

Effects of Participant Displeasure on the Social-Psychological Study of Power on Amazon's Mechanical Turk

SAGE Open
July-September 2019: 1–13
© The Author(s) 2019
DOI: 10.1177/2158244019876268
journals.sagepub.com/home/sgo


R. Gordon Rinderknecht¹ 

Abstract

Recall-based power priming is a popular research design that is widely disliked by Amazon Mechanical Turk (MTurk) workers. This article assesses the potential consequences of such displeasure through a conceptual replication of Fast et al. on MTurk. Specifically, this article assesses the extent to which recall-based priming can elicit a sense of high power and positive emotion. Findings indicate that being primed with a sense of high power through recall does not elicit the expected positive change in emotion. Findings also indicate that recall-based priming is a less effective manipulation of power than an alternative priming method with which participants were more willing to participate. Unlike the recall-based prime, this alternative prime also replicated Fast et al.'s original findings. These results are attributed to the incompatibility between feeling powerful and participating in a disliked study design. Findings highlight the importance of addressing worker displeasure in power research, and this article suggests how displeasure can be avoided as well as how such displeasure may be a detriment to other areas of research.

Keywords

power, emotion, MTurk, priming, replication

Introduction

Amazon's Mechanical Turk (MTurk) is a platform on which employers (referred to as "requesters") advertise tasks (referred to as "Human Intelligence Tasks [HITs]") to an online workforce for hire (Paolacci & Chandler, 2014). By using MTurk, researchers can save time and money while collecting data that meet or exceed the quality of data provided by offline samples (Behrend, Sharek, Meade, & Wiebe, 2011; Berinsky, Huber, & Lenz, 2012; Buhrmester, Kwang, & Gosling, 2011; Casler, Bickel, & Hackett, 2013; Hauser & Schwarz, 2016; Mason & Suri, 2012). MTurk participants also appear just as attentive and truthful as offline participants, and MTurk samples are far more diverse than are samples of university students (Paolacci & Chandler, 2014; Paolacci, Chandler, & Ipeirotis, 2010). Overall, researchers benefit from MTurk workers' participation, but the act of participation may at times be unpleasant for participants for a variety of reasons, including poor pay and feelings of exploitation (Busarovs, 2013; Pittman & Sheehan, 2016). This is generally approached as an ethical concern (Gleibs, 2017); I extend this research to evaluate the empirical consequences of experiencing displeasure as an MTurk worker. Research finds that difficult studies tend to elicit high dropout rates with online samples relative to laboratory samples

(Zhou & Fishbach, 2016), but how does the unpleasantness of such research affect those who choose to complete the study despite its difficulty? All difficult research may risk suffering from high dropouts, but the present article explores how research studying the effects of power may suffer additional risks. Specifically, I examine how dislike for a popular experimental design—recall-based power priming—may conflict with the manipulation's ability to elicit a sense of power and positive emotion.

Recall-Based Power Priming and Worker Displeasure

In social psychology, power is typically defined as the relative capacity to achieve one's goals by providing or withholding resources or punishments (Keltner, Gruenfeld, & Anderson, 2003), and power results from the resources and

¹University of Maryland, College Park, USA

Corresponding Author:

R. Gordon Rinderknecht, Department of Sociology, University of Maryland, College Park, 2112 Art-Sociology, College Park, MD 20742-1315, USA.
Email: rrinderk@umd.edu



punishments one can leverage in pursuit of their goals (Emerson, 1962; Fiske, 1993). Of particular importance here, power is also experienced psychologically, and the resulting psychological state manifests itself cognitively, emotionally, and behaviorally (Galinsky, Gruenfeld, & Magee, 2003). A popular method for putting someone in such a state involves asking participants to recall a time when they possessed power over others or where others possessed power over them (Galinsky et al., 2003; Sturm & Antonakis, 2015). Participants then describe, in writing, what happened and how they felt (Galinsky et al., 2003). This method is referred to as recall-based power priming.

While life as an MTurk worker can elicit complaints over broad issues, including underpayment and mistreatment, communities of workers also express displeasure toward particular research designs. Recall-based power priming is one research design complained about within these communities. As an example of such dislike, an MTurk worker asked an online community of workers:

Ever have those HITs where you . . .

“return” them as soon as you see what it is? For me, it’s those ones that start with “Write a paragraph describing an event where you had power over another individual. Please use as much detail as possible.” As soon as I read those words, my mouse hand is already reaching for the “return” button. (airhef, 2015)

Several workers replied expressing similar tendencies to “return” (i.e., drop out) recall-based power-priming HITs, and other conversations in this community feature similar complaints (e.g., bicZoid, 2016; claccnt01, 2015). Consistent with this sentiment, Zhou and Fishbach (2016) found high drop-out rates in a recall-based power-priming study conducted on MTurk. However, displeasure may have further negative effects that are particular to research on power. Workers’ displeasure for particular methods may interact with the theoretically predicted emotional consequences of power. Research finds that high and low power elicit positive and negative emotions, respectively (Anderson & Berdahl, 2002; Berdahl & Martorana, 2006; Keltner et al., 2003), but displeasure experienced as a consequence of the prime may attenuate the positive emotion expected from being primed with a high sense of power. Furthermore, as doing something one dislikes is inconsistent with feeling powerful, displeasure directed at the prime may undercut recall-based priming’s usefulness for eliciting a sense of power. The present article assesses the usefulness of MTurk samples for recall-based power priming via a conceptual replication of Fast, Sivanathan, Mayer, and Galinsky (2012). First, however, it is important to conduct two pretests to identify (a) whether recall-based power priming is in fact disliked by MTurk workers, (b) why it is disliked, and (c) how commonly

recall-based power priming is encountered. The first will tell us whether anecdotal examples of displeasure (airhef, 2015; bicZoid, 2016; claccnt01, 2015) are indicative of widespread displeasure, the second will tell us the criterion for selecting a power prime that may avoid such displeasure, and the third will indicate the importance of assessing the effectiveness of recall-based power priming with MTurk samples.

Pretest 1

In Pretest 1, I sought to (a) quantify the popularity of recall-based power priming and (b) determine why recall-based power priming elicits displeasure.

Materials and Methods

Participants. I ran Pretest 1 as part of a larger, unrelated study administered to 220 American MTurk participants (see Survey 2 in Supplementary Information for an overview of this previous study). The study paid US\$2 for an advertised 15-min participation ($M_{\text{age}} = 33.16$, $SD_{\text{age}} = 10.28$, 39% women). All participants were U.S. residents who had previously completed at least 1,000 HITs and had a HIT acceptance rate of at least 99%. All participants provided informed consent by selecting a button indicating that they accepted the terms of the consent form.

Design and procedure. Pretest 1 relied on survey measures to evaluate workers’ level of exposure to recall-based power priming and relied upon both survey and open-ended questions to assess workers’ dislike for the prime.

Exposure to recall-based power primes. Participants reported how often they encounter recall-based power priming (1 = *daily* to 7 = *never*). Participants who selected “never” were directed to the debriefing page.

