



Full length article

Microworkers as research participants: Does underpaying Turkers lead to cognitive dissonance?



Bingjie Liu*, S. Shyam Sundar

Media Effects Research Laboratory, Donald P. Bellisario College of Communications, Penn State University, 115 Carnegie Building, University Park, PA 16802, USA

ARTICLE INFO

Keywords:

Amazon's Mechanical Turk
Underpayment
Cognitive dissonance
Enjoyment
Effort

ABSTRACT

Social science researchers increasingly rely on microworkers to serve as study participants, paying them very little compared to participants recruited from other venues. This has raised ethical concerns and questioned the validity of research based on microworkers. Informed by cognitive dissonance theory, we conducted two between-subjects experiments to examine the effects of underpaying Amazon Mechanical Turk workers (Turkers) on their perceptions and their actual performance on criteria crucial to online social science research. Data show that underpaid Turkers experienced 'cognitive dissonance' such that those paid as low as \$0.25 rated their participation as more important than those who were paid higher. Perceived importance was associated with other positive perceptions and demand characteristics. Nevertheless, underpaying Turkers increased dropout rate, reduced their level of effort in answering open-ended questions and undermined perceived agency. We discuss the ethical and practical implications of underpaying microworkers.

1. Introduction

Social science research is increasingly relying on recruiting participants from platforms of microworkers such as Amazon's Mechanical Turk (MTurk, <https://www.mturk.com>) (Proaps, Landers, Reddock, Cavanaugh, & Kantrowitz, 2014; Schneider & Cheung, 2013; Shay et al., 2014, 2016). There are more than half a million workers doing human intelligence tasks (HITs) on MTurk (Paolacci & Chandler, 2014; Pittman & Sheehan, 2016). The large pool of subjects constitutes a "global, on-demand, 24 × 7 workforce" (Amazon Mechanical Turk, 2017), affording fast iteration between experimental tests (Mason & Suri, 2012). As advertised on its website, MTurk requestors get "thousands of HITs completed in minutes" (Amazon Mechanical Turk, 2017).

Past research suggests that responses from Turkers as participants are generally as reliable as participants recruited from conventional sources (Buhrmester, Kwang, & Gosling, 2011; Casler, Bickel, & Hackett, 2013; Holden, Dennie, & Hicks, 2013). However, compared with participants recruited from other sources, online microworkers are rewarded much less for completing the very same task (Mason & Suri, 2012). The large supply of cheap labor has made MTurk the go-to place for researchers to "obtain high-quality data inexpensively and rapidly" (Buhrmester et al., 2011, p. 3). The average MTurk worker (Turker) gets paid 5 cents for a 10- to 15-min survey and earns about \$3 per hour by completing a series of studies (Pittman & Sheehan, 2016). The typical compensation ranges from \$0.01 to \$0.10 per HIT regardless of

the nature of the HITs and whether it takes seconds or hours to complete (Mason & Watts, 2010). More than two thirds of Turkers considered themselves exploited (Busarovs, 2013; Dholakia, 2015) and some call MTurk "the digital sweatshop" (Scholz, 2012). Turkers have written to Amazon's CEO protesting "I am a human being, not an algorithm" seeking for more institution-level protection of their rights for getting fair payment (Harris, 2014).

Being paid so little, these workers might be undergoing "cognitive dissonance"—the gap between effort and reward might induce an unpleasant state, in order to overcome which individuals are known to adjust their attitudes in line with their behaviors so that they are consistent with each other (Festinger, 1962). Such psychological and behavioral outcomes from underpaying Turkers raise both ethical concerns regarding their well-being and pragmatic concerns regarding data quality such that low payment might reduce Turkers' engagement during participation and thereby threaten the validity of the study.

In response to the ethical and practical implications of underpaying Turkers, the current study aims to answer the following two questions: 1) how Turkers' perceptions about their participation vary as a function of their pay rates, a question that has been neglected in existing research on MTurk, and 2) how their effort varies as a function of their pay rates.

* Corresponding author.

