Journal of Personality and Social Psychology

The Pitfall of Experimenting on the Web: How Unattended Selective Attrition Leads to Surprising (Yet False) Research Conclusions

Haotian Zhou and Ayelet Fishbach

Online First Publication, June 13, 2016. http://dx.doi.org/10.1037/pspa0000056

Journal of Personality and Social Psychology

0022-3514/16/$12.00 http://dx.doi.org/10.1037/pspa0000056

The Pitfall of Experimenting on the Web: How Unattended Selective Attrition Leads to Surprising (Yet False) Research Conclusions

Haotian Zhou and Ayelet Fishbach
University of Chicago

The authors find that experimental studies using online samples (e.g., MTurk) often violate the assumption of random assignment, because participant attrition—quitting a study before completing it and getting paid—is not only prevalent, but also varies systemically across experimental conditions. Using standard social psychology paradigms (e.g., ego-depletion, construal level), they observed attrition rates ranging from 30% to 50% (Study 1). The authors show that failing to attend to attrition rates in online panels has grave consequences. By introducing experimental confounds, unattended attrition misled them to draw mind-boggling yet false conclusions: that recalling a few happy events is considerably more effortful than recalling many happy events, and that imagining applying eyeliner leads to weight loss (Study 2). In addition, attrition rate misled them to draw a logical yet false conclusion: that explaining one's view on gun rights decreases progun sentiment (Study 3). The authors offer a partial remedy (Study 4) and call for minimizing and reporting experimental attrition in studies conducted on the Web.

Keywords: dropout rate, selective attrition, Web-experiment, Mechanical Turk

The impact of Internet technology on how experimental psychologists conduct their research is undeniable. In particular, the field of social psychology has witnessed an exodus to cyberspace, characterized by an explosion in the popularity of the Web-experiment method. Running experiments online has increased in recent years, catalyzed by the advent of crowdsourcing online labor markets such as Amazon's Mechanical Turk (MTurk) and user-friendly data-collection Web apps such as Qualtrics. To quantify social psychologists' growing reliance on the Web-experiment method, we selected two of the field's leading empirical journals—*Journal of Personality and Social Psychology* (*JPSP*) and *Personality and Social Psychology Bulletin* (*PSPB*)—and tallied the percentage of papers that included at least one MTurk study during each semiannual epoch over the three-and-a-half-year run from January 2012 to June 2015. We also included the data concerning *Psychological Science*—the leading nonspecialty psychology journal—over the same period as the benchmark. The results are plotted in Figure 1. Clearly, in comparison to other subfields within psychology, the Web is on the way to becoming a major data source for social psychology.

Haotian Zhou and Ayelet Fishbach, Booth School of Business, University of Chicago.

The authors thank Kaushal Addanki, Daniel Walco, Franklin Shaddy, Janina Steinmetz, Jessica Bregant, Juliana Schroeder, Kaitlin Woolley, Natalie Wheeler, Oliver Sheldon, Sarah Molouki, Shoham Choshen-Hillel, Yanping Tu, Nira Liberman, Minjung Koo, Tal Eyal, Jane Risen, Alex Shaw, Ed O'Brien, and Eugene Caruso for sharing their data and insights with us.

Correspondence concerning this article should be addressed to Haotian Zhou or Ayelet Fishbach, Booth School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637. E-mail: zhouhaotian@uchicago.edu or ayelet.fishbach@chicagobooth.edu

This exodus to cyberspace is unsurprising considering the many advantages the Internet allegedly offers over the physical lab or the field. In an early review of online research, Reips (2000), who first coined the term *Web experiment*, identified no fewer than 10 benefits associated with running psychological experiments online, among which are easy access to larger and more diverse samples and the cost saving of lab space, person-hours, equipment, and administration. In fact, these days MTurk often allow researchers to complete data collection in a matter of hours while incurring minimal cost (Goodman, Cryder, & Cheema, 2013; Paolacci & Chandler, 2014).

The Manageable Nuisances

The long list of advantages Web experiments boast does not mean they are trouble free. In the same early review, Reips (2000) called attention to a set of drawbacks for which psychologists seeking to adopt the Web-experiment method should be on the lookout. Recently, these drawbacks have been put under close scrutiny in the context of MTurk because MTurk is becoming the primary venue for online research.

For instance, researchers have examined the extent to which MTurk workers are honest, attentive, or motivated (Buhrmester, Kwang, & Gosling, 2011; Goodman et al., 2013; Hauser & Schwarz, 2016; Paolacci & Chandler, 2014; Ross, Irani, Silberman, Zaldivar, & Tomlinson, 2010). So far, the verdict is mostly positive. MTurk's participant pool, though not perfect, is regarded as a serviceable replacement for more traditional data sources, such as college students and other similar convenience samples (Paolacci & Chandler, 2014). In fact, after taking into account its low cost and high efficiency, MTurk might even be considered preferable to the traditional venues.

Moreover, the few documented drawbacks of MTurk samples are not particularly bothersome relative to the lab samples. For

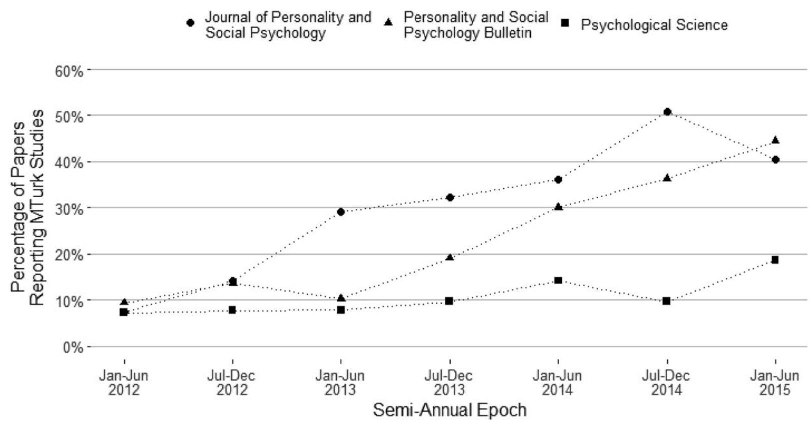

Figure 1. The percentage of empirical papers that reported at least one MTurk study during each semi-annual epoch over an extended period of time for *Journal of Personality and Social Psychology*, *Personality and Social Psychology Bulletin*, and *Psychological Science*.

instance, although MTurk samples are not representative of the general population, they are nonetheless more diverse than college student samples (Berinsky, Huber, & Lenz, 2012). Another caveat regarding MTurk is the issue of varying sample composition. Because MTurk workers are free to choose the studies in which they want to participate, different studies end up sampling different subpopulations within MTurk (Paolacci & Chandler, 2014). Clearly, this caveat also applies to lab studies, in which participation is voluntary and participants self-select to studies. A third major concern regarding MTurk data is the issue of cross-contamination: participants enroll in studies using procedures similar or identical to those they have experienced before (Horton, Rand, & Zeckhauser, 2011). Admittedly, preventing participants from working on related studies is easier when they are recruited from the campus or local community, yet cross-contamination can still sneak in under different guises. For instance, the students might have learned about similar studies in their courses, especially when a course-credit requirement mandates their participation. In addition, students and local residents are more likely to talk to each other about the studies they recently completed than MTurk workers who are mostly strangers.