Evaluations of recall-based power primes. Participants reported whether they usually enjoyed participating in recall-based power-priming studies (“usually yes,” “usually no,” “not sure,” or “no opinion”). In addition, I asked participants who dislike recall-based power priming to explain their dislike as an open-ended response.

Response coding. Seven explanations for disliking recall-based power priming emerged from participants’ responses, which I refer to as dislikes recall, dislikes writing, asks too much, boring, depressing (i.e., participants felt generally unhappy after exposure to the prime), ineffective (i.e., participants felt the method does not achieve its intended goals), and dislikes power. After identifying these explanatory categories, I again read through each response and coded each as expressing one or more of the seven previously mentioned explanations.

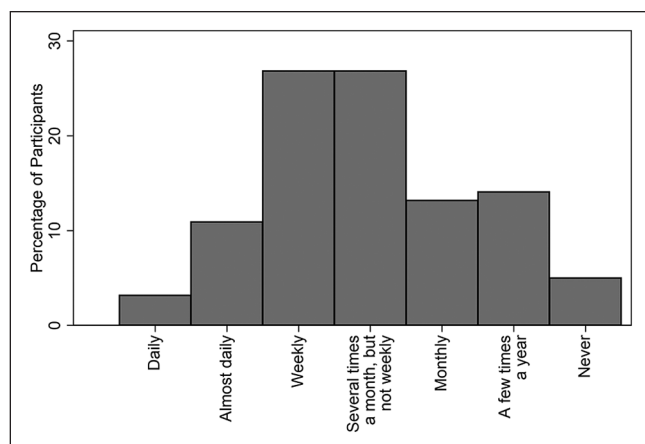


Figure 1. Percentage of participants in each category of recall-based power-priming exposure.

Results

Exposure to power primes. Figure 1 shows participants' self-reported exposure to recall-based power priming; 41% of participants reported encountering such primes on at least a weekly basis, and 81% of participants reported encounters on at least a monthly basis. Eleven participants (5%) reported having never encountered such priming. Percentages reported in the following analyses exclude these 11 participants.

Evaluations of power primes. When asked whether they enjoyed participating in recall-based power-priming research, 54% of participants reported "usually no," 19% reported "usually yes," 7% reported "not sure," and 20% reported having "no opinion." Of the 113 participants who reported disliking such priming, 107 produced usable responses, and these 107 responses contained 143 explanations categorized into the seven explanatory categories reported below from most to least commonly encountered. These categories include (a) disliking recall, (b) disliking writing, (c) asking too much from participants, (d) finding the research boring, (e) depressing, (f) ineffective, and (g) generally disliking power.

Participants' most common explanation for not enjoying recall-based power priming broadly involved disliking recalling events involving power (29% of all explanations). Some of these participants found such recall tedious or intrusive, but the most common reason, given by 52% of these participants, involved participants' difficulty coming up with a relevant situation. Some claimed to not have the necessary life experiences, while others expressed a desire to not repeat stories used in previous studies. As one participant put it, "I can only write so many without repeating the same one again. My life is just not full of power struggles."

The next most common explanation involved disliking writing (24% of all explanations). However, 31% of these explanations provided details suggesting the participant had

a particular dislike for writing in recall-based power-priming or similar studies. One such participant wrote, "I used to make my living as a writer, and even I don't like to write about such experiences." The next most common explanation involved researchers asking too much of workers (20% of all explanations), and 50% of these explanations explicitly mentioned poor pay.

Beyond the top three explanations for not enjoying recall-based power priming, 19 participants stated that such priming research is boring, tedious, and/or repetitive (13% of explanations). One such participant complained that "... I'm constantly re-telling the same story ..." and that "... it just becomes dumb and repetitive over time." Eight participants expressed that recall-based power priming makes them depressed or uncomfortable (approximately 1% of explanations). Six participants stated in some way that the manipulation seemed flawed (<1% of explanations). One such participant responded, "... after seeing [recall-based power primes] so often, I don't know that they have the effect our requesters hope for them to have." And finally, six participants stated a general dislike of power—either using it, having it used against themselves, or both (<1% of explanations).

Pretest 2

Pretest 1 indicates that recall-based power priming is commonly encountered and disliked, but these data do not indicate whether this method is more commonly encountered or disliked than other popular research designs. To address this, Pretest 2 measured how relatively common and disliked recall-based power priming is by dividing participants into three groups where participants either (a) ranked how commonly they encounter several research paradigms, (b) ranked how much they liked participating in these research paradigms, or (c) reported their two most disliked questions and/or requests often encountered on MTurk. I randomly assigned participants to separate groups to prevent exposure to one question affecting responses to other questions.

Materials and Methods

Participants. I attached a questionnaire to a brief, unrelated survey administered to 750 MTurk participants spread evenly across three groups (see Survey 3 in Supplementary Information for an overview of this previous study). The study paid 25 cents for an advertised 1- to 2-min participation ($M_{age} = 36.41$, $SD_{age} = 11.79$, 47% women). All participants were U.S. residents who had previously completed at least 500 HITs and had a HIT acceptance rate greater than 95%. All participants provided informed consent by selecting a button indicating that they accepted the terms of the consent form.

Design and procedure. Pretest 2 relied on a ranking to assess recall-based power priming's popularity relative to other

popular research designs, and both ranking and qualitative coding were used to measure workers' tendency to dislike recall-based power priming relative to other common research paradigms. Group 1 had participants rank the relative popularity of recall-based power priming. Group 2 directly measured dislike through a ranking of participants' most liked paradigms. Note that for Groups 1 and 2, the paradigms used in the ranking were arbitrarily chosen by previous research (J. Chandler, Mueller, & Paolacci, 2014) and therefore may neglect to compare recall-based power priming against other commonly encountered research paradigms. To determine whether recall-based power priming is more or less disliked than paradigms or procedures not captured in Group 2, Group 3 asked participants to report via free responses their least liked requests often encountered in studies on MTurk.

Group 1: Popularity of research paradigms. Participants ranked descriptions of four randomly ordered research paradigms based on "how commonly you encounter them," with the most commonly encountered study being placed first. Along with recall-based power priming, participants ranked three paradigms identified by J. Chandler et al. (2014) as commonly encountered on MTurk: prisoner's dilemmas, dictator/ultimatum games, and moral dilemmas.

Group 2: Enjoyment of research paradigms. Participants ranked the same previously mentioned four research paradigms: "... in order of how much you like participating in these kinds of studies." Of the 250 participants who completed this ranking, the first nine responded to a question asking them to rank paradigms based on "dislike" rather than "like," with the most disliked paradigm placed first. Of these nine participants, three ranked recall-based power priming first (most disliked) and six ranked such priming last (most liked). This tendency toward extremes suggested some participants were incorrectly assuming last meant least liked. I altered the question wording for the remaining 241 participants because ranking based on liking seemed less likely to be misinterpreted by participants. All analyses of this group are based on the remaining 241 participants.