E-mail addresses: bxl5252@psu.edu (B. Liu), sss12@psu.edu (S.S. Sundar).

2. Literature review

“Incentives are the essence of economics” (Prendergast, 1999, p. 7). Monetary reward influences individuals’ satisfaction, motivation and task performance especially when reward and performance are linked (Cerasoli, Nicklin, & Ford, 2014; Heyman & Ariely, 2004; Lazear, 2000; Prendergast, 1999; Scott, 2000). Companies have been trying to design the best incentive scheme to maximize the motivational effect of monetary reward (Prendergast, 1999).

According to the literature on the psychology of motivation, there are two types of motivation, intrinsic motivation and extrinsic motivation, which primarily drive human activities and influence an individual’s performance jointly (Cerasoli et al., 2014). Intrinsic motivation is about individuals’ interest and enjoyment in an activity, rather than the outcome of the activity, for example, studying a topic because you find it interesting. On the other hand, extrinsic motivation is based on the outcome of the activity, such as tangible rewards, social rewards, verbal rewards, and so on, for example, studying a topic in order to get good grades (Deci, Koestner, & Ryan, 1999; Deci & Ryan, 1985).

Although some research suggests that tangible rewards might undermine intrinsic motivation (Deci, 1971; Deci & Ryan, 1985), several meta-analyses have found that reward does not harm intrinsic motivation (Cameron & Pierce, 1994). Most applied studies have found positive or null effects of reward on intrinsic motivation (Eisenberger & Cameron, 1996; Eisenberger, Pierce, & Cameron, 1999). A more comprehensive meta-analysis suggests that the negative effect of tangible rewards only occurs when individuals are highly interested in the activity (Cameron, Banko, & Pierce, 2001).

Based on this literature, the possibility of tangible rewards undermining task performance is more concerning when the payment level is high and the intrinsic interest level is high to begin with, which seems not to be the case for Turkers who are normally paid too low. When it comes to Turkers, although many do HITs out of intrinsic motivation (Paolacci & Chandler, 2014; Paolacci, Chandler, & Ipeirotis, 2010; Pittman & Sheehan, 2016), more and more Turkers treat doing HITs as a real job to make their ends meet (Mason & Suri, 2012; Ross, Irani, Silberman, Zaldivar, & Tomlinson, 2010). However, the pay rate is much lower than the federal minimum wage of \$ 7.25 per hour in the United States (United States Department of Labor, 2018). The discrepancy between the payment they get from doing HITs and the fact they have to spend effort to get the HITs done might lead to cognitive dissonance.

2.1. Cognitive dissonance theory

According to cognitive dissonance theory, there is a natural human drive towards ‘assonance’ among their cognitions such as their beliefs, attitudes, thoughts and behaviors (Festinger, 1962). In one of the classic studies on cognitive dissonance (Festinger & Carlsmith, 1959), participants completing a monotonous task were paid to tell other future participants that the experiment that they just went through was very interesting for either \$1 or \$20. Counterintuitively, participants who were paid \$1 reported more enjoyment, more perceived scientific importance of the experiment, and more interest in future participation than those in the \$20 condition.

As cognitive dissonance theory explains, when individuals behave in ways that are inconsistent with their attitudes, there is a natural drive towards ‘assonance,’ and so they adjust their attitudes, post-behavior, in order to account for the behavior. In the aforementioned study, when study participants are induced to do something that is contrary to their private opinion, such as telling others that certain tasks were interesting when in fact it was boring, for a reward that does not merit such a lie (a mere dollar in this case), cognitive dissonance occurs, which puts individuals in an unpleasant motivational state, driving them to reduce the dissonance by adjusting their attitudes in accordance with their behavior (Eisenberger et al., 1999). Resolving cognitive dissonance by

changing behaviors or attitudes is a reliable phenomenon, verified by many empirical studies in subsequent decades that have shown support for cognitive dissonance theory (Adams & Rosenbaum, 1962; Aronson & Mills, 1959; Brehm, 1956; Cooper, 2007; Dickerson, Thibodeau, Aronson, & Miller, 1992; Hinojosa, Gardner, Walker, Coglisier, & Gullifor, 2017).

