Computers in Human Behavior 29 (2013) 2156–2160

Contents lists available at SciVerse ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing ☆

Krista Casler *, Lydia Bickel, Elizabeth Hackett

Franklin & Marshall College, Department of Psychology, P.O. Box 3003, Lancaster, PA 17604-3003, USA

ARTICLE INFO

Article history:
Available online 31 May 2013

Keywords:
MTurk
Social media
Methodology
Recruitment
Crowd-sourcing

ABSTRACT

Recent and emerging technology permits psychologists today to recruit and test participants in more ways than ever before. But to what extent can behavioral scientists trust these varied methods to yield reasonably equivalent results? Here, we took a behavioral, face-to-face task and converted it to an online test. We compared the online responses of participants recruited via Amazon's Mechanical Turk (MTurk) and via social media postings on Twitter, Facebook, and Reddit. We also recruited a standard sample of students on a college campus and tested them in person, not via computer interface. The demographics of the three samples differed, with MTurk participants being significantly more socio-economically and ethnically diverse, yet the test results across the three samples were almost indistinguishable. We conclude that for some behavioral tests, online recruitment and testing can be a valid—and sometimes even superior—partner to in-person data collection.

1. Introduction

The most common participants in behavioral, psychological research are American college students. This group has the benefit of convenience, but many scholars argue—understandably and rightly—for much more diverse sampling (Henrich, Heine, & Norenzayan, 2010; Henry, 2008; Sears, 1986). Thanks to the Internet and the proliferation of Internet-enabled devices in the hands of people from many walks of life worldwide, researchers are now tapping into a more diverse population than ever before. Recruiting over social networks, through online message boards, and using targeted email lists (to name just a few possibilities) has brought an unprecedented variety of participants in touch with researchers' online surveys and tests.

The benefits of diversity and spread acknowledged, there have been growing pains associated with increased use of Internet-based methods of recruitment and testing, and many behavioral scientists have yet to take advantage of online participants. Concerns have taken several forms, but they primarily have involved uncertainty over whether these methods yield trustworthy and equivalent results compared to more traditional methods. For instance, one early apprehension was that Internet participants were more depressed or maladjusted than traditional samples (Kraut, Patterson, Lundmark, Kiesler, Mukophadhyay, & Scherlis, 1998)—a misconception thoroughly refuted by Gosling, Vazire, Srivastava, and John (2004). Others have wondered whether technology-sourced participants might be less invested or motivated than traditional ones, or whether the anonymity provided by Internet participation negatively affects responding. Reassuringly, little evidence has turned up in support of these worries, either. Internet participants appear to be no less motivated than their traditional peers (Gosling et al., 2004); widespread methods such as "instructional manipulation checks"[1] (e.g., Oppenheimer, Meyvis, & Davidenko, 2009) successfully ensure that participants are paying attention while completing online surveys and tests; and if anything, the cloak of Internet anonymity appears to be more of a benefit than a detriment in many online surveys (Joinson, 1999; Joinson, Woodley, & Reips, 2007; Simsek & Viega, 2001). The conclusion seems to be that with sensible safeguards and manipulation checks in place, online participation is no greater a concern to data integrity than the other biases and demand characteristics researchers guard against in more standard methods of data collection. Online distribution of surveys today is popular and increasingly trusted (Davidov & Depner, 2011; Smyth & Pearson, 2011; Tuten, Urban, & Bosnjak, 2002).

Beyond traditional message boards and email, there is a new and fast-growing method of Internet recruitment available today—marketplace crowdsourcing, primarily done through

☆ This research was supported in part by grants from the Franklin & Marshall College Committee on Grants.

* Corresponding author. Tel.: +1 717 291 3828.
E-mail addresses: kcasler@fandm.edu (K. Casler), lbickel@fandm.edu (L. Bickel), ehackett@fandm.edu (E. Hackett).

[1] These often consist of questions that appear to be part of the questionnaire or task, but they tell participants to respond with "I have read the instructions" or another given response rather than to truly answer the question.