Group 3: Most disliked questions and/or requests. Participants reported their "... two least favorite, commonly encountered questions and/or requests in survey HITs ..." as free responses. In addition, I specified that "I do not mean what you dislike about these HITs generally (e.g., underpayment, wasting time, 'bubble hell') but rather "... what sort of specific questions do you dislike answering or requests do you dislike following?" Participants in this group never encountered mentions of the four research paradigms presented to Groups 1 and 2.

Results

Group 1: Popularity of research paradigms. Participants identified dictator/ultimatum games as their most commonly

encountered paradigm (M rank = 1.84), followed by recall-based power priming (M rank = 2.28), prisoner's dilemmas (M rank = 2.78), and moral dilemmas (M rank = 3.11).

Group 2: Enjoyment of research paradigms. Participants liked dictator/ultimatum games the most (M rank = 1.61), followed by prisoner's dilemmas (M rank = 2.12), moral dilemmas (M rank = 2.63), and, finally, recall-based power priming (M rank = 3.64).

Group 3: Most disliked questions and/or requests. In total, 250 participants each reported their two least favorite questions and/or requests, resulting in 500 total responses; 88 of these responses mentioned recall-based manipulations—81 expressed a general dislike for such primes, and 20 of these responses specifically mentioned power. Other responses referenced recalling something negative (e.g., "a particularly difficult time," "something you don't like thinking about," "a time that I felt inferior," "a time when you were ashamed or felt guilty"), innocuous (e.g., "a specific event," "a past situation"), or did not provide an example. The remaining seven responses focused on factors associated with recall-based methods (e.g., timers and word-count requirements). For example, one such participant wrote about disliking: "Short writing prompts about a vague emotion in a rigid time frame, i.e., tell me about a time in the last 6 weeks where you felt a sense of dread." It was unclear from this response if recalling and writing about vague emotions would be acceptable without the "rigid time frame." Overall, 121 participants (48%) mentioned in some way disliking research involving writing, and 70% of these participants expressed a particular dislike for recall-based priming research. (Nine participants complained about writing in both of their responses.) By comparison, four participants (2%) disliked moral dilemmas, one participant disliked dictator/ultimatum games, and no participants disliked prisoner's dilemmas. In addition, workers' dislike for recall-based power priming appears to be aimed at recall-based methodology and not power specifically. Although 23% of responses in Group 3 that mentioned recall-based priming also mentioned power, no responses mentioned power without mentioning recall-based manipulations.

Pretest Discussion

Together, these pretests indicate that recall-based power priming is commonly disliked. In Pretest 1, a majority of participants familiar with recall-based power priming reported not enjoying participation in such research. Pretest 2 also found recall-based priming methodology to be less liked than other common research paradigms.

Pretest 1 revealed the elements of the recall-based power prime that appear to elicit displeasure, with the two main complaints revolving around disliking recalling events involving power and disliking writing. Avoiding displeasure

may therefore require the use of a method that can prime power while avoiding these sources of displeasure.

Finally, these pretests highlight the popularity of recall-based power-priming research on MTurk, both in how routinely it is encountered (Pretest 1) and relative to other common research paradigms (Pretest 2). However, because the comparison categories in Pretest 2/Group 1 were arbitrarily chosen by previous research (J. Chandler et al., 2014), it is difficult to know whether recall-based power priming is among the most commonly encountered paradigms on MTurk. Still, findings from Pretest 1 and 2 indicate that a significant amount of recall-based power-priming research is being conducted on MTurk. It is therefore important to determine whether a power prime lacking the disliked qualities of the recall-based prime functions as a more effective manipulation of power.

The Main Study

This study evaluates MTurk's usefulness for recall-based power-priming research through a conceptual replication of a recall-based power-priming study (Fast et al., 2012), which was undertaken to demonstrate that the experience of power leads to overconfident decision-making, and showed with data collected from MTurk that high power increases confidence in one's decisions relative to low-power and control conditions. I chose to replicate this study for multiple reasons. First, I pursued this replication primarily to compare two different power primes and not as a challenge to the original research. Fast et al. (2012) was therefore an appealing study to replicate due to being well supported—It has been replicated both in the original study and in more recent research (Lammers, Dubois, Rucker, & Galinsky, 2017). Second, Fast et al. was appealing to replicate due to the ease of conducting it with an MTurk sample and with additional, alternative power-priming conditions, as well as due to the availability of documentation on the language used in the original study.

My central goal with this research was to see whether MTurk workers responded as prior research would expect to the recall-based prime, and whether an alternative prime functioned as a better manipulation of power. These expectations pertained to (a) confidence, (b) sense of power, and (c) emotion. Specifically, if the prime worked as it had in prior research, participants assigned to the high-power condition should have felt more confident and powerful than participants assigned to the low-power and control conditions. Theory also indicates that participants in the high-power condition should also have experienced more positive emotion relative to how they felt before exposure to the prime (Keltner et al., 2003). In contrast, participants assigned to the low-power condition should have felt less confident and powerful than participants assigned to the high-power and control conditions. Participants should also report more negative emotions after exposure to the low-power prime relative to how

they felt before the prime. However, I anticipated that workers' dislike for the recall-based prime may conflict with the high-power prime because feeling powerful is inconsistent with doing something one dislikes.

For comparison, I included two additional conditions utilizing an alternative power prime. This alternative method is referred to as priming via imagined hierarchical role (Dubois, Rucker, & Galinsky, 2012; Galinsky, Rucker, & Magee, 2015)—hereafter referred to as *hierarchical-role* priming. Without considering the possible consequences of displeasure, the high- and low-power hierarchical-role conditions should elicit similar confidence, sense of power, and emotion responses as the high- and low-power recall conditions, respectively. However, because hierarchical-role priming avoids the aspects of the recall-based method disliked by MTurk workers (e.g., writing and recalling events), I anticipated that the high-power hierarchical-role condition would be a more effective manipulation of power than the high-power recall prime.

I treat recall-based and hierarchical-role priming as equivalent manipulations of power, with the only important difference being the extent of displeasure they elicit from participants. Specifically, I treat them as equivalent manipulations of *social* power. Social power refers to the ability to affect others, as opposed to *personal* power, which refers to autonomy from others (Galinsky et al., 2015). Treating these manipulations as equivalent manipulations of social power appears warranted, given that priming a sense of high power in both recall-based and hierarchical-role priming focuses on control over others rather than independence from others. Treating these manipulations as equivalent is further supported by prior research using the two manipulations interchangeably (Dubois et al., 2012) and by (Galinsky et al., 2015) reviewing of both primes without noting differences between them in what they manipulate.