In short, extant research suggests little grounds for being more critical of experiments conducted online than those conducted in physical labs. However, this complacency might be less justified than many researchers would like to believe. In the present article, we argue that a pernicious problem that tends to disproportionally afflict Web experiments—participant attrition—has been largely overlooked.

The Problem With Dropouts

Experimental studies are conducted to test causal theories or mechanisms. The internal validity of an experiment is predicated on successful random assignment, which allows for unambiguous causal inference by statistically removing myriads of confounds. Whenever participants drop out of different conditions of an experiment for different reasons—that is, condition-dependent or selective dropout—confounds of the experimental manipulations are likely to be introduced, which would compromise the experiment's internal validity and expose the experimenter to the risk of drawing false conclusions about the causal effect. Evidently, a high dropout rate opens the door for condition-dependent dropout to creep in.

Indeed, in his early review, Reips (2000) expressed the concern that although the majority of the drawbacks of Web experiments can be addressed more or less satisfactorily through some ingenuity and technical knowhow on the part of the researchers, one problem is not only extremely pernicious but also lacks agreed-upon solutions, that is, a high dropout or attrition rate (see also a report on online research by the American Psychiatric Association's Board of Scientific Affairs' Advisory Group, Kraut et al., 2004). Long before MTurk started to gain popularity, Musch and Reips (2000) reported the mean dropout rate of a typical Web experiment is around 34% and, depending on the specifics of a given study, the rate can range from 1% to 87%.

Given the drastic change in Internet technology as well as the way Web experiments are conducted has undergone since Musch and Reips' research (2000) more than 15 years ago, one cannot be faulted for questioning the pertinence of their data. To gauge the prevalence of participant attrition on MTurk, we collected a convenience sample of 88 Web studies conducted on MTurk that contained a total of 22,260 responses (i.e., individual participants). We obtained these data by emailing 17 social psychology researchers affiliated with the University of Chicago, inviting them to share the MTurk studies they had conducted in the past two years with us. By granting us direct access to their Qualtrics surveys in which the studies were embedded, we were able to collect information about dropouts in these studies (notably, those researchers did not check for dropouts, and the majority of them were unaware of how to obtain such information).

A simple tally of these MTurk studies revealed that whereas only one study (1.13%) in this sample was free of attrition (i.e., 0% dropout rate), more than 20% of the 88 studies had a dropout rate that exceeded 30%. To put these numbers into perspective, we also obtained a convenience sample of 82 single-session lab studies conducted by 46 different research teams at the University of Chicago during the present academic year (n = 7,861). In sharp contrast to the MTurk samples, 79 of the 82 (96%) lab studies were free of attrition, and of the three studies that did encounter dropouts, the highest dropout rate was 4.74%.

Participants of Web experiments appear to be much less committed to completing their tasks than those of lab experiments for at least two reasons. First, the sunk cost on the part of the participants is much lower on the Internet than in the lab. To try out a Web experiment, only a reliable Internet connection and a few clicks of the mouse are necessary, whereas trying out a study in a lab typically requires the participants to make an appointment with the lab, commute to the lab at the scheduled time, and then possibly wait in the waiting area until the experimenter is ready to interact with them. Second, the social cost associated with quitting a study is much lower on the Internet than in the lab. The highly impersonal and anonymous nature of cyberspace greatly attenuates the awkwardness and embarrassment participants might experience when backing out of a study in which they have consented to participate.

Where Have All the Dropouts Gone?

The preceding discussion might strike the reader as particularly innocuous. After all, researchers should know of the potential havoc participant attrition might wreak and should be cognizant of the norm of monitoring and reporting attrition when reporting a study, Web experiment or not. Further, the empirical fact that the attrition is higher online than offline is hardly likely to turn heads. The stark contrast between this halcyon backdrop and what we uncovered in an exploratory project, however, is surprising.

Assisted by a text-analysis program coded in R, we combed all the empirical papers published in *JPSP* in 2014 and 2015. For each calendar year, we first identified the individual studies (instead of papers) that were run on MTurk. If a study was MTurk-based, we then coded whether it disclosed any information concerning dropouts by scanning its methods and results sections.[1]

In 2014, we identified a total of 147 studies conducted on MTurk. Of these, only two mentioned information pertaining to participant attrition. The situation was not much different in 2015. Specifically, of a total of 142 MTurk studies in 2015, only four included information concerning dropouts. What these data could potentially suggest is that, at least when conducted on MTurk, Web experiments are generally free of the issue of participant attrition. Yet this rather rosy picture directly contradicted the rather grim image Musch and Reips (2000) portrayed, as well as that emerging out of our convenience sample of 88 MTurk studies.

The most likely reason dropouts are rarely disclosed in published papers is that authors are simply unaware that attrition is happening in their studies. Indeed, researchers in our convenience sample were unaware they had attrition. This lack of awareness can be traced to three main sources, all of which have to do with the fact that participant attrition is less visible on the Internet than in physical labs. First, keeping track of participant attrition requires no forethought or special plans in physical labs, because quitters have to notify the experimenter once they decide to discontinue. By contrast, quitters on the Web normally do not contact the researchers when they drop out. As a result, unless the researchers take certain proactive measures, they might never find out that attrition has occurred. For instance, some applications that help researchers creating Web-based experiments (e.g., jsPsych) by default save participant-generated data locally on the participants' computers and only post the data to the server when the participants get to the finish line. Because researchers can only access data that are on the server, partial data from the dropouts would be lost, thereby creating an illusion of zero attrition and successful random assignment.

Second, these days a large number of behavioral scientists, especially those in academia, tend to implement their online studies in Qualtrics, which has become the behavioral scientists' tool of choice for data collection on the Web. Indeed, our data, which we complied through both Google and Yahoo search engines, show that each of the 108 American universities classified as RU/VH (Research Universities/Very High Research Activity) in the Carnegie Classification of Institutions of Higher Education are customers of Qualtrics. In addition, at least 158 non-RU/VH universities in the United States alone use Qualtrics to collect data online. Thankfully, Qualtrics automatically posts participant data to the server at multiple points during the course of a study instead of only at the finish line. Yet, using Qualtrics, partial responses from someone who drops out of a study would, by default, be available only after a week has passed since she first loaded the study to her browser. Because MTurk often fulfills the requested participant quota (i.e., predetermined sample size) within a span of a few hours, the researchers who typically do not wait for a week to download and analyze the data would only see complete responses in their data files unless they force the partial responses into the data files by deactivating the survey.[2]

Third, because quitters do not count toward requested participant quotas on MTurk, the researchers typically find the number of complete responses exactly matches or even slightly exceeds the requested quota (as in the case of all experiments reported in the present research). Coupled with the fact that MTurk quitters have zero impact on the researchers' budget because they are not getting paid, the notion that many researchers never suspect that the issue of high participant attrition plagues their studies is unsurprising.

Clearly, if researchers are unaware of dropouts, published studies might unknowingly suffer from compromised internal validity. However, a less pessimistic alternative account for the near ab-

[1] The text-analysis program first looked for an extensive set of word stems, words, and phrases that might signal the disclosure of attrition information in the text of that MTurk study. The word stems included in the search set are attrit, incomplet exclu, non-respon, and nonrespon. The words included in the search set are quit, drop-out, dropout, unfinished, refuse, and skip. The phrases included in the search set are drop out, decline to respond/complete/finish/answer, fail to respond/complete/finish/answer, and do not respond/complete/finish/answer. The program searched for all the inflections of the verbs and nouns listed above. For instance, in the case of fail to respond, the program searched for *fail to respond*, *failed to respond*, *failing to respond*, and *fails to respond*. Then, if a study was a hit by the text-analysis program, Haotian Zhou read the text again to verify whether that study actually disclosed attrition information.

