectation at each level of sample (Table 1). Consistent with the predictions that MTurkers are more attentive than subject pool participants, expectation had the strongest effect on MTurk participants, $F(1, 235) = 30.07, p < .001, \eta_p^2 = .11, 95\% \text{ CI }$

Table 1 Mean ranked willingness to pay (WTP; with *SD*) and median WTP by expectation and sample

Sample	Expectation	
	Run-Down Grocery Store	Fancy Resort
MTurk		
Mean (<i>SD</i>)	75.1 (59.8)	129.4 (62.5)
Median	\$1.50	\$3.00
Subject Pool		
Mean (<i>SD</i>)	140.2 (61.5)	160.8 (56.3)
Median	\$3.00	\$3.50

[34.8, 73.8], for the simple effect of expectation. As is shown in the top portion of Table 1, MTurkers were willing to pay substantially more for the soda when it was sold by a fancy resort than when it was sold by a run-down grocery store.

Also as predicted, expectation had a weak effect on the subject pool participants, $F(1, 235) = 2.62, p = .107, \eta_p^2 = .01, 95\% \text{ CI } [-4.5, 45.7]$, for the simple effect of expectation. As is shown in the bottom portion of Table 1, the subject pool participants were willing to pay only marginally more for the soda when it was sold by a fancy resort than when it was sold by a run-down grocery store. Since more-attentive samples show stronger effects on well-established tasks that rely on minor wording variations (Oppenheimer et al., 2009; Peer et al., 2014), this further confirms that MTurk participants are more attentive than subject pool participants. This also casts doubt on attentiveness differences due to IMC-identifying heuristics, since MTurkers demonstrated more attentiveness than did subject pool participants on a task that has no resemblance to common IMCs.

Additionally, we found a significant main effect of sample on ranked WTP, $F(1, 235) = 35.83, p < .001, \eta_p^2 = .13, 95\% \text{ CI } [16.2, 32.1]$: Subject pool participants were willing to pay more for the soda ($M = 150.5, SE = 6.4$) than were MTurkers ($M = 102.3, SE = 4.9$), which may have reflected age differences between the populations.

General discussion

In three studies, MTurkers were consistently more likely to pass IMCs than were subject pool participants under comparable online data collection conditions. These experimental results are consistent with other observations (Fig. 1) and bear on the use of MTurk as well as subject pools.

MTurk

Despite mixed evidence in the past, it currently appears that MTurkers are indeed more attentive to instructions than are

undergraduate samples. These results challenge the familiar concern that MTurkers are less attentive than traditional samples (see the informal survey in Chandler et al., 2014). However, that concern has never enjoyed strong empirical support, despite its popularity. Only two studies exist that have experimentally compared undergraduate to MTurk participants on attentiveness to instructions in attention checks, and only one of those studies found MTurkers behaving less attentively than undergraduates (Goodman et al., 2013, Study 2). The other study observed no differences in attentiveness (Paolacci et al., 2010).

Furthermore, in the single study in which MTurkers were less attentive than undergraduate participants, the researchers placed no country restrictions on MTurk for an English-language survey. As a result, the majority of the MTurk sample were nonnative English speakers, who were compared on IMC pass rates against a sample of college undergraduates who were mostly native English speakers. Since passing an English-language IMC is heavily reliant on comprehending it (evident in an IMC pass rate of 71% for native English speakers vs. 29% for nonnative English speakers; Goodman et al., 2013, Study 2), these sample discrepancies make it difficult to draw firm conclusions regarding attentiveness differences between MTurkers and college undergraduates. Additionally, MTurk offers simple and often-recommended avenues for restricting the country of participants (Peer et al., 2014), making it unnecessary to recruit MTurk participants with potentially poor English language skills for English-language surveys.

MTurkers' high attention to instructions may constitute a mixed blessing for social science research on MTurk. On the one hand, it lends further support to the use of MTurk samples by showing that participants on the site are quite attentive. This suggests that MTurk is a viable avenue for collecting survey data, crowdsourcing tasks, and even psychological tasks that require somewhat complicated instructions. On the other hand, it also suggests that this population may be going through somewhat different mental processes when approaching tasks than do traditional subject pools. MTurkers may pay close attention to minor aspects of question wording, looking for IMCs, which may lead to different question interpretations than researchers intended (for a review, see Schwarz, 1994). Prolonged exposure to IMCs may also prompt MTurkers to treat surveys, tasks, and individual questions with suspicion, which can have pronounced cognitive effects (Mayo, Alfasi, & Schwarz, 2014; Schul, Mayo, & Burnstein, 2004). Finally, attentiveness can be a moderation condition for many psychological effects, and effects found on MTurk may not hold for populations and conditions with lower attentiveness. Our Study 3 illustrates this, demonstrating that Thaler's (1985) soda-pricing task showed strong effects with an attentive sample (MTurkers), but relatively weak effects with an inattentive sample (subject pool participants).