It is important to note that this is a conceptual replication and not a direct replication of Fast et al. (2012). In Fast et al.'s original study, participants were randomly assigned to a high-power, low-power, or control condition. After responding to the prime, participants next completed the power and confidence measures, followed by the Positive and Negative Affect Schedule (PANAS) and demographic questions, and finally they were debriefed. Apart from the inclusion of the two hierarchical-role primes, there are three central deviations from Fast et al.'s original design that are important to note. First, Fast et al. utilized the PANAS as a posttest measure indicating participants' emotional response to the prime (Watson, Clark, & Tellegen, 1988). The present study instead uses pre- and posttest measures of emotion derived from Lawler, Shane, and Yoon (2008). I utilize a pre- and posttest design to increase statistical power. I utilize the alternative measure because Fast et al.'s original study reported no significant differences in the experience of negative emotion and only a marginal difference between conditions in the experience of positive emotion as collected by

the PANAS. PANAS itself captures a wide range of emotions, both positive and negative, and some affective states, like “alert” and “active,” which do not appear relevant to the present article. I hoped to increase the likelihood of detecting emotional change as a result of exposure to the prime by utilizing a measure that included a greater proportion of emotions relevant to the experience of displeasure. I discuss limitations to this approach in the “Discussion” section following the replication. Second, the present study utilizes a lower number of participants per condition—40, as opposed to the approximately 50 participants per condition reported by Fast et al. Furthermore, this study employed exclusion criteria that differed from Fast et al., including manipulation check wording and the exclusion of participants who likely copy/pasted responses. Third, participants were instructed to type at least 100 words for their response. This differed from Fast et al., which appears to have not had a word requirement.

Due to differences in emotion measurement, sample size, and exclusion criteria, I focus predominantly on comparisons between the recall-based and hierarchical-role priming conditions in the present study and less on comparisons between the findings of this study and Fast et al.’s (2012) original study.

Materials and Methods

Participants. In total, 200 MTurk workers completed the study after responding to an advertisement asking them to “Please participate by answering some questions about yourself.” The advertisement then reported payment details and estimated study duration; 43 participants dropped out before completion. However, six participants dropped out after condition assignment (and are therefore counted in the 43 drop-outs), restarted the study, and then proceeded to complete their participation (and are therefore counted in the 200 participants who completed the study). I discuss these six participants further in the “Results” section of the Main Study, and I ultimately exclude their data (both complete and incomplete) from all analyses. Participants who finished the study received US\$1.25 for an advertised 10-min participation ($M_{\text{age}} = 33.10$, $SD_{\text{age}} = 9.19$, 42% women). All participants were residents of the United States who had previously completed at least 500 HITs and who had a HIT acceptance rate greater than 95%. All participants provided informed consent by selecting a button indicating that they accepted the terms of the consent form.

Design and procedure. This replication consisted of five conditions. For the first three, the manipulation wording matched the wording used by Fast et al.’s (2012) third study. For the (1) high-power recall condition, participants were instructed to

Please recall a particular incident in which you had power over another individual or individuals. By power, we mean a situation

in which you controlled the ability of another person or persons to get something they wanted, or were in a position to evaluate those individuals. Please describe in detail this situation in which you had power: events, feelings, thoughts, etc.

For the (2) low-power recall condition, participants were instructed to

Please recall a particular incident in which someone had power over you. By power, we mean a situation in which someone controlled your ability to get something you wanted, or was in a position to evaluate you. Please describe in detail this situation in which you lacked power: events, feelings, thoughts, etc.

For the (3) control condition, participants were instructed to “Please recall your day yesterday. Please describe in detail your day: events, feelings, thoughts, etc.” Two additional conditions instructed participants to imagine being in either (4) a boss in a company (high power) or (5) an employee in a company (low power). For the (4) high-power hierarchical-role condition, participants read the following:

We would like to imagine you are a boss at a company. Read about the role below and try to vividly imagine what it would be like to be in this role (i.e., how you would feel, think, and act). As a boss, you are in charge of directing your subordinates in creating different products and managing work teams. You decide how to structure the process of creating products and the standards by which the work done by your employees is to be evaluated. As the boss, you have complete control over the instructions you give your employees. In addition, you also evaluate the employees at the end of each month in a private questionnaire—that is, the employees never see your evaluation. The employees have no opportunity to evaluate you.

For the (5) low-power hierarchical-role condition, participants read the following:

We would like to imagine you are an employee at a company. Read about the role below and try to vividly imagine what it would be like to be in this role (i.e., how you would feel, think, and act). As an employee, you are responsible for carrying out the orders of the boss in creating different products. The boss decides how to structure the process of creating these products and the standards by which your work is to be evaluated. As the employee, you must follow the instruction of the boss. In addition, you are evaluated by the boss each month, and this evaluation will be private, that is, you will not see your boss’s evaluation of you. This evaluation will help determine the bonus reward you get. You have no opportunity to evaluate your boss.

The manipulation in Conditions 4 and 5 matched the wording used by Dubois et al. (2012).

Cheating by copy/pasting. Participants may cheat when responding to the recall-based prime by copy/pasting pre-written responses, rather than recalling and writing their

responses during the study. I identify copy/pasting through JavaScript code designed to detect when characters are typed or copy/pasted into the study's textbox. However, copy/pasting behavior alone may not indicate the participant cheated. Some participants might prefer to write their response in a text processor rather than in the textbox provided to them by the study. To account for this possibility, I calculate the minimum words per minute (wpm) the copy/pasted response must have been typed, had it been typed in a word processor. I categorize a participant as having cheated if they copy/pasted a significant amount of their response and—had it been typed during the study—must have been typed faster than 120 wpm. I chose 120 wpm as a cutoff because it is a speed only achieved by highly advanced typists (Ayres & Martínás, 2005).

Cheating by restarting the study. The design of this study incentivized a form of cheating I had not expected. The consent form informed participants that they “. . . might either be asked to share some of your own life experiences or read about hypothetical experiences.” Some participants restarted the survey after being assigned to a recall-based condition, apparently in hope of gaining access to a hierarchical-role condition. I identified this behavior by embedding JavaScript into the MTurk advertisement. This script appended workers' unique worker ID into the study URL, allowing me to identify multiple submissions from the same participant and the timing of each submission—that is, I was then able to identify at which point they dropped out during their first run throughout the study, as well as tie them to subsequent submissions. I contacted the first participant who engaged in such behavior and she explained, “I cleared the browser history and refreshed because I didn't want to write about my day.”

Sense of power and confidence. Immediately following the manipulation, participants answered the same eight power questions ($\alpha = .94$; $M = 4.71$, $SD = 1.31$) and four confidence questions ($\alpha = .88$; $M = 5.26$, $SD = 1.12$) used by Fast et al. (2012). All measures ranged from 1 = *strongly disagree* to 7 = *strongly agree*. For analysis, responses were recoded so that 7 indicated the greatest sense of power or confidence.

Emotional response. Participants responded to the same battery of 12 emotion measures immediately prior to condition assignment ($\alpha = .92$; $M = 6.06$, $SD = 1.25$) and immediately following the power and confidence questions ($\alpha = .93$; $M = 6.01$, $SD = 1.40$). Each of these emotion measures consisted of an 8-point bipolar scale, with poles including displeased/pleased, unhappy/happy, and so on (Lawler et al., 2008; see Main Study in Supplementary Information for all scale items). For both measures, larger values indicate a more positive change in emotion.