0747-5632/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.chb.2013.05.009

Amazon's Mechanical Turk (MTurk) program. This means of recruitment has been less studied than other online methods involved in the studies referred to above. MTurk is an Internet marketplace where employers post "Human Intelligence Tasks" (HITs) for paid workers to complete, with typical HITs consisting of small tasks such as comparing product images, transcribing podcasts, copying business card text into a database, or responding to online questions. The marketplace principally was designed for employers ("requesters") to hire workers to do simple jobs that are better suited to human than computer-based labor. Of late, scholars also have found it to be a fruitful forum for recruiting participants to complete computer-based tasks.

From the standpoint of diversity, MTurk respondents are much more demographically varied than participants recruited through traditional methods and slightly more diverse even than other Internet samples, with workers residing in dozens of countries worldwide (Berinsky, Huber, & Lenz, 2012; Buhrmester, Kwang, & Gosling, 2011). Paolacci, Chandler, and Ipeirotis (2010) found that American and Indian workers comprise a majority of MTurk workers, and that approximately 65% of U.S. based workers are female, which is consistent with demographics of other Internet samples. They also found the average age of workers to be 36 years, slightly younger than the U.S. population and the population of Internet users but significantly higher than standard college samples commonly used in psychology testing. Approximately 14% of U.S. based workers reported MTurk as their primary source of income and 61% reported that earning extra money was an important motivating factor in their participation (Paolacci et al., 2010). Interestingly, lower rates of pay have been shown to make it somewhat more difficult for requesters to recruit participants, but the data quality has remained unaffected (Buhrmester et al., 2011).

Diversity noted, the big question is whether MTurk-sourced data are reliable: can a largely financially motivated, online, worldwide sample provide trustworthy results for the academic researcher? Preliminary data have been promising. For instance, researchers found no differences in results of a prisoner's dilemma task, a priming task, and a framing effects task whether they were completed online by MTurk workers or on a computer by students in a laboratory (Horton, Rand, & Zeckhauser, 2011). MTurk also has been used in studies examining the effects of pay rate on the quality and quantity of output data (Mason & Watts, 2009), gender differences in risk taking (Eriksson & Simpson, 2010), body size and satisfaction (Gardner, Brown, & Boice, 2012), and human cooperation over networks (Suri & Watts, 2011), all with outcomes consistent with those obtained via standard recruitment.

These reports are uniformly positive. What is more, recent papers are available that include "how to" information for researchers who wish to try out MTurk labor (see, for example, Mason & Suri, 2012). However, there is a significant limitation in the literature that has kept many behavioral scientists from capitalizing on this source of participants: In comparing to "typical" samples (i.e., college undergraduates), researchers thus far have pitted MTurk participants against participants who likewise complete a task *on a computer* in the lab. This procedure importantly minimizes differences in testing across recruitment conditions (i.e., all participants complete identical computerized tasks, with the only difference being whether they are college recruits in a laboratory room or online recruits at home), but it also means that the type of face-to-face, behavioral methods so central to experimental psychology have been excluded from comparisons thus far. To our knowledge, researchers have not attempted to compare behavioral, face-to-face types of methods with versions adapted for MTurk's widespread, online reach. We put forward the following project as a small, first comparison of this sort.

In the investigation described here, we took an in-person task and customized it for online participation. We compared three recruitment and testing conditions: traditional college undergraduates completing the behavioral task one-on-one with an experimenter in a university lab test room, adults recruited via social media postings and participating at their leisure online, and adults crowd-sourced via MTurk and also participating online. The task was an object selection and categorization task originally designed to see how adults learn about the functions of tools, although the theoretical particulars of the task itself are largely irrelevant to the current question of interest. We wondered how adults completing the task on a computer would compare to adults completing the task at a table with a human experimenter.

2. Materials and methods

2.1. Participants

2.1.1. Eighty-six participants were recruited in total

2.1.1.1. Traditional undergraduate recruits. Twenty-four traditional undergraduate students participated in the in-person, lab-based condition. They were recruited via word-of-mouth on a college campus. The sample was 58% female, ranged in age from 18 to 23 (mean age = 20.5; SD = 1 year), and reported the following ethnic make-up: White = 67%, Asian = 12.5%, Hispanic = 4%, Black = 12.5%, Other or did not report = 4%. These participants received $5USD at the conclusion of their lab visit.