[2] Deactivating a Qualtrics survey not only takes it offline so that people will not be able to take the survey, but also closes and records any partial responses so that data from the dropouts will be immediately available in the dataset downloaded from Qualtrics. Researchers should take the following two steps to deactivate a survey:

1. Log into the Qualtrics account and in the My Surveys tab, click the green checkmark next to the survey name.
2. A popup window will appear. The researcher should type "close" in the text field of the window and then click Deactivate Survey button at the lower right corner of the window.

sence of attrition in published papers might exist. Perhaps the studies that made it to the press tended to be better designed and implemented than the studies in our convenience sample.

Present Research

We conducted this research to explore the prevalence and consequences of attrition rates and suggest some remedies. Our first goal was to test for dropout rates in several social psychology paradigms commonly used online. In particular, we sought to discern if published paradigms are indeed less susceptible to high attrition as one might be tempted to conclude after perusing MTurk-based studies reported in journal papers, which rarely mention dropouts. Accordingly, in Study 1, to ascertain the size of attrition, we replicated the procedures of several MTurk studies that made it to the press but did not report attrition rates. We were specifically interested in paradigms that are cognitively taxing (e.g., ego-depleting) and those involving writing an open essay (e.g., recall low- vs. high-power episode). Using these common manipulations, we should be concerned if dropout rates are substantial (e.g., 20% or above), because the threat of compromised internal validity due to selective attrition—participants self-selecting to opt out of an experiment for reasons related to the condition they were assigned to—increases with dropout rates. Suppose only 1% of participants select to drop out from a two-condition experiment. Even if all the dropouts are concentrated in one condition, thereby constituting selective attrition, their impact on the experiment's internal validity is likely to be negligible.

Our second goal was to test whether attrition rates in Web experiments conducted on MTurk are some innocuous nuisance or, alternatively, capable of rendering a between-subjects experiment worthless. Specifically, in Study 2, we tested whether substantial dropout rates can yield absurd research conclusions that clearly defy common sense and psychological theories. We assumed that such outlandish conclusions could result from condition-dependent dropout. For example, if an ease-of-retrieval study asks people to recall many versus few happy events, and unhappy people, for whom the recall task is hard, drop out in the "many" condition, we could find that recalling many events is reported to be no less easy than recalling only a few.

Our third goal was to test whether, in addition to outlandish inferences, selective attrition in Web experiments can also lead to apparently sensible conclusions that are consistent with extant literature and therefore more likely to beguile the untrained eyes. We test this possibility in Study 3.

Finally, in Study 4, we turned our attention to potential remedies to the insidious problem of a high attrition rate on the Web. Specifically, we devised and evaluated a nearly costless strategy involving three elements: (a) prewarning (e.g., telling participants the experiment would include an open-ended questions), (b) personalization (e.g., asking for individuating information), and (c) appealing to conscience (e.g., explaining that dropping out could affect the quality of data we receive). Notably, such strategies would only provide partial remedies. In addition, we suggest in our General Discussion that researchers should rethink the practice of using certain manipulations in Web experiments and call for more openness in monitoring and reporting dropouts.

Study 1: Attrition Rates Caused by Common Paradigms

We designed Study 1 to determine whether dropout rates are substantial in social psychology paradigms used online and whether they could vary across conditions (i.e., a higher percentage of participants quit one condition than the other condition). When participants self-select to opt out of an experiment, even when the dropout rates are comparable across conditions, attrition might be condition dependent (e.g., the personal characteristics of the dropouts might vary across conditions). Further, when the dropout rates are not comparable across the different conditions, attrition is almost certainly condition dependent.

We selected six popular paradigms in social psychology with widely adopted manipulation procedures: terror management, construal level, power, regulatory focus, ego depletion, and political attitude. We then identified published papers that used these paradigms and met the following two criteria: (a) the paradigm in the paper was implemented in at least one MTurk experiment, and (b) the paper provided sufficient information regarding the procedure of the experiment so that it could be unambiguously replicated. Notably, none of these six source papers reported information regarding dropouts in the experiments we selected to replicate.

Method

For each of the six paradigms that we chose to replicate, we created an MTurk HIT (Human Intelligence Task), which in our case was a Qualtrics survey that an MTurk Worker can work on and collect a reward for completing. We described all six HITs as "a simple anonymous survey that takes about 5 minutes to complete." For each HIT, we requested 100 participants and paid participants who completed the study a fixed 50 cents in compensation. We determined the 10-cents-per-minute rate in accordance with the recommendation by the Guideline for Academic Requesters Project, a joint effort by academic researchers and the MTurk-worker community. We allowed only MTurk workers residing in either the United States or Canada to take the HITs.

In each replication experiment, participants first clicked a button to indicate their consent. When computing dropout rates, we took into account only participants who consented. These participants were then taken to a second page where they were invited to complete the task intended as an experimental manipulation in the original experiment. In our replication experiments, we only had participants complete the experimental manipulation (i.e., we skipped all the other steps in the original procedures, including manipulation checks and dependent variables). The manipulation task was followed by a very short demographic questionnaire. However, the participants did not know beforehand that the survey they consented to consisted of only the manipulation task and the demographic questionnaire.

Results

Table 1 lists the sources of the six MTurk experiments we selected to replicate, the synopsis of the experimental manipulations implemented by each study, and the condition-wise dropout rates. We determined dropout rates by dividing the number of participants who were assigned to a given condition and completed

Table 1
Condition-Wise Dropout Rates of the Six Replication Experiments in Study 1, as Well as Their Sources and Manipulation Tasks

Replication experiment	Condition	Dropout rates
A. Terror management (Study 2 in Wisman, Heflick, & Goldenberg, 2015)	*Mortality salience*: Writing down thoughts and feelings about one's own death	30.7%
	Control: Writing down thoughts and feelings about physical pain	34.6%
B. Construal Level (Experiment 2 in Henderson, 2013)	*Abstract construal*: Describing why one wants to accomplish three goals in one's life	44.0%
	Concrete construal: Describing how one is to accomplish three goals in one's life	38.2%
C. Power (Study 5 in May & Monga, 2014)	*Powerful*: Recalling a past episode where one was in a powerful position	34.1%
	Powerless: Recalling a past episode where one was in a powerless position	33.8%
D. Regulatory focus (Study 4 in Woltin & Yzerbyt, 2015)	*Promotion-focus*: Writing about one's aspirations	29.9%
	Prevention-focus: Writing about one's obligations	34.3%
E. Ego depletion (Study 1 in Yam, Chen, & Reynolds, 2014)	*Ego-depletion*: Writing a 100-word paragraph without using letters A and N	77.6%
	No-depletion: Writing a 100-word paragraph without using letters X and Y	22.8%
F. Elaboration mode (Experiment 2 in Fernbach, Rogers, Fox, & Sloman, 2013)	*Reason*: Enumerating reasons for one's attitude toward certain public policies	33.7%
	Mechanism: Explaining the mechanisms by which the same public policies work	58.8%

the entirety of their task by the number of those who were assigned to the same condition and at least gave their consent. As shown, we observed high dropout rates (>20%) in all conditions of all experiments. We next provide the details of each individual experiment.