Manipulation check. To check to make sure participants paid attention to the prime, participants in recall-based

conditions reported whether they had been asked to write about (a) “A time in which you had power over others,” (b) “A time in which others had power over you,” or (c) “Details about your previous day.” Participants in hierarchical-role conditions reported whether they had been asked to imagine being (a) “A boss in a company,” (b) “An employee in a company,” or (c) “Someone self-employed.”

Reasons for dropping out. Participants who dropped out were invited to participate in a brief follow-up survey, which asked them to report basic demographic information as well as to explain why they dropped out. Several participants dropped out and then restarted the study. I do not count these participants as having dropped out and, apart from the first participant to engage in such behavior, I did not follow up with them.

Reasons for considering dropping out. Participants reported whether they ever considered dropping out of the study (either “yes” or “no”). Participants who selected “yes” then explained why they considered dropping out and why they decided to stay in the study.

Analytic strategy. I begin by using ordinary least squares (OLS) regression analyses to compare differences in confidence and power between conditions. Next, I rely on Wilcoxon's signed rank sum tests to evaluate the change in pre- and posttest measures of emotion. Nonparametric tests are necessary here due to the presence of outliers in the emotional response measure, though findings in this section do not differ from findings produced by comparable parametric tests.

Results

Descriptive data

Dropouts. In this section, I report both incidence of dropping out and considered dropping out of the study. In both cases, I focus on how such rates differed between recall-based and hierarchical-role conditions, as well as participants' reported reasons for dropping out or considering dropping out. I do not focus on the five participants who dropped out of the study prior to condition assignment, nor do I include the six participants who dropped out of the study after condition assignment and then proceeded to restart the study. In what follows, I first discuss these six participants further, explain why they are not included, and report the sample size following their exclusion. Next, I report incidence of dropping out and considering dropping out.

Six participants dropped out after being assigned to a recall-based condition, then proceeded to restart the study and completed their participation in a different condition. Of these six participants, five proceeded to complete the study after being assigned to a hierarchical-role condition and one completed the study after again being assigned to a recall-based

Table 1. Count of Participants Excluded by Reason and Condition.

Condition	Original <i>N</i>	Cheating		Poor quality		Final <i>N</i>
		Copy/pasted	Repeated participation	Failed manipulation check	Unacceptable response	
Recall						
High	40	−3	0 (1)	−1	−1 (1)	35
Low	40	−1	0	−1	−1	37
Control	40	−2	0	0 (1)	0 (1)	38
Hierarchical role						
High	40	—	−3	−3 (2)	—	34
Low	40	—	−2	−2 (1)	—	36

Note. Numbers in parentheses refer to the count of participants in each category already excluded based on prior criteria; negative numbers and zeros refer to the number of participants removed from the dataset who had not already been excluded.

condition. I remove all six participants' data from all analyses, including six instances of dropping out and six instances of completing the study. I remove their data from analyses of dropouts because (a) their inclusion complicates the analyses without substantively altering conclusions, and (b) I did not attempt to follow-up with them to identify why they chose to drop out. I remove their data from analyses of completed data due to their behavior violating random assignment. Overall, their removal leaves 149 participants having been assigned to a recall-based condition and 77 having been assigned to a hierarchical-role condition.

Of the 149 participants assigned to a recall-based condition, 30 (25%) dropped out. Of these 30 participants, 16 responded to the follow-up survey. When asked why they dropped out, six focused on insufficient payment, nine expressed a general dislike for writing, three of which expressed a lack of events to draw upon for the recall prime, and one reported having ran out of time to complete the study. In comparison, of the 77 participants assigned to a hierarchical-role condition, only two (3%) dropped out. The one such participant who responded to the follow-up survey explained he dropped out due to fearing he was going to fail the manipulation check. Overall, participants assigned to a recall-based condition were significantly more likely to drop out than were participants assigned to a hierarchical-role condition, $\chi^2(1, N = 226) = 12.8, p < .001$. Dropouts did not significantly differ between recall-based conditions, $\chi^2(1, N = 149) = 2.1, p = .353$.

Of the 119 participants assigned to a recall-based condition who completed the study, 21 (18%) reported considering dropping out at some point during the study. When explaining why they considered dropping out, three focused on insufficient payment while the majority focused on a general dislike for writing. In explaining why they did not drop out, the majority focused or alluded to sufficient payment. In comparison, of the 75 participants assigned to a hierarchical-role condition who completed the study, only one (approximately 1%) participant considered dropping out, which he explained as being due to "... feeling a little grouchy ..."

and insufficient payment. Overall, among participants who completed the study, participants assigned to a recall-based condition were significantly more likely to consider dropping out than were participants assigned to a hierarchical-role condition, $\chi^2(1, N = 194) = 12.2, p < .001$. Considering dropping out did not significantly differ between recall-based conditions, $\chi^2(1, N = 119) = 0.4, p = .81$.

Data quality. Table 1 reports the number of participants in each condition removed for cheating and poor quality; 16 participants copy/pasted a significant amount of text into their responses. Had these responses been typed in a word processor, four of them would have exceeded the cutoff speed of 120 wpm by typing between 164 and 364 wpm, indicating that it is extremely unlikely that these participants typed out their responses during the study and instead relied on pre-written responses. Three additional participants copy/pasted responses that they possibly typed in or near a range achieved by the average professional typist (50-80 wpm; Ayres & Martínás, 2005). These three participants were kept in the data. Their inclusion or exclusion does not affect the conclusions of my analyses. Finally, I labeled two participants as having cheated because their copy/pasted responses were nonsensical. Regarding dropouts based on other criteria, I exclude all participants who restarted the study after encountering the manipulation and/or failed the manipulation check. And although I keep responses in the data even if they fall a bit below the required 100-word requirement, I exclude one response for being only 33 words long. I also exclude a participant instructed to recall a period of high power but who mostly complained about the study's word-length requirement, stating, "i [sic] think you have all the power here." All other participants responded to the prime as expected. Unless noted otherwise, all further analyses are conducted on the remaining 180 participants.

Sensitivity and power analysis. Given the total sample size of 180 spread across five groups, the minimum detectable effect is .26 with a power of .80 and .30 with a power of

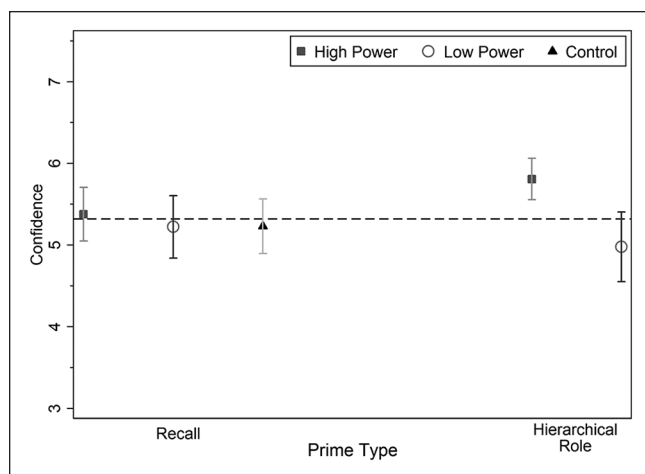


Figure 2. Mean confidence by prime type and level.
Note. Dashed line indicates mean confidence across all conditions.