2.1.1.2. Social media recruits. Thirty adults were obtained via social media postings on Facebook, Reddit, Twitter, and StumbleUpon. This sample was 83% female, ranged in age from 19 to 73 (mean age = 26; SD = 14 years), and reported the following ethnic make-up: White = 93%, Asian = 3.5%, Hispanic = 3.5%.

2.1.1.3. Amazon MTurk recruits. Thirty-two adults were recruited with a HIT posted on MTurk. They were 35% female, ranged in age from 18 to 60 (mean age = 33; SD = 11 years), and reported the following ethnic diversity: White = 37.5%, Asian = 40.5%, Hispanic = 6%, Black = 6%, Other or did not report = 10%. These participants received $0.50USD in their Amazon.com account for successfully completing the HIT, a high rate of pay compared to similar tasks on MTurk.

2.2. Procedure

In-person testing took place in a comfortable test room in a university psychology lab. Participants sat across from a female experimenter at a small table. The experimenter presented four pairs of simple tools, one at a time, to the participant. In each pair, one of the tools was a familiar object, and it was always demonstrated performing its standard function (e.g., a paintbrush, used for painting). The other tool was always a novel object that was highly similar in overall shape and appearance to the known tool. The novel tool was either named and demonstrated performing its action ("teaching" trials) or named but only described according to non-functional features such as its color or place of manufacture ("non-teaching" trials). Regardless of condition, however, the participant was asked to hold, try out, and handle every tool twice. After learning about a pair of tools, participants were introduced to a new, unrelated task and asked to choose which tool—the novel or the known one—they needed to complete that goal. Both tools could do the job equally well. This method was followed for three additional tool pairs. Sessions lasted 10–12 min.[2]

[2] The theoretical particulars and predictions of the task itself are not central to the present paper's aims and therefore are described only in broad terms here; for a complete description, see Casler, Hackett, & Bickel, 2013, or contact the first author.

Internet recruits completed the task in their homes, places of work, or elsewhere at their convenience. The online test was hosted by Qualtrics.com, a highly equipped, web-based survey and research software. Participants clicked a link that brought them to an information page and online consent form. The behavioral procedures were converted to online presentation by video-recording the experimenter's demonstrations and spoken questions, editing the footage for clarity of presentation, and adding subtitles where necessary. After consenting, participants viewed the first pair of tool videos (one novel, one known); the videos used identical language and pacing as the in-person methods. Although participants could not physically handle and try out the tools, the demonstrations were repeated three times in the video (to mirror the single demonstration plus two attempts made by undergraduates in the in-person condition), using close-up shots clearly depicting the handling and features of the objects. A third video then introduced the unrelated task, and participants were asked the test question audibly as a voice-over in the video. Participants indicated their tool selection by clicking a radio button beside a photograph of their tool choice. This method was followed for all four tool pairs. The online session required a similar amount of time to complete as an in-person one.

To ensure that online participants were capable of completing the task adequately and were taking their involvement seriously, we included two manipulation safeguards. First, after each video, we provided a small text box for them to reply to this prompt: "Name one thing you learned about a [tool name] from the video". We did this to be sure participants (a) had paid attention to the video, (b) had their Internet device's sound turned on, and (c) understood English sufficiently such that they could follow the very simple narration and requests (we did not expect or require perfect English, nor did we expect responses to be more than a few words). Second, we controlled the deployment of test questions such that participants could not move on from a video to the next video or task until the length of the current video's time had elapsed. This helped ensure that online participants were not skipping past videos or watching them only in part before responding to the test questions.

3. Results

Our goal was to explore to what extent traditional, in-person, behavioral testing with undergraduate students in the lab—long the backbone of experimental psychology—could be successfully brought to today's online platform. We compared behavioral testing with crowd-sourced labor as well as with online recruitment using social media.

The expected behavioral outcome was that adults would prefer the novel object significantly more often than predicted by chance on non-teaching trials, but that they would select at chance when picking between the novel versus the known tool on teaching trials (readers are encouraged to refer to Casler et al., 2013, for theoretical discussion). This pattern was indeed obtained with the behavioral, in-person sample (see Fig. 1). More central to the current question, however, a one-way ANOVA found no differences across the three recruitment/testing conditions in terms of how often participants chose the novel tool on non-teaching trials, $F(2,79) = .606$, *n.s.*, or on teaching trials, $F(2,78) = .538$, *n.s.* (see Fig. 1). Averaged across the three conditions, participants preferred the novel tool in non-teaching trials significantly greater than would be predicted by chance, $M = 72.5\%$, SD = 36%, $t(1,81) = 5.64$, $p < .0001$, $d = .63$. In contrast, participants showed no preference for the novel tool in teaching trials, choosing it equally to the familiar tool, $M = 49.3\%$, SD = 32%).