Experiment 1A: Terror management. A total of 156 MTurk workers consented to take part in this experiment before the MTurk HIT quota was fulfilled. They were randomly assigned to write about either their own death (i.e., mortality salience) or their own pain (i.e., control). Fifty-one of these participants dropped out of the survey once they learned what their first task (i.e., the experimental manipulation) entailed: 30.7% (23/75) in the mortality-salience condition and 34.6% (28/81) in the control condition. The 105 participants who saw their task through took approximately 2.97 min (i.e., median) to complete the survey. Among these 105 nonquitters ($M_{\text{age}} = 35$), 52.38% were males.

Experiment 1B: Construal level. A total of 173 MTurk workers consented to take part in this experiment before the MTurk HIT quota was fulfilled. They were randomly assigned to write about why they wanted to (i.e., abstract construal) or how they planned to (i.e., concrete construal) accomplish certain self-nominated life goals. Seventy-one of these participants dropped out of the survey once they learned what their first task (i.e., the experimental manipulation) entailed: 44.0% (37/84) in the abstract-construal (i.e., why) condition and 38.2% (34/89) in the concrete-construal (i.e., how) condition. The 102 participants who finished the study took approximately 4.77 min (i.e., median) to complete the survey. Among these 102 nonquitters ($M_{\text{age}} = 34$), 57.84% were males.

Experiment 1C: Power. A total of 156 MTurk workers consented to take part in this experiment before the MTurk HIT quota was fulfilled. They were randomly assigned to recall a memory in which they were in a powerful position or in a powerless position. Fifty-three of these participants dropped out of the survey once they learned what their first task (i.e., the experimental manipulation) entailed: 34.1% (28/52) in the powerful condition and 33.8% (25/74) in the powerless condition. The 103 participants who finished the study took approximately 3.50 min (i.e., median) to complete the survey. Among these 102 nonquitters ($M_{\text{age}} = 37$), 48.54% were males.

Experiment 1D: Regulatory focus. A total of 147 MTurk workers consented to take part in this experiment before the MTurk HIT quota was fulfilled. They were randomly assigned to write about their hopes and aspirations (i.e., promotion-focus) or duties and obligations (i.e., prevention-focus). Forty-seven of these participants dropped out of the survey once they learned what their first task (i.e., the experimental manipulation) entailed: 29.9% (23/77) in the promotion-focus condition and 34.3% (24/70) in the prevention-focus condition. The 100 participants who finished the study took approximately 4.38 min (i.e., median) to complete the survey. Among these 100 nonquitters ($M_{\text{age}} = 35$), 48.0% were males.

Experiment 1E: Ego depletion. A total of 208 MTurk workers consented to take part in this experiment before the MTurk HIT quota was fulfilled. They were randomly assigned to write a 100-word paragraph without using the letters A and N (i.e., depletion) or without using the letters X and Y (i.e., no-depletion). One hundred six of these participants dropped out of the survey once they learned what their first task (i.e., the experimental manipulation) entailed. The dropout rate in the depletion condition, 77.6% (83/107), is significantly higher than no-depletion condition, 22.8% (23/101), $\chi^2 = 62.43$, $p < .01$. The 102 participants who finished the study took approximately 7.42 min (i.e., median) to complete the survey. Among these 102 nonquitters ($M_{\text{age}} = 40$), 45.1% were males.

Experiment 1F: Elaboration modes. A total of 189 MTurk workers consented to take part in this experiment before the MTurk HIT quota was fulfilled. They were randomly assigned to either elaborate on the reasons for their positions on two public policies (i.e., reason condition) or elaborate on the mechanisms via which the same two policies work (i.e., mechanism condition). Eighty-five of these participants dropped out of the survey once they learned what their first task (i.e., the experimental manipulation) entailed. The dropout rate in the reason condition, 33.7% (35/104), is significantly lower than mechanism condition, 58.8% (50/85), $\chi^2 = 11.97$, $p = .018$. The 104 participants who finished the study took approximately 5.7 min (median) to complete the survey. Among these 104 nonquitters (M_{age} = 36), 51.9% were males.

Discussion

Despite adhering to the accepted fair-payment practice, we find that in all six replications, overall dropout rates exceed 30% (31.9% to 51% across studies). As previously mentioned, none of the six source studies, which we replicated, disclosed information about participant attrition (published MTurk studies rarely do); however, our data suggest that any study using these paradigms was likely susceptible to a high attrition rate.

Dropouts undermine internal validity whenever they are condition dependent. One diagnostic sign of condition-dependent attrition is differential dropout rates across conditions. Thus, our results clearly show that researchers should not administer ego-depletion and elaboration-mode (i.e., reason- vs. mechanism-based elaboration) manipulations online, because these manipulations most certainly are conducive to selective attrition. Specifically, the remaining 22.4% in the depletion (41.2% in the mechanism) condition likely had different personal characteristics (e.g., more motivated, educated) than the remaining 77.2% in the no-depletion (66.3% in the reason) condition. Indeed, in a follow-up study that used the same ego-depletion manipulation (n = 272), we found that among nonquitters, those assigned to the depletion condition self-reported less fatigue prior to manipulation (M = 1.85, SD = 0.83, 95% CI = [1.61, 2.08]) than those in the no-depletion condition (M = 2.19, SD = 0.92, 95% CI = [2.02, 2.36]), $t(110)$ = 2.40, p = .02, Cohen's d = 0.39. These results indicate that at the outset of the study, participants in the depletion condition as a whole were less fatigued than those in the no-depletion condition, which might counteract the intended effect of the manipulation, namely, making the former group more mentally fatigued than the latter.

More importantly, attrition can still be condition dependent even if a similar number of participants drop out in each condition (Birnbaum & Mellers, 1989; Reips, 2000). For example, the remaining 65.9% in the power condition might have had different life experiences (e.g., higher positions) than the remaining 66.2% participants in the powerless condition. Indeed, to prove that the attrition observed in an experiment is independent of the experimental manipulations is essentially infeasible because enumerating all the reasons for quitting that are correlated with experimental manipulation, let alone ruling them all out, is practically impossible. In Study 2, we presented a pair of severe cases of internal-validity violation on MTurk, likely due to selective dropout and, in one case, the dropout rates were similar across conditions.

Study 2: Arriving at Surprising (yet False) Research Conclusions

To explore whether participant attrition on MTurk can indeed pose a threat to the internal validity of Web experiments, we conducted two between-subjects experiments on MTurk. In each experiment, we first administered an experimental manipulation that could potentially lead to condition-dependent attrition, thereby introducing a certain confound. We then measured a dependent variable (DV) that, in theory, should have one type of relation with the attrition-induced confound (if selective dropout did occur) but a qualitatively different type of relation with the actual experimental manipulation. As a result, if we observe that the DV varies across the two conditions in a manner that was predicted by the confound-DV relationship, rather than manipulation-DV relationship, we could infer that the internal validity of the experiment had been compromised by condition-dependent dropout.