.90. The difference in confidence between the high/low and high/control contrasts in Fast et al.'s (2012) original study exceeded these minimum thresholds. (Fast et al., 2012, reported these contrasts as differing by .41 and .34, respectively.) This indicates that the current study is sufficiently powered to detect the effects found in Fast et al.'s original study. Furthermore, given the effect size observed for perceptions of confidence between the high- and low-power recall conditions in the replication below, a power analysis indicates that achieving a significant difference between these conditions at the .05 level would require an additional 1,472 participants across both conditions. I take these findings as evidence that the replication below is sufficiently powered, despite the difference in confidence observed between the recall-based conditions falling below the .26 threshold.

When assessing differences in sense of power, the difference between the low-power recall and low-power hierarchical-role conditions failed to reach the .26 minimum. I note this lack of power in both the reporting and discussion of this finding.

Replication

There was no significant association between sex and confidence, sense of power, or both emotions before and after the prime. Age associated significantly with confidence, $r = .15$, $p = .045$, but did not associate significantly with sense of power or both emotions before or after the prime. Including age as a covariate in the replication analyses below did not alter results or affect levels of significance. Therefore, I do not discuss this variable further.

Fast et al. (2012) hypothesized and found that participants primed with a sense of power reported greater confidence in their decisions. I was unable to replicate Fast et al. using their original recall-based prime and confidence measures. Figure 2 shows differences in mean confidence between all five

conditions. Participants' confidence in the high-power recall condition ($M = 5.38$, $SD = 0.99$) did not differ significantly from participants' confidence in the low-power recall condition ($M = 5.22$, $SD = 1.19$), $t(107) = -0.61$, $p = .543$, $d = -.12$. The high-power recall participants' confidence also did not differ significantly from participants in the control condition ($M = 5.23$, $SD = 1.06$), $t(107) = -0.59$, $p = .56$, $d = -.11$. The combination of the low-power recall and control conditions ($M = 5.23$, $SD = 1.11$) also did not differ significantly from the high-power recall condition, $t(107) = -0.69$, $p = .492$, $d = -.13$. Fast et al.'s hypothesis replicated among the hierarchical-role conditions. The high-power hierarchical-role participants ($M = 5.81$, $SD = 0.75$) reported significantly more confidence in their decisions than did the low-power hierarchical-role participants ($M = 4.98$, $SD = 1.30$), $t(68) = -3.23$, $p = .002$, $d = -.78$.

Figure 3 shows differences in mean perceived power between all five conditions. Regarding participants' sense of power, high- and low-power conditions differed significantly and in the expected direction across both recall-based and hierarchical-role conditions. However, the high-power hierarchical-role prime was a more effective manipulation of power than was the high-power recall prime. The high-power hierarchical-role condition ($M = 5.92$, $SD = 0.89$) elicited a significantly greater sense of power than did the high-power recall condition ($M = 4.94$, $SD = 1.10$), $t(175) = 3.45$, $p = .001$, $d = .52$. Participants' sense of power in the low-power hierarchical-role condition ($M = 4.02$, $SD = 1.53$) did not significantly differ from those in the low-power recall condition ($M = 4.19$, $SD = 1.19$), $t(175) = -0.58$, $p = .563$, $d = -.09$. Note, however, that the present study lacked the sufficient sample size to detect a difference between these two low-power conditions, if such a difference existed. Furthermore, the high-power recall condition ($M = 4.94$, $SD = 1.10$) differed significantly from the low-power recall condition, $t(175) = -2.67$, $p = .008$, $d = -.40$. The control condition ($M = 4.61$, $SD = 1.14$) did not differ significantly from either the high- or low-power recall conditions, but did differ significantly from the low-power hierarchical-role condition, $t(175) = -2.11$, $p = .036$, $d = -.32$. Finally, the combination of the low-power recall and control conditions ($M = 4.40$, $SD = 1.17$) differed significantly from the high-power recall condition, $t(107) = -2.27$, $p = .025$, $d = -.44$.

Overall, the hierarchical-role conditions produced the effects observed by Fast et al. (2012), while the recall-based conditions only partially behaved as they had in Fast et al.'s original paper. Neither the high- nor low-power recall conditions differed significantly from the control condition for either sense of power or confidence. I anticipated that the high-power condition might not differ from the control condition due to disliking the method being incompatible with feeling powerful. This incompatibility does not explain why the low-power recall condition also did not differ significantly from the control condition. It is worth noting, however, that the low-power condition did not differ significantly

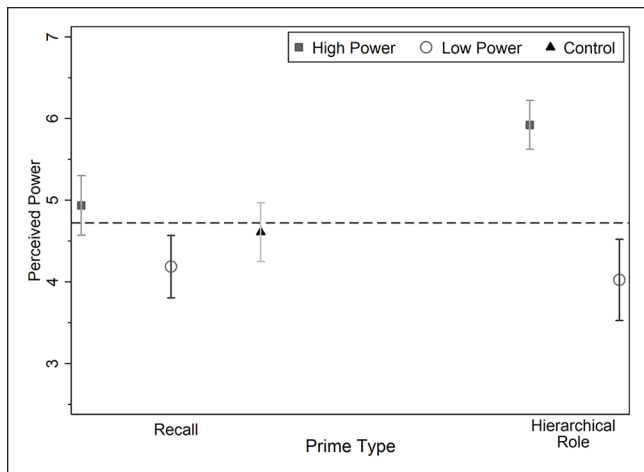


Figure 3. Mean perceived power by prime type and level.
Note. Dashed line indicates mean perceived power across all conditions.

from the control condition for sense of power or confidence in Fast et al.'s original study.

Emotional response to the prime. Research indicates that being primed with a sense of high power should evoke positive emotional reactions, while being primed with a sense of low power should evoke negative emotional reactions (Keltner et al., 2003). This expectation was confirmed for all conditions apart from the high-power recall condition. Participants in the high-power hierarchical-role condition felt more positive after the prime than they felt prior to the prime ($Z = 3.46, p < .001$), while participants in the high-power recall condition reported no such change ($Z = 0.97, p = .333$). Participants in both low-power recall ($Z = -2.73, p = .006$) and hierarchical-role ($Z = -3.81, p < .001$) conditions felt worse after exposure to the prime relative to how they felt before the prime. Participants in the control condition reported no significant change ($Z = 0.61, p = .54$). Overall, participants responded emotionally as theory expected in every condition except the high-power recall condition (Keltner et al., 2003).

Discussion

Fast et al.'s (2012) primary hypothesis—that power increases confidence—replicated convincingly in hierarchical-role conditions. As expected, participants assigned to read a scenario about being a boss reported significantly more confidence in their decisions than did those assigned to read a scenario about being an employee. This same hypothesis failed to replicate using Fast et al.'s original recall-based prime. This failure is unusual given that Fast et al. not only successfully used recall-based priming in their original study but also replicated these findings with a different MTurk sample. What could have changed between Fast et al.'s original studies and the present replication? One issue may lie in

the present study being a conceptual rather than direct replication, which makes it difficult to make comparisons with Fast et al.'s original work. However, assuming the failed replication is not the product of design choices, it is possible that displeasure toward recall-based priming has grown since Fast et al. originally conducted their research in 2009.