Differences were found in measures of diversity, however. First, the recruitment groups varied in their degree of self-identified ethnic diversity, χ^2 (10, $N = 86$) = 26.57, $p < .01$. Within the MTurk sample, only 37.5% of respondents identified themselves as non-Hispanic Caucasian, as compared to 67% of traditional participants and 93% of social media recruits. Similarly, the groups varied significantly in terms of economic status. We did not collect average family income data from the in-person undergraduate participants, but the mean family income level for the social media recruits was in the $125,000–$150,000 range and the mean for the MTurk participants was significantly lower, in the $25,000–$50,000 range, $t(60) = -6.188$, $p < .0001$, $d = 1.56$. The groups also differed in distribution of males and females, χ^2 (2, $N = 86$) = 15.25, $p < .0001$, with the MTurk sample containing the most males (65%), the social media sample containing the fewest males (17%), and the traditional undergraduate students in the middle (42% male). Finally, the three samples differed significantly in age distribution, $F(2,83) = 8.73$, $p < .0001$, $\eta^2 = .174$. Independent samples *t*-tests revealed that social media ($M = 26$ years, SD = 14) and MTurk ($M = 33$ years, SD = 11) samples were similar to one another in age distribution, $t(60) = -1.96$, *n.s.* However, both online groups were significantly older than the traditional undergraduate participants ($M = 20.5$ years, SD = 1): social media, $t(52) = 2.00$, $p = .05$, $d = .59$; MTurk, $t(54) = 5.28$, $p < .0001$, $d = 1.49$.

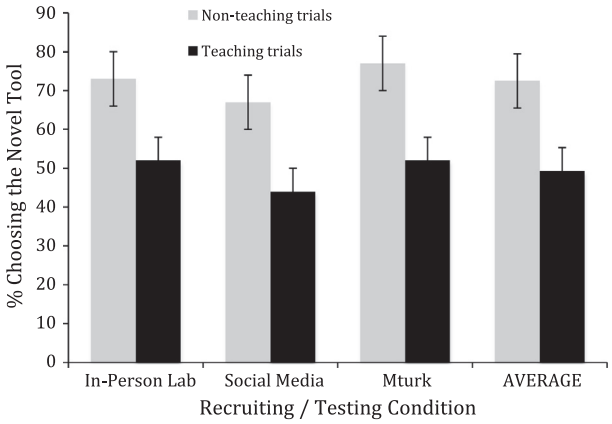

Fig. 1. Mean percentage of the time participants chose the novel tool over the known tool according to condition. Error bars depict the standard error of the mean.

4. Discussion

This was a simple study but one with an important suite of messages: (a) crowd-sourced participants can provide high quality data; (b) they bring a highly desirable degree of diversity to the researcher's table; and, most importantly (c) they and other online recruited participants can successfully complete certain behavioral-type tasks traditionally thought of as requiring in-person testing. Given recent urgings for behavioral researchers to reduce reliance on "WEIRD" participants (Western, Educated, Industrialized, Rich, and Democratic)—who may well be less representative of the general population than has previously been supposed (Henrich et al., 2010)—this is very good news. Certainly there are many instances where dialogue, contingent responding, and face-to-face interaction between researcher and participant are key. However, Mturk recruitment and online testing is an excellent option when those requirements are absent, and we encourage our colleagues who typically conduct in-person testing to consider adding it to their toolkit. We argue that with some creativity and care, behavioral tasks can successfully be adapted to the screen. The results of the current study—the first comparison of its kind that we know of—indicated no downsides to such presentation; the choice patterns were identical for online and face-to-face test takers.