Specifically, in Experiment 2A, we predicted that an experiment that assigns participants to recall many versus few happy events would result in a biased sample consisting of mainly happy people in the many-events condition, because happy events come to mind easily for these people, whereas the less-happy people in this condition would have to quit this difficult task. As a result of this experimental attrition, recalling many happy events could feel easier than recalling fewer happy events. In Experiment 2B, we predicted that an experiment that assigns participants to imagine applying eyeliner (vs. applying aftershave cream) would end up with a sample that is disproportionally female. As a result, participants assigned to imagine applying eyeliner would report weighing less than those assigned to imagine applying aftershave. We summarize the results of the two experiments in Table 2 before describing each study in detail separately.

Table 2
Summary of the Results of the Two Experiments in Study 2

Experiment	Condition	Dropout rates	Main results
A. Can doing more feel like less work?	*Many*: Listing 12 happy events from the past year	69.0%	*Rated recall difficulty*: M = 2.74, SD = 2.02
	Few: Listing 4 happy events from the past year	26.0%	*Rated recall difficulty*: M = 3.97, SD = 2.90
B. Can imagining applying eyeliner help one lose weight?	*Eyeliner*: Describing how applying versus not applying eyeliners would make one feel differently	32.4%	*Reported body weight*: M = 159.64 lbs., SD = 46.78
	Aftershave cream: Describing how applying versus not applying aftershave cream would make one feel differently	24.3%	*Reported body weight*: M = 182.08 lbs., SD = 44.78

Experiment 2A: Can Doing More Feel Like Less Work?

All else being equal, recalling 12 happy events from the past year requires more effort than recalling four such events (e.g., Schwarz et al., 1991). Therefore, participants who recall 12 events should report experiencing a higher level of difficulty than those recalling four events. On the other hand, remembering 12 happy events from the previous year would be difficult for people who either did not lead a very happy life in the past year or have a higher threshold for reporting happiness, inducing these people to quit. Thus, our manipulation could potentially lead to selective dropout such that the sample that was required to recall 12 events would consist of more people who had many happy life events in the past year than the sample that recalled four events, thereby introducing a confound. Because of this confound, participants in the 12-event condition might end up reporting experiencing a lower level of difficulty than those in the four-event condition.

Method.

Participants. We requested 100 participants on MTurk and paid them a fixed 40 cents in compensation. We allowed only MTurk workers residing in either the United States or Canada to participate in this study.

Procedure. We launched a Web experiment on MTurk, advertised as "an anonymous survey consisting a few simple tasks." The experiment was programmed in Qualtrics and randomly assigned participants who accessed the survey to either the *few* or the *many* condition. The experimental manipulation stipulated that participants in the few condition should recall and list four happy events from the past 12 months, whereas those in the many condition should list 12 events. As the DV, all participants reported how difficult they found the memory task to be on a 7-point scale ranging from 1 (*not difficult at all*) to 7 (*extremely difficult*).

Results. A total of 196 MTurk workers consented to take part in this experiment before the MTurk HIT quota was fulfilled. Ninety-four of these participants dropped out of the survey once they learned what their first task (i.e., the experimental manipulation) entailed. The quitters neither completed the manipulation nor the DV. The dropout rate in the many condition, 69% (69/100), is significantly higher than in the few condition, 26% (25/96), $\chi^2 = 36.22$, $p < .01$. The 102 participants who finished the study took approximately 4.13 min (i.e., median) to complete the survey. Among these 102 nonquitters ($M_{\text{age}} = 36$), 47.1% were males.

Because dropout rates differed across condition in this experiment, we can infer that the attrition was selective, thereby violating random assignment. Turning to the dependent variable, we further find that participants in the many condition regarded the memory task to be less difficult than their counterparts in the few condition ($M = 2.74$, $SD = 2.02$, 95% CI = [2.00, 3.48] vs. $M = 3.97$, $SD = 2.90$, 95% CI = [3.29, 4.66]), $t(81) = 2.25$, $p = .016$, Cohen's $d = 0.46$. This pattern is more consistent with the confound–DV than the manipulation–DV relation, corroborating the argument that the interval validity of the current experiment was compromised. Yet, this seemingly paradoxical finding might dumbfound someone who ignores participant attrition altogether.

Experiment 2B: Can Imagining Applying Eyeliner Help one Lose Weight?

All else being equal, imagining using eyeliner versus aftershave cream should have no impact on body weight. We predicted, however, that imagining applying eyeliner would be difficult or even aversive for average adult males, inducing them to quit. Such selective dropout would mean that the sample in the eyeliner condition would consist of more females than the aftershave cream condition, thereby introducing a confound. Given that females generally weigh less than males, the self-reported body weights might be lower in the eyeliner condition than in the aftershave cream condition.

Method.

Participants. We requested 100 participants on MTurk and paid them a fixed 50 cents in compensation. We allowed only MTurk workers residing in either the United States or Canada to participate in this study.

Procedure. We launched a Web experiment on MTurk, advertised as "an anonymous survey consisting a few simple tasks." The experiment was programmed in Qualtrics and randomly assigned participants to either the eyeliner or the aftershave cream condition. The experimental manipulation stipulated that participants in the eyeliner condition should describe how applying versus not applying eyeliner could make them feel differently, whereas those in the aftershave cream condition answered a parallel question about aftershave cream. As the DV, all participants reported their current body weights in pounds.

Results. A total of 144 MTurk workers consented to take part in this experiment before the MTurk HIT quota was fulfilled. Forty-one of these participants dropped out of the survey once they learned what their first task (i.e., the experimental manipulation) entailed. The quitter neither completed the manipulation nor the DV. The dropout rates were comparable across the two conditions: 32.4% (24/74) in the eyeliner condition and 24.3% (17/70) in the aftershave cream condition, $\chi^2 = 1.17$, $p = .883$. The 103 participants who finished the study took approximately 2.23 min (i.e., median) to complete the survey. Among these 102 nonquitters ($M_{\text{age}} = 33$), 64.1% were males.

Because the dropout rates were comparable across the two conditions, we examined how the DV covaried with the conditions to determine if condition-dependent dropout had violated the random assignment. We found that participants in the eyeliner condition reported weighing significantly less ($M = 159.64$ lbs., $SD = 46.78$, 95% CI = [146.35, 172.93]) than participants in the aftershave cream condition ($M = 182.08$ lbs., $SD = 44.78$, 95% CI = [169.73, 194.42]), $t(100) = 2.5$, $p = .01$, Cohen's $d = 0.49$, a pattern that was more consistent with the confound–DV relation than manipulation–DV relation. These results, which clearly defy common sense, indicated that internal validity had been breached likely due to the confound induced by selective dropout. In fact, the eyeliner condition indeed had more females than the aftershave cream condition, 42% versus 30%. Yet, this counterintuitive finding might perplex someone who ignores participant attrition altogether.

Discussion. In a pair of MTurk experiments with rather innocuous manipulations, we again noticed substantial participant attrition (48% in Experiment 2A and 28.5% in Experiment 2B). Moreover, we showed that in both experiments, the attrition was likely to have violated the random assignment by introducing confounds. Notably, this breach of internal validity happened even when dropout rates did not differ significantly across the conditions (Experiment 2B), which suggests that detecting whether or not attrition is condition independent requires more than tallying dropouts in different conditions.