The main focus of the replication was on comparison between the recall-based and hierarchical-role primes, which revealed the hierarchical-role prime to have several advantages. Both the recall-based and hierarchical-role conditions produced a significant difference in the sense of power between high- and low-power conditions. However, the high-power hierarchical-role prime elicited a significantly greater sense of power than did the high-power recall prime, while no significant difference was found between the low-power conditions. That being said, the low-power conditions were insufficiently powered to detect a difference in participants' sense of power, if a difference exists. Furthermore, all conditions elicited the changes in emotion theory expected (Keltner et al., 2003) except for the high-power recall-based prime, which produced no significant change in emotion. These findings support the proposition that displeasure directed at recall-based priming may be conflicting with the high-power recall prime.

Changes in emotion do not always result from manipulations of power (Galinsky et al., 2015). Most relevant to the present article, recent research finds that priming a sense of high power increases positive affect in positive contexts (e.g., a summer day) and neutral contexts, but not in negative contexts (e.g., an exam day) (Leach & Weick, 2018). Findings from both pretests indicate that recall-based power priming may qualify as a negative context for a significant portion of MTurk participants. And although comparisons are complicated by differences in emotion operationalization, findings from the main study generally support the conclusions of Leach and Weick (2018) that context affects the emotional consequences of power. The present study further indicates that context may also shape other consequences of power (e.g., confidence). Future recall-based power-priming research using MTurk samples may therefore benefit from identifying participants for whom recall-based methods constitute a negative context, then either controlling for these attitudes or preventing such participants from enrolling in the study.

Pretest 1 identified seven primary sources of displeasure with recall-based power priming, with the most common sources being a general dislike for recalling past events, disliking writing, and feeling like researchers were asking too much from participants. Participants' reasons for dropping out or considering dropping out of a recall-based condition largely match these dominant reasons, though participants who considered dropping out focused predominantly on dislike for writing and low payment. The less common reasons revealed in Pretest 1 include recall-based power priming being boring, depressing, seeming ineffective, and that some

workers simply disliked having power or having power used against them. Researchers may struggle to make recall-based power priming less effortful (however, see Lammers et al., 2017), but some of these non-effort-related sources may be tied to a perception of meaninglessness in workers' participation, which researchers can address. Online conversations among MTurk workers even mock the apparent pointlessness of some recall-based priming studies. For example, one worker, in response to being asked, "Which HITs do you stay away from?" responded, "The ones that ask you to recall a time when nobody liked you, not even your dog, and they want you to write about it for 5 minutes, and then answer questions about salad dressing . . ." (luxical, 2016). Explaining the purpose of the task may be an effective method for addressing such meaninglessness (D. Chandler & Kapelner, 2013). Given that a sense of meaninglessness degrades the quality of participants' work, these complaints should be taken seriously (Paolacci & Chandler, 2014). However, researchers would need to take care to make sure attempts at reducing meaninglessness do not interfere with the prime—for example, by introducing demand effects.

Participants assigned to recall-based conditions could cheat by either copy/pasting pre-written responses or by restarting the study in the hope of being assigned to a more preferable condition. Overall, this behavior was uncommon. Although participants engaging in such behavior tended to fail the manipulation check at the end of the study, these findings still indicate that researchers should be cognizant of such cheating behavior and have procedures in place for detecting it.

Drop-out rates differed dramatically between recall-based and hierarchical-role conditions: Nearly no one dropped out of the hierarchical-role conditions. However, dropouts did not differ dramatically between recall-based conditions, indicating these dropouts may not pose a serious threat to internal validity (Zhou & Fishbach, 2016). Hierarchical-role conditions, given their design, also were not at risk of the cheating observed in the recall-based conditions. However, it is important to note that despite hierarchical-role priming's strengths, this article does not demonstrate that such priming is the ideal manipulation of power. Rather than manipulating power, both recall-based and hierarchical-role power priming may communicate to participants the sort of outcomes researchers are seeking, with which participants then comply (Sturm & Antonakis, 2015). MTurk samples may be especially vulnerable to these demand effects. Although the between-group designs typical of priming research should help obscure the exact variable being manipulated (Paolacci & Chandler, 2014), workers' prior experience with every level of the independent variable undermines this key strength. Furthermore, it can be challenging to verify that the prime was effective without introducing further demand effects (Sturm & Antonakis, 2015). Despite these concerns, this article's findings are still valuable. Recent research challenges such criticisms of priming (Lammers et al., 2017),

and, given priming's tremendous popularity on MTurk, many power researchers appear undaunted by priming's critics.

Beyond power, eliciting the expected emotional responses may be important for other concepts manipulated by recall-based methodology, like attachment and status. In fact, existing research already shows how emotion can interfere with status processes (Lovaglia & Houser, 1996). Displeasure is also not exclusive to recall-based priming. Given how widespread displeasure appears to be among MTurk workers, the empirical consequences of such displeasure deserve further scrutiny.

Limitations

Representativeness. One limitation of this research is that it is not representative of the entire MTurk population. This is due to workers self-selecting into each study, more experienced workers enrolling more quickly (Arechar, Kraft-Todd, & Rand, 2017; Casey, Chandler, Levine, Proctor, & Strolovitch, 2017), and because I required workers to meet or exceed a minimum "reputation" (i.e., percent of work deemed acceptable by previous requesters), which, to be meaningful, required workers to have completed a minimum number of HITs (Peer, Vosgerau, & Acquisti, 2014). A possible disadvantage of not examining relatively inexperienced workers is that displeasure in recall-based methods may partly be a function of repeated exposure. A less experienced sample may therefore associate with fewer dropouts and stronger results. However, given that complaints regarding the "repetitiveness" of recall-based power priming were not among the most commonly provided explanations for disliking the manipulation in Pretest 1, I expect the more common sources of displeasure (i.e., disliking recalling events involving power, disliking writing, and feeling like requesters are asking too much) would continue to elicit displeasure regardless of worker experience—but with the data available, I cannot speak to how findings are affected by the non-representativeness of my sample.

Main study design. There were several limitations specific to the replication. Although I propose that the source of the recall-based prime's disadvantage relative to the hierarchical-role prime is displeasure directed at the recall prime, I do not directly measure dislike directed at either prime during the replication. Therefore, I cannot directly assess the role displeasure played in explaining the hierarchical-role prime's advantages relative to the recall-based prime. Next, regarding the emotion measure in the replication, I used a scale provided by Lawler et al. (2008); however, it may have been more useful to continue to use the PANAS, as Fast et al. (2012) had, due to its strong psychometric properties and validity (Watson et al., 1988). Beyond measurement, the design of the study could have been strengthened. For example, I selected the hierarchical-role prime because it lacked the most significant sources of displeasure identified by Pretest 1. However, this approach did not assess if any specific

aspect of the recall-based prime was responsible for its weakness relative to the hierarchical-role prime (e.g., writing, recalling potentially unpleasant events). An analysis that independently varied these potential sources of displeasure may have been useful for constructing new power primes. Finally, it is conceivable that the control condition should have produced a negative change in emotion due to the inclusion of writing, which was one source of displeasure identified in Pretest 1. A replication study that varied the elements of the recall-based power prime that elicited displeasure would have been better suited for explaining why the control condition functioned as expected in this case.