The norm of low payment leaves little choice for Turkers, especially for those who count on MTurk to make ends meet. According to cognitive dissonance theory then, Turkers induced to do HITs with too low a payment should experience cognitive dissonance characterized by the inconsistency between the effort they expend and the small reward they get. As information about the task is limited in the beginning, cognitive dissonance and reappraisal might occur when the task unfolds itself and some effort is sunk. Therefore, those who are underpaid should adjust their attitude to the HIT to be more positive so as to keep their attitude and behavior consonant, which might further lead them to have more positive experience and spend more effort in participation.

Rationalization as a solution to cognitive dissonance is most likely to be adopted when after accepting the task, participants realize that performing the task, with some time and effort already sunk, is not congruent with the low payment. Alternatively, participants’ dissonance could also be resolved by changing behaviors. For example, in a study by Dickerson et al. (1992), cognitive dissonance among participants was induced by making them feel hypocritical about their showering habits as a result of being reminded of their past shower habits and making a public commitment urging others to take shorter showers. Those participants ended up reducing actual shower time to keep their behaviors in line with the public commitment they made earlier.

For Turkers who are underpaid, as the HIT unfolds itself, they can simply reduce the effort they put into their participation without going through the process of rationalization, to be in line with the cognition of low reward. For example, higher monetary reward increases attentional effort in the flanker task, a test of selective visual attention which asks participants to judge whether a target stimulus belongs to a particular category (Eriksen & Eriksen, 1974; Hübner & Schlösser, 2010). Likewise, it is possible that Turkers adjust the effort they expend based on the reward, in order to reduce dissonance and achieve assonance.

2.2. Turkers’ performance and subjective experience

Past research on the effect of pay rate on Turker’s performance has yielded inconsistent findings. Research on Turkers has examined effect of payment on performance in both objective and subjective tasks. Overall, research has found that higher payment expedites data collection (Buhrmester et al., 2011; Feng, Besana, & Zajac, 2009). Among research on objective tasks, in which questions have a correct answer, some found accuracy of answering math and text questions increased as payment increased from \$4 to \$8 and \$10 per hour (Aker, El-Haj, Albakour, & Kruschwitz, 2012), whereas others did not find an effect when the payment differential was small. For example, participants were only rewarded \$0.10 and \$0.30 for a 2–5 min survey (Rouse, 2015), \$0.01, \$0.02, \$0.03, and \$0.10 per HIT (Aker et al., 2012), \$0.01, \$0.05, and \$0.10 per image set in an image ordering task (Mason & Watts, 2010), and \$0.02, \$0.10, and \$0.50 for completing a personality questionnaire (Buhrmester et al., 2011). While accuracy may not vary, it is not clear if low payment amount induced cognitive dissonance.

For social science research with self-report measurements, participants’ effort is manifest in their engagement with the stimuli (in the case of online experiments) and questions. Sufficient engagement can be indicated by completion of the study, attention paid to content on the questionnaire, which also means longer time spent on the questionnaire—in general, following researchers’ instructions and meeting researchers’ requirements in the questionnaire.

In the current study, we included an indicator of completion of the

study (dropout rate), measures of attention captured by attention check questions, the dispersion among items of a scale, time spent on the questionnaire (Cameron & Pierce, 1994), measures of effort put towards meeting researchers' requirements (which was captured by number of words in response to open-ended questions), and demand characteristics, i.e., the tendency among participants to provide responses that they deem as being helpful for the purpose of the research (Orne, 1962). Although researchers generally prefer to have participants spend more effort during participation, effort leading to demand characteristics is undesirable for researchers, as they lead to biased and inauthentic responses.