The present study clearly demonstrated that Web experiments on MTurk face the real danger of having selective dropout compromise their internal validity. Yet, if most Web experiments with compromised internal validity due to selective dropout are similar to the two experiments described here (i.e., leading to completely absurd conclusions), the danger in remaining oblivious to the dropout issue might be minimal. After all, if a study makes some preposterous claims that clearly violate common sense (e.g., recalling 4 events is cognitively more taxing than recalling 12 events), it is less likely to sneak past the peer-review process to disseminate misinformation and befuddle the field even if people were temporarily unable to pinpoint the exact culprit. In the next study, we report an experiment suffering from internal-validity breach due to selective dropout. However, unlike those in Study 2, the possibly invalid finding of this experiment seemed to be perfectly sensible and therefore unlikely to be refuted without placing dropout under close scrutiny.

Study 3: Arriving at Potentially Meaningful (yet Likely False) Research Conclusions

In Study 3, we further explored the deleterious effect of high attrition rates on Web experiments. We designed an experiment that in the event of selective dropout could engender invalid evidence for an intuitively justifiable prediction: having to explain one's views on gun rights in a short essay will turn participants against gun rights.

Writing an essay to justify one's view on the issue of gun restriction might cause opponents of gun restriction to examine their preestablished opinions more critically and therefore realize certain limitations on gun ownership in the United States is certainly called for. Thus, one could predict that explaining one's stance on gun restriction will on the whole reduce opposition toward gun restriction. However, articulating reasons for one's position on gun restriction might be a more cognitively demanding task for the opponents of gun restriction than for the proponents because, in general, the former group tends to be less educated than the latter group (Wolpert & Gimpel, 1998). Therefore, we predicted that an experiment in which the experimental condition requires people to explain their stances on gun restriction in writing (vs. a control condition that does not) might end up with a biased sample consisting of a disproportionally high number of gun-restriction proponents in the experimental versus control conditions, because the opponents of gun restriction will quit the study. As a result of this selective attrition, we should be able to show that explaining one's view on gun restriction makes people more supportive of gun restriction even if the writing task per se has no attitudinal effect at all. In other words, if more gun-rights supporters drop the study when they have to explain (vs. not) their views, that fact alone can explain why a writing task seemingly makes people more opposed to gun rights.

Birnbaum and Mellers (1989) pointed out that by checking for any correlations between experimental conditions and relevant demographic or personality variables assessed before the administration of manipulations, researchers would be able to determine whether the participant attrition in their experiments is condition dependent or condition independent. Following this suggestion, we also measured certain demographic covariates prior to manipulation to obtain more direct evidence of selective attrition.

Method

Participants. We requested 160 participants on MTurk and paid them a fixed 50 cents in compensation. We allowed only MTurk workers residing in either the United States or Canada to accept this study.

Procedure. We launched a Web experiment on MTurk, advertised as "an anonymous survey consisting a few simple tasks." The experiment was programmed in Qualtrics and randomly assigned participants to either the writing or the control condition. Prior to the manipulation, we had all participants answer a yes/no question about whether they thought Americans' right to own firearms should be subject to restriction. After this binary question, participants in the writing condition were asked to explain the reason for their positions in the form of a 120-word essay. However, participants in the control condition were exempted from this writing task. Afterward, all participants rated their degree of agreement with an antirestriction (i.e., progun) statement—"Individuals' right to own and possess firearms should remain unfettered by governmental regulation" ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Results

A total of 198 MTurk workers consented to take part in this experiment and completed at least the demographic questionnaire before the MTurk HIT quota was fulfilled. Thirty-six of these participants dropped out of the survey once they learned what the experimental manipulation entailed. The quitters did not complete the manipulation but did answer the premanipulation binary question. The dropout rate in the writing condition, 36% (36/100), was significantly higher than the control condition, 0% (0/98), $\chi^2 = 43.12$, $p < .001$. The 162 participants who finished the study took approximately 2.8 min (median) to complete the survey. Among these 162 nonquitters ($M_{\text{age}} = 33$), 59.26% were males.

We summarize the main results in Table 3. We find that the simple-writing exercise significantly reduced participants' agreement with the antirestriction statement, that is, made them more supportive of gun restriction ($M_{writing} = 2.52$; $SD = 1.65$, 95% CI = [2.10, 2.93] vs. $M_{control} = 3.46$; $SD = 2.04$, 95% CI = [3.05, 3.87]), $t(153) = -3.23$ $p = .001$, Cohen's $d = 0.50$. At first glance, this result seemed to suggest that the antigun movement could potentially advance its cause by encouraging active discourse.

However, when we examined the covariation between the premanipulation gun-restriction stance and the experimental manipulations, we noticed that the percentage of gun-restriction proponents was higher in the writing condition than in the control condition (89.1% vs. 73.5%), $\chi^2 = 4.88$, $p = .027$. Yet, when we examined both nonquitters and dropouts together, the difference in the proportions of gun-restriction proponents between the two

Table 3
Summary of the Results of Study 3: Can Deliberation Increase Support for Restriction on Gun Ownership?

Condition	Dropout rates	Main result
Writing: Explaining the reasons for one's position on gun-restriction	36.0%	Agreement with the anti-restriction (i.e., pro-guns) statement: $M = 2.52$, $SD = 1.65$
Control: Doing nothing	0%	Agreement with the anti-restriction (i.e., pro-guns) statement: $M = 3.46$, $SD = 2.04$

conditions disappeared (81% vs. 73.5%), $\chi^2 = 1.198$, $p = .27$. Together with the fact that attrition rates differed across conditions, this result clearly indicates that the attrition we observed in the present experiment was selective in nature, with antirestriction participants being more likely to quit the *writing* condition. Thus, the ostensibly psychological finding that justifying one's position could increase one's support of gun restriction could also be attributed to a nonpsychological reason that the writing (vs. control) condition contained more gun-restriction proponents thanks to selective dropout.

Discussion

We again obtained evidence of condition-dependent attrition for a common manipulation. Specifically, progun participants were more likely to quit the writing condition than the control condition, resulting in unbalanced samples that rendered unequivocal causal inference impossible. In other words, the seemingly interesting and sensible psychological finding that elaborating on one's gun-related attitude makes one more supportive of gun control could be equally accounted for by the difference in participant compositions across conditions. Therefore, for someone who is not on the lookout for the threat of participant attrition, a Web experiment might reveal some apparently surprising new insights into the working of the human mind that are not about mental processes at all.

Now we have shown that online participants such as MTurk workers often drop out of an experiment they have started at a rate that cannot be safely ignored. More importantly, the reasons for dropping out can be directly related to the experimental manipulation; thus, the next logical step is to consider the possible remedies to this pernicious problem.

Study 4: Seeking Remedy

Making the manipulations in an experiment equal on every possible aspect so that attrition is unlikely to be condition dependent is easier said than done. Often, the aspect of the manipulations that is most likely to result in selective dropout is exactly the aspect that has to vary across the conditions. For instance, researchers need participants in one condition to work on a more taxing task than the other condition when studying ego depletion. Similarly, researchers need participants in one condition to relive more upsetting memories than the other condition when studying social rejection.

Seeking practical solutions, Horton et al. (2011) suggested the researchers could place a long and tedious "warm-up" task before the introduction of manipulations such that those participants who make it to the point of manipulation would have incurred sufficient sunk cost and would therefore be unlikely to quit. Reips (2000) reported that by adding a 20-min "warm-up" task, he was able to reduce the dropout rate at the point of manipulation to 9.7%. The clear drawback of this strategy is that the study is made unusually long—most Web experiments tend to be very short—and the researchers need to compensate participants for useless work (i.e., the "warm-up" task), both of which effectively annul two main benefits of conducting Web experiments— high efficiency and low cost.