What makes this finding especially exciting is that the MTurk participants were very favorably diverse. Corroborating previous reports (Buhrmester et al., 2011; Paolacci et al., 2010), this sample represented a greater range of ages and a significantly higher mean age than usually provided by university pools of Introduction to Psychology students. Likewise, they came from a widespread geographic range and represented a much more varied distribution of ethnic and economic backgrounds than is typically seen in academic research. Indeed, the MTurk participants in this study were significantly more diverse even than the other group of online respondents recruited via social media.

We consider it a powerful thought that researchers do not always need to travel long and far conducting special cross cultural research in order to confirm findings across groups of participants. We heartily support cross-cultural work, and especially that which is theoretically motivated based on documented cultural differences. However, we also find the idea refreshing that when no such differences are predicted, experimenters do not need to treat people groups as ''conditions'' or ''independent samples'' but instead can simply tap into a wide pool of diverse people from the start. In the research described here, for instance, we can conclude with some confidence that the cognitive biases under investigation are likely common to the *human mind*, not just to primarily white, middle class, well-off, young college students in the United States.

We note that where the current MTurk sample fell short in diversity was in gender distribution; the in-person sample, with a 42–58% male-female split, was more balanced than the MTurk sample at 65% male. However, we note that although the MTurk sample was weighted toward males, other studies of MTurk have reported slightly higher female response rates (Buhrmester et al., 2011; Paolacci et al., 2010), mitigating concerns that this method simply yields male-dominated samples across the board.

Beyond diversity, there is a benefit to MTurk recruiting that does not come across in the data but instead is embedded in the methodology itself. Traditional in-person, lab-based testing is a time consuming process, and one which requires great care on the part of researchers to maintain consistency in deployment across participants and time. In the current study, we spent several weeks recruiting, scheduling, and testing each lab-based participant individually, sometimes waiting fruitlessly for ''no-show'' participants, and maintaining stimuli due to wear and tear. In contrast, we spent a few days carefully preparing the videos, setting up the online test, trying it out, and establishing an MTurk account, but once the HIT went live, all participant slots were completed in under an hour. Social media recruitment also was faster than in-person testing, but it spanned several days and was not as immediate as the MTurk response. Scaling up sample sizes is often daunting with behavioral tasks, but this issue is decreased with social media recruiting, and it nearly disappears altogether with crowd-sourced labor. What is more, the pay rate of $0.50USD on MTurk was a desirable, competitive wage, yet it cost just 10% of our payment to in-person participants in the lab ($5USD). The savings of both time and money have potential to be substantial.

On the other hand, there are costs to consider using crowd-sourced labor. With any online participation, responses must be screened carefully for ''red flags'' of incompetence or non-investment (Oppenheimer et al., 2009). After all, one of the benefits of seeing one's participants in person is having confidence that they are motivated, cognitively capable, and paying attention. Fortunately, with thoughtful and creative manipulation checks in place, researchers usually can discard online participants who have not taken the task seriously or who had insufficient skills to complete it correctly. MTurk also allows posters to restrict access to their HITs only to workers who meet certain requirements, such as having a certain approval rating based on previous adequate completion of HITs or affirming that they speak a certain level of English.

The ratings aspect is a particular benefit of MTurk compared to recruiting via social media or other more traditional online venues. Due to anonymity, many online test or survey takers have no reason—outside of their own curiosity in the subject matter or personal ethics—to take their participation seriously. It is not uncommon for researchers to find that participants have raced through questions, skipped portions of the task, or filled text boxes with a few meaningless keystrokes. In contrast, MTurk workers retain the benefit of anonymity, but slack performance has the potential to reduce their approval ratings, which can carry a penalty of restricting future access to higher quality and better paying HITs. This serves as an incentive for MTurk participants to take the time to complete HITs adequately. We encourage readers to consult Mason and Suri (2012) for more specific description of techniques and suggestions for conducting research with crowd-sourced subjects.

In summary, we see few disadvantages to using crowd-sourced labor to expand the stretch of one's behavioral research. Many tasks can be deployed online as easily as with traditional testing at the researcher's home university or community. Naturally, care and inventiveness may be required to transfer tasks to a successful online format, and it is also the case that future research will need to involve different types of behavioral tasks and use other sources of online samples. However, the current investigation suggests that the potential for savings of both time and money can be considerable, and the resulting benefits of diversity speak for themselves. Although the diversity distribution with MTurk is still not perfect (for instance, unless multiple language versions are created, all participants still must speak English), it is nonetheless superior to the samples commonly utilized in face-to-face studies.