Existing research on the effect of financial incentives on Turkers has not paid enough attention to their subjective experience. Subjective experience with an activity is often evaluated based on perceived importance of the activity, interest/enjoyment, tension/pressure, and self-determination/agency in research on motivation and task performance (Arnold, 1976; Cherrington, Reitz, & Scott, 1971; Eisenberger et al., 1999). In the current study, Turkers' subjective experience is evaluated based on perceived value of their participation (Eisenberger et al., 1999), enjoyment (Arnold, 1976; Cherrington et al., 1971), (reduced) tension, and agency (perceived choice) (Eisenberger et al., 1999) in their participation. If cognitive dissonance is in play, then these aspects of subjective experience would be rated higher by participants when they are paid less, as they signal an attempt at post-behavioral rationalization of their decision to participate.

Based on the discussion above, we propose the following hypotheses:

H1. Compared with higher payment, low payment enhances perceived importance of the activity of doing the study.

H2. Compared with higher payment, low payment enhances positive perceptions of the experience in terms of a) enjoyment, and b) perceived choice, and reduces negative experience in terms of c) tension.

H3. Compared with higher payment, low payment enhances effort expended on participation as a result of enhanced perceived importance in terms of a) attention, b) number of words, c) time, d) dispersion within scales, and e) demand characteristics.

Alternatively, it is also possible that Turkers, when underpaid, simply reduce their effort without changing their private attitudes and subjective experience.

H4. Compared with higher payment, low payment reduces effort expended on participation in terms of a) study completion (measured by dropout rate), b) attention, c) number of words, d) time, e) dispersion within scales, and f) demand characteristics.

This last hypothesis represents an alternate, behavioral path for achieving the assonance when experiencing cognitive dissonance theory. Therefore, H3 and H4 are competing hypotheses in this study.

2.3. Individual differences

Although both rationalizing the low payment and reducing effort can reconcile inconsistencies among cognitions for Turkers, not everyone can achieve assonance through rationalization. Those who see intrinsic value in doing the HITs are more likely to engage in rationalization whereas those who see the HITs as a vehicle for obtaining external rewards (be it in the form of compensation or approval rating) are likely to strategically moderate expenditure of their effort.

Reward on MTurk is essentially performance-based, with the researcher approving or rejecting a worker's response. Although most HITs in social science research collect subjective responses from workers, researchers have come up with objective criteria to judge the quality of workers' performance. For example, they use attention check questions such as "please choose 1 for this item" among scale items

(Hauser & Schwarz, 2016). Failure to follow such instructions suggests a lack of attention. HIT requesters could reject their work if they see poor performance on such objective criteria, which not only results in denial of the monetary reward for Turkers but also brings down their approval rating on MTurk (Mason & Watts, 2010), which is visible to other researchers and might cause one to be excluded from future HITs if one's historic approval rate is too low. For Turkers who care more about extrinsic tangible rewards, enhanced positive attitudes towards the intrinsic value of doing the HITs as a result of rationalization may not be sufficient for them to achieve assonance between the considerable amount of effort and the paltry compensation for that effort. Instead, reducing effort is the more viable approach to reduce dissonance and achieve cognitive assonance. Therefore, we also hypothesize.

H5. Compared with those who do not care about getting approval, Turkers who care about getting approval are less likely to rationalize the underpayment by distorting their perceptions of a) perceived importance of their participation, b) enjoyment, c) perceived choice, and d) tension.

H6. Compared with those who do not care about getting approval, Turkers who care about getting approval are more likely to reduce effort when underpaid in terms of a) attention, b) number of words, c) time, d) dispersion within scales, and e) demand characteristics.

We conducted two experiments in the paradigm of classic cognitive dissonance studies to detect if cognitive dissonance is induced by underpaying Turkers.

3. Experiment 1

3.1. Method

The task in the current study was completing a questionnaire on Qualtrics about individuals' media use and their perceptions of various environments, which takes about 15–20 min to complete. Participants were randomly assigned to either the high-payment condition (\$1.50) or the low-payment condition (\$0.50) for the same task. To allow for randomization and avoid the problem of self-selection, the reward was announced as "\$0.50" on MTurk when publishing the HIT. Payment manipulations were implemented at the very beginning of the questionnaire, with participants in the high-payment condition being told "Thanks for your interest! We just learned of additional