Another seemingly viable strategy to reduce the attrition rate is to increase payment. It appears intuitive that by giving a larger reward to participants who finish, a researcher should be able to ensure the majority of the participants would see their task through. However, a recent investigation of online panels by Goritz (2014) showed that monetary rewards increased people's willingness to consent to an online survey (i.e., starting rate) but had no impact on people's willingness to finish the survey once they consented (i.e., dropout rate). Moreover, overpayment could alter the nature of motivation for participants to take part in a Web experiment, which might unexpectedly interact with the experimental manipulations, not to mention that it greatly increases the cost of the study.

Reips (2000) proposed three other nearly costless strategies that could nudge participants who have consented from backing out: (a) prewarning (e.g., telling participants that "you will be writing about the goals you want to accomplish in your life"); (b) personalization (e.g., asking for individuating information such as email); and (c) appealing to conscience (e.g., telling participants that dropping out could affect the quality of data and that science needs good data). However, given this proposal was not tested, here we evaluate its efficacy. Because from a practical point of view, gauging the relative merits of the three strategies Reips proposed is of little value, we combined all three strategies to create a cocktail remedy. We assessed the efficacy of this remedy in reducing dropout rates in a Web-experiment.

Method

Participants. We requested 150 participants on MTurk and paid them a fixed 55 cents in compensation. We allowed only MTurk workers residing in either the United States or Canada to accept this study.

Procedure. We launched a Web experiment on MTurk, advertised as "an anonymous survey consisting a few simple tasks." The experiment was programmed in Qualtrics and randomly assigned participants to either the remedy or the no-remedy condition. On the first page of the survey, participants in the remedy condition read the following message:

> This is an anonymous survey consisting of multiple questions. A few questions are open-ended questions where you need to type a few

> sentences or a short paragraph or two. Many MTurk workers do not like answering open-ended questions and tend to quit a survey once they see such questions. **If a sizable number of people quit a survey halfway, the data quality of that survey would be compromised. However, our research depends on good quality data.** Thus, please make sure you do not mind open-ended questions before taking this survey.

Afterward, they were told to type a short sentence—"I will answer open-ended questions"—into a text field on the same page if they intended to take the survey. By contrast, participants in the no-remedy condition, only saw this message on the first page:

> This is an anonymous survey consisting of multiple questions of various formats. If you'd like to participate, click ≫ to enter the survey.

Essentially, participants in the remedy condition experienced both prewarning (i.e., typing the short sentence) and appealing-to-conscience (i.e., the bolded portion of the message) strategies on the first page. The second page of the survey contained the consent form. Toward the end of the consent form, we asked participants in the remedy condition to enter their MTurk ID, which embodied the personalization strategy, if they agreed to the terms of the consent form. However, participants in the no-remedy condition were not asked to provide their MTurk ID.

All participants then saw the ostensibly first task—a goal-description task—of the survey, which employed the same experimental manipulation in Experiment 1B to induce an abstract-construal mindset. Specifically, participants were asked to indicate three goals they set for themselves and explain why they wanted to accomplish each goal. Once participants completed the goal-description task, they only needed to fill out a short demographic questionnaire before they reached the end of the survey.

Results

A total of 238 MTurk workers consented to take part in this experiment before the MTurk HIT quota was fulfilled. Eighty-eight of these participants dropped out of the survey once they learned the goal-description task. The 150 participants who finished the study took approximately 4.6 min (median) to complete the survey. Among these 150 nonquitters ($M_{\text{age}} = 34$), 46.67% were males. The dropout rate in the remedy condition, 20.6% (21/102), was significantly lower than in the no-remedy condition, 49.3% (67/136), $\chi^2 = 20.57$, $p < .001$. This finding suggests that our cocktail remedy was partially effective at attenuating dropout problems.

To see if our remedy had any side effects, we first analyzed the time nonquitters in either condition spent on the goal-description task. Because the writing-time distributions in both conditions were positively skewed, we applied log transformation before subjecting the data to a *t* test. We found that nonquitters in the remedy condition ($M_{log\text{-}transformed} = 5.29$, $SD = 0.65$, 95% CI = [5.14, 5.43]) spent almost the same amount of time on the goal-description task as their counterparts in the control condition ($M_{log\text{-}transformed} = 5.19$, $SD = 0.54$, 95% CI = [5.06, 5.32]), $t(150) = 0.97$, $p = .34$, Cohen's $d = 0.16$. Then, we examined the number of words nonquitters typed in the goal-description task. Because of the same issue of skewed distribution, we also applied log transformation to word count prior to statistical analysis. We found that nonquitters in the remedy condition ($M_{log\text{-}transformed} = 4.40$, $SD = 0.51$, 95% CI = [4.29, 4.51]) did not write any more words than their counterparts in the control condition ($M_{log\text{-}transformed} = 4.36$, $SD = 0.57$, 95% $CI = [4.22, 4.30]$), $t(150) = 0.47$, $p = .64$, Cohen's $d = 0.08$. Our cocktail remedy does not seem to have affected participants' motivation to work on their task.

Discussion

In the present study, we evaluated the efficacy of a cocktail remedy that combines the three nearly costless strategies Reips (2000) proposed to reduce the dropout rate at the point of manipulation. Overall, adding the remedy decreased the dropout rate by more than half. In addition, this remedy did not induce the participants to treat their task more seriously. Although our cocktail remedy was far from optimal (we still observed a 20% dropout rate), the promising results suggest that, by provoking early dropout, researchers could ensure continued participation after participants make the decision to stay, while incurring little extra time or monetary cost.

One caveat regarding the proposed remedy is that it might have some adverse impact on the external validity of an experiment using it. By promoting early dropout so that only participants who are more likely to see a task through would get to the point of manipulation, we limit the population to those who are more motivated or more serious. As a result, the results can only be generalized to this subpopulation. Therefore, for researchers who are concerned with generalizability of their experimental findings, this remedy might not be appropriate.

General Discussion

This research draws attention to high attrition rates, a problem to which Web experiments are especially vulnerable. We documented high dropout rates using standard social psychological paradigms (Study 1), showed that such attrition rates can result in completely absurd (Study 2) as well as potentially interesting (Study 3) yet false conclusions, and evaluated some remedy (Study 4).

We have little doubt that experimental psychologists are well aware of the many detrimental effects associated with high dropout rates. However, in practice, few researchers are cognizant that Web-experiment participants often drop out from a study at a rate that cannot be safely ignored, as evidenced by the fact that researchers rarely provide information regarding dropout rate when reporting Web experiments (mainly MTurk) in academic journals. As we elaborated in the Introduction, we believe such a lack of awareness is mainly rooted in the fact that participant attrition is less visible on the Internet than in physical labs. Apparently, a researcher is unlikely to worry about her experiment's internal validity being compromised by condition-dependent dropout if she sees no evidence of attrition having occurred. As a result, she could end up investing valuable time and financial resources to pursue something that is completely false as we have demonstrated in this research. Therefore, raising awareness among researchers is an urgent matter critical to the healthy development of the field.