However, we hasten to add that we do not advocate abandoning traditional testing in the lab—far from it. We leave readers with the suggestion that a face-to-face, in-person sample is a critical companion group to include whenever one is adapting a behavioral task for online completion. If differences should emerge between online and in-person test takers, this attention-grabbing outcome alerts researchers to look more closely at their methods and investigate whether aspects of the mode of presentation or differences in the participants themselves are driving the unequal responding. In this way, scholars can build toward a more complete and nuanced understanding of human behavior and cognition across people and place.

References

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

Buhrmester, M., Kwang, T., & Gosling, S. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3–5. http://dx.doi.org/10.1177/1745691610393980.

Casler, K., Hackett, E., & Bickel, L. (2013). Perceiving parts and fixing functions: Why an object's function is less than the sum of its parts, submitted for publication.

Davidov, E., & Depner, F. (2011). Testing for measurement equivalence of human values across online and paper-and-pencil surveys. *Quality and Quantity: International Journal of Methodology, 45*, 375–390. http://dx.doi.org/10.1007/s11135-009-9297-9.

Eriksson, K., & Simpson, B. (2010). Emotional reactions to losing explain gender differences in entering a risky lottery. *Judgment and Decision Making, 5*, 159–163.

Gardner, R., Brown, D., & Boice, R. (2012). Using Amazon's Mechanical Turk website to measure accuracy of body size estimation and body dissatisfaction. *Body Image, 9*, 532–534.

Gosling, S., Vazire, S., Srivastava, S., & John, O. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist, 59*(2), 93–104. http://dx.doi.org/10.1037/0003-066X.59.2.93.

Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61–83. http://dx.doi.org/10.1017/S0140525X0999152X.

Henry, P. J. (2008). College sophomores in the laboratory redux: Influences of a narrow data base on social psychology's view of the nature of prejudice. *Psychological Inquiry, 19*(2), 49–71. http://dx.doi.org/10.1080/10478400802049936.

Horton, J., Rand, D., & Zeckhauser, R. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics, 14*, 399–425. http://dx.doi.org/10.1007/s10683-011-9273-9.

Joinson, A. (1999). Social desirability, anonymity, and internet-based questionnaires. *Behavior Research Methods, Instruments, and Computers, 31*, 433–438. http://dx.doi.org/10.3758/BF03200723.

Joinson, A., Woodley, A., & Reips, U. (2007). Personalization, authentication and self-disclosure in self-administered Internet surveys. *Computers in Human Behavior, 23*, 275–285. http://dx.doi.org/10.1016/j.chb.2004.10.012.

Kraut, R., Patterson, V., Lundmark, M., Kiesler, S., Mukophadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist, 53*, 1017–1031.

Mason, W., & Watts, D. (2009). Financial incentives and the "performance of crowds". In: *HCOMP'09: Proceedings of the ACM SIGKDD Workshop on Human Computation*, (pp. 77–85).

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

Oppenheimer, D., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867–872. http://dx.doi.org/10.1016/j.jesp.2009.03.009.

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

Sears, D. (1986). College sophomores in the lab: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology, 51*, 515–530. http://dx.doi.org/10.1037/0022-3514.51.3.515.

Simsek, Z., & Viega, J. (2001). A primer on internet organizational surveys. *Organizational Research Methods, 4*, 218–235. http://dx.doi.org/10.1177/109442810143003.

Smyth, J., & Pearson, J. (2011). Internet survey methods: A review of strengths, weaknesses, and innovations. In: M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the Internet: Advances in applied methods and, research strategies* (pp. 11–44).

Suri, S., & Watts, D. (2011). Cooperation and contagion in web0based, networked public goods experiments. *PLoS ONE, 6*(3), e16836. doi: 10.1371/journal.pone.0016836.

Tuten, T., Urban, D., & Bosnjak, M. (2002). Internet surveys and data quality: A review. In B. Batinik, U. Reips, & M. Bosnjak (Eds.), *Online social sciences* (pp. 7–26). Ashland, OH, US: Hogrefe & Huber Publishers.