Conclusion

This article finds recall-based power priming to be commonly encountered, broadly disliked, and furthers our knowledge of the challenges researchers face when conducting recall-based power-priming studies on MTurk. These challenges include high drop-out rates, cheating, and potentially an inability to replicate past research. This article not only demonstrates the advantages of an alternative prime for overcoming these challenges but also indicates several actions that may be beneficial for researchers who prefer to continue using recall-based primes to study power with MTurk samples. These actions include measuring the extent to which recall-based power priming qualifies as a “negative context” for participants (Leach & Weick, 2018), communicating the importance of participants’ responses (D. Chandler & Kapelner, 2013), and employing techniques for detecting the reuse of previously written responses. This article therefore provides multiple avenues which may strengthen the study of power on MTurk.

Acknowledgments

I thank Jeffrey Lucas, Long Doan, and my anonymous reviewers for their constructive feedback, and I thank Nathanael Fast for providing the documentation necessary for the replication study.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Partial funding for open access provided by the UMD Libraries’ Open Access Publishing Fund.

ORCID iD

R. Gordon Rinderknecht  <https://orcid.org/0000-0002-5917-406X>

Supplemental Material

Supplemental material for this article is available online.

References

- airhef. (2015, May 15). Re: Which HITs do you stay away from? [Online forum]. Retrieved from https://www.reddit.com/r/mturk/comments/362cc3/ever_have_those_hits_where_you/
- Anderson, C., & Berdahl, J. L. (2002). The experience of power: Examining the effects of power on approach and inhibition tendencies. *Journal of Personality and Social Psychology*, 83, 1362-1377.
- Arechar, A. A., Kraft-Todd, G., & Rand, D. G. (2017). Turking overtime: How participant characteristics and behavior vary over time and day on Amazon Mechanical Turk. *Journal of the Economic Science Association*, 3, 1-11.
- Ayres, R. U., & Martinás, K. (2005). 120 wpm for very skilled typist. In R. U. Ayres (Ed.), *On the reappraisal of microeconomics: Economic growth and change in a material world* (p. 41). Cheltenham, UK: Edward Elgar Publishing.
- Behrend, T., Sharek, D., Meade, A., & Wiebe, E. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods*, 43, 800-813.
- Berdahl, J. L., & Martorana, P. (2006). Effects of power on emotion and expression during a controversial group discussion. *European Journal of Social Psychology*, 36, 497-509.
- 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.
- bicZoid. (2016, February 5). Re: Which HITs do you stay away from? [Online forum]. Retrieved from https://www.reddit.com/r/mturk/comments/448a6y/which_hits_do_you_stay_away_from/
- 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.
- Busarovs, A. (2013). Ethical aspects of crowdsourcing, or is it a modern form of exploitation. *International Journal of Economic & Business Administration*, 1, 3-14.
- Casey, L. S., Chandler, J., Levine, A. S., Proctor, A., & Strolovitch, D. Z. (2017). Intertemporal differences among MTurk workers: Time-based sample variations and implications for online data collection. *SAGE Open*, 7, 1-15.
- Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon’s MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, 29, 2156-2160.
- Chandler, D., & Kapelner, A. (2013). Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior & Organization*, 90, 123-133.
- Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, 46, 112-130.
- clacent01. (2015, August 24). Re: Top 10 Turking gripes [Online forum]. Retrieved from https://www.reddit.com/r/mturk/comments/3i8fw3/top_10_turking_gripes/
- Dubois, D., Rucker, D. D., & Galinsky, A. D. (2012). Super size me: Product size as a signal of status. *Journal of Consumer Research*, 38, 1047-1062.
- Emerson, R. M. (1962). Power-dependence relations. *American Sociological Review*, 27, 31-41.
- Fast, N. J., Sivanathan, N., Mayer, N. D., & Galinsky, A. D. (2012). Power and overconfident decision-making. *Organizational Behavior and Human Decision Processes*, 117, 249-260.

- Fiske, S. T. (1993). Controlling other people: The impact of power on stereotyping. *American Psychologist*, 48, 621-628.
- Galinsky, A. D., Gruenfeld, D. H., & Magee, J. C. (2003). From power to action. *Journal of Personality and Social Psychology*, 85, 453-466.
- Galinsky, A. D., Rucker, D. D., & Magee, J. C. (2015). Power: Past findings, present considerations, and future directions. In M. Mikulincer & P. R. Shaver (Eds.), *APA handbook of personality and social psychology* (Vol. 3, pp. 421-460). Washington, DC: American Psychological Association.
- Gleibs, I. H. (2017). Are all "research fields" equal? Rethinking practice for the use of data from crowdsourcing market places. *Behavior Research Methods*, 49, 1333-1342.
- Hauser, D., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than subject pool participants. *Behavior Research Methods*, 48, 400-407.
- Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review*, 110, 265-284.
- Lammers, J., Dubois, D., Rucker, D. D., & Galinsky, A. D. (2017). Ease of retrieval moderates the effects of power: Implications for the replicability of power recall effects. *Social Cognition*, 35, 1-17.
- Lawler, E. J., Shane, R. T., & Yoon, J. (2008). Social exchange and micro social order. *American Sociological Review*, 73, 519-542.
- Leach, S., & Weick, M. (2018). From grumpy to cheerful (and back): How power impacts mood in and across different contexts? *Journal of Experimental Social Psychology*, 79, 107-114.
- Lovaglia, M. J., & Houser, J. A. (1996). Emotional reactions and status in groups. *American Sociological Review*, 61, 867-883.
- luxical. (2016, February 5). Re: Which HITs do you stay away from? [Online forum]. Retrieved from https://www.reddit.com/r/mturk/comments/448a6y/which_hits_do_you_stay_away_from/
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44, 1-23.
- Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, 23, 184-188.
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5, 411-419.
- Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, 46, 1023-1031.
- Pittman, M., & Sheehan, K. (2016). Amazon's Mechanical Turk a digital sweatshop? Transparency and accountability in crowdsourced online research. *Journal of Media Ethics*, 31, 260-268.
- Sturm, R. E., & Antonakis, J. (2015). Interpersonal power: A review, critique, and research agenda. *Journal of Management*, 41, 136-163.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063-1070.
- Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology*, 111, 493-504.

Author Biography

R. Gordon Rinderknecht is a PhD student in the Department of Sociology at the University of Maryland. He is broadly interested in power, status, interpersonal relationships, and online research methodology. His current research focuses on the situational factors shaping the relationship between social isolation and loneliness.