At this juncture, we note that Web (vs. lab) experiments might actually be less vulnerable to a subtler form of attrition, namely, item attrition. Online studies are exclusively computer based; hence, the researcher can completely block the option to skip questions, especially those serving as the manipulation or DV measures. By contrast, to the extent that a lab study implements its manipulations in paper-and-pencil format, participants might opt to skip the critical item given that they cannot as easily walk away from the study entirely like their Web counterparts. However, researchers can easily spot any skipped items and be prompted to consider the possible negative impact item attrition might have on interpretation of the results. By contrast, on the Web, researchers who do not proactively monitor attrition might remain completely oblivious to the issue.

Because proving attrition that has already happened is condition independent is not feasible in practice, researchers conducting Web experiments should strive to minimize dropout at the point of manipulations—dropouts prior to manipulation would be condition independent by definition and would only affect external validity rather than internal validity. Although increasing payment and adding a long warm-up task before the manipulations have proven to cut down dropout rates (Horton et al., 2011; Reips, 2000), both strategies more or less neutralize one of the main appeals of online research, namely, saving cost. We showed in Study 4 that drastically reducing attrition while adding barely noticeable overhead is possible. Future research would be needed to explore how to further optimize and augment the cocktail strategy we tested in the present article.

Aside from implementing dropout-reduction strategies, before launching a study, a researcher should be mindful of how both the intended and unintended discrepancies between the manipulations in different conditions might differentially attract or repel people of certain characteristics or traits, thereby making the study susceptible to selective attrition rooted in individual differences. For example, in an experiment in which the two conditions differ in how demanding they are (e.g., Study 3), more agreeable and more conscientious people are likely to be overrepresented in the more demanding conditions. A researcher could place certain demographic and personality measures (e.g., agreeability and conscientiousness) prior to manipulation to discern if the attrition creates incomparable samples across conditions in terms of certain participant characteristics. Moreover, equipped with such knowledge, the researchers could better decide whether selective attrition can be reduced to a negligible level through realistic tinkering of the original design or it is time to move back to the physical lab.

Conclusions and Recommendations

Internet technology is transforming every aspect of human society, and scientific practice is no exception. Given that Web-experimentation methodology is here to stay for the foreseeable future, we recommend researchers who intend to collect experimental data online consider the following four main actions to better manage the threat to internal validity arising from the major pitfall of the online research, namely, a high attrition rate.

1. *Minimizing condition-dependent attrition*. Researchers should consider implementing a battery of remedies such as those tested in Study 4 to provoke potential quitters (less motivated or conscientious participants) to drop out of their experiments before the manipulations. Although promoting premanipulation attrition might limit the generalizability of experimental findings, it can help minimize the threat to internal validity arising from condition-dependent attrition.

2. *Gaining insights into causes of condition-dependent attrition*. Researchers should strive to design their experiments in such a way that would give them insight into why participants are quitting their experiments, especially when the decision to discontinue turns out to be condition dependent. The nature of online study precludes researchers from tracking down the dropouts and directly interviewing them. Thus, we recommend measuring certain demographics and personality variables prior to the manipulation. By comparing participants in different conditions on these premanipulations measures, researchers can gain insights into causes of attrition.

3. *Increasing the visibility of attrition*. Researchers should make sure their data-collection tool is posting respondents' data to the server at as many points as possible during the course of an experiment. This way, they will have partial responses from dropouts on record and be able to know at what point a dropout quit the study. Researchers using Qualtrics, which is posting responses at multipoints by default, should make sure to de-activate their surveys (see Footnote 2) before downloading datasets so that partial responses from dropouts are included in their working data files.

4. *Reporting attrition*. Regardless of the size of the attrition rate, when reporting a study, researchers should always disclose information regarding (a) overall attrition rate and (b) condition-based attrition rate. Whereas a zero attrition rate is rare on the Web, by consistently collecting and reporting this information, we can, as a field, get a better idea of when Internet studies become problematic.

To summarize, researchers should not only implement dropout-reduction strategies, but also proactively explore reasons for, record, and report participant attrition. Zero attrition is a rare outcome in Web studies, and therefore, attrition should be more openly reported and discussed.

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. http://dx.doi.org/10.1093/pan/mpr057

Birnbaum, M. H., & Mellers, B. A. (1989). Mediated models for the analysis of confounded variables and self-selected samples. *Journal of Educational Statistics, 14,* 146–158. http://dx.doi.org/10.3102/10769986014002146

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. http://dx.doi.org/10.1177/1745691610393980

Fernbach, P. M., Rogers, T., Fox, C. R., & Sloman, S. A. (2013). Political extremism is supported by an illusion of understanding. *Psychological Science, 24,* 939–946. http://dx.doi.org/10.1177/0956797612464058

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. http://dx.doi.org/10.1002/bdm.1753

Göritz, A. S. (2014). Determinants of the starting rate and the completion rate in online panel studies1. In R. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research* (pp. 154–170). Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9781118763520.ch7/summary10.1002/9781118763520.ch7

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods, 48,* 400–407.

Henderson, M. D. (2013). When seeing the forest reduces the need for trees: The role of construal level in attraction to choice. *Journal of Experimental Social Psychology, 49,* 676–683. http://dx.doi.org/10.1016/j.jesp.2013.03.001

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

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: Report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. *American Psychologist, 59,* 105–117. http://dx.doi.org/10.1037/0003-066X.59.2.105

May, F., & Monga, A. (2014). When Time Has a Will of Its Own, the Powerless Don't Have the Will to Wait: Anthropomorphism of Time Can Decrease Patience. *Journal of Consumer Research, 40,* 924–942. http://dx.doi.org/10.1086/673384

Musch, J., & Reips, U.-D. (2000). A brief history of Web experimenting. In *Psychological experiments on the Internet* (pp. 61–87). San Diego, CA: Academic Press. http://dx.doi.org/10.1016/B978-012099980-4/50004-6

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

Reips, U.-D. (2000). The Web experiment method: Advantages, disadvantages, and solutions. *Psychological Experiments on the Internet,* 89–117.

Ross, J., Irani, L., Silberman, M., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers?: Shifting demographics in Mechanical Turk. In *CHI'10 extended abstracts on Human factors in computing systems* (pp. 2863–2872). Retrieved from http://dl.acm.org/citation.cfm?id=1753873

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61,* 195–202. http://dx.doi.org/10.1037/0022-3514.61.2.195

Wisman, A., Heflick, N., & Goldenberg, J. L. (2015). The great escape: The role of self-esteem and self-related cognition in terror management. *Journal of Experimental Social Psychology, 60,* 121–132. http://dx.doi.org/10.1016/j.jesp.2015.05.006

Wolpert, R. M., & Gimpel, J. G. (1998). Self-interest, symbolic politics, and public attitudes toward gun control. *Political Behavior, 20,* 241–262. http://dx.doi.org/10.1023/A:1024814624070

Woltin, K.-A., & Yzerbyt, V. (2015). Regulatory focus in predictions about others. *Personality and Social Psychology Bulletin, 41,* 379–392. http://dx.doi.org/10.1177/0146167214566188

Yam, K. C., Chen, X.-P., & Reynolds, S. J. (2014). Ego depletion and its paradoxical effects on ethical decision making. *Organizational Behavior and Human Decision Processes, 124,* 204–214. http://dx.doi.org/10.1016/j.obhdp.2014.03.008

Received February 22, 2016
Revision received April 15, 2016
Accepted April 18, 2016 ■