Our results also have implications for the sustainable use of MTurk samples in social science research. As has been suggested by our data and by others (Chandler et al., 2014), MTurk is a subject pool that *learns*, and its users often know more about social science research procedures than researchers may like. Even when researchers exclude MTurkers who participated in their own previous studies, many MTurkers have seen common measures in other surveys (Chandler et al., 2014). Hence, researchers may want to avoid using measures that may cease to tap the intended psychological construct after repeated administration. Furthermore, when they are incentivized for paying close attention to instructions, MTurkers unsurprisingly become quite attentive over time. Hence, researchers should avoid incentivizing practices that they do not wish to become norms on MTurk. For instance, issuing MTurkers bonuses (additional compensation) for completing longer surveys may lower attrition, but it may also encourage the practice of staying in surveys longer than necessary, which could affect the results in persistence tasks.

One caveat for the present results, however, deserves mention. We followed the current standard practice by restricting MTurk samples to workers with a high reputation across a large number of HITs. MTurkers who have successfully completed numerous surveys (i.e., those who have high reputations) are more attentive than less-experienced MTurkers (Peer et al., 2014). We would undoubtedly expect less drastic attentiveness differences between the samples if we had not made such restrictions. However, these restrictions were necessary in order to be representative of the criteria that are typically used in psychological research on MTurk (Berinsky et al., 2012; Goodman et al., 2013; Paolacci et al., 2010) and widely recommended in the MTurk literature (Peer et al., 2014). Thus, whereas our research demonstrates that the typical MTurk participants in psychological research are more attentive than comparable subject pool participants, not all MTurkers are highly attentive, especially those with low reputations.

Subject pools

For many readers, the largest surprise in our data may be the very poor attention observed under subject pool conditions. In both experiments, the majority of our subject pool participants, who received course credit for their participation, failed the IMC. Inspection of Fig. 1 shows that this poor performance is not exceptional and is also observed for many other subject pools. These observations challenge the belief that subject pool participants may do a better job than MTurkers. Instead, differences between subject pool and MTurk results may reflect the opposite of what is often assumed: dismal attention to detail in the subject pool, and high attention on MTurk. Future research may fruitfully explore whether the

difference in attention favors results that reflect heuristic processing under subject pool conditions, but systematic processing under MTurk conditions.

Author Note We thank Aashna Sunderrajan and Madhuri Natarajan for assisting with Study 1, and the UMich OLab for their valuable insight.

References

- 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, 351–368.
- 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. doi:10.3758/s13428-013-0365-7
- Clifford, S., & Jerit, J. (2014). Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies. *Journal of Experimental Political Science*, 1, 120–131. doi:10.1017/xps.2014.5
- Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE*, 8, e57410. doi:10.1371/journal.pone.0057410
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. doi:10.3758/BF03193146
- 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. doi:10.1002/bdm.1753
- Kapelnier, A., & Chandler, D. (2010). Preventing satisficing in online surveys: A “kapcha” to ensure higher data quality. In *Proceedings of CrowdConf 2010*. Available at www.academia.edu/2788541/Preventing_Satisficing_in_Online_Surveys
- Kittur, A., Chi, E. H., & Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. In M. Czerwinski & A. Lund (Eds.), *Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems* (pp. 453–456). New York: ACM.
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., & Nosek, B. A. (2014). Investigating variation in replicability. *Social Psychology*, 45, 142–152.
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44, 1–23. doi:10.3758/s13428-011-0124-6
- Mayo, R., Alfasi, D., & Schwarz, N. (2014). Distrust and the positive test heuristic: Dispositional and situated social distrust improves performance on the Wason rule discovery task. *Journal of Experimental Psychology: General*, 143, 985–990.
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45, 867–872.
- Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, 23, 184–188.
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5, 411–419.
- 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. doi:10.3758/s13428-013-0434-y
- Schul, Y., Mayo, R., & Burnstein, E. (2004). Encoding under trust and distrust: The spontaneous activation of incongruent cognitions. *Journal of Personality and Social Psychology*, 86, 668–679.
- Schwarz, N. (1994). Judgment in a social context: Biases, shortcomings, and the logic of conversation. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 26, pp. 123–162). San Diego: Academic Press.
- Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science*, 4, 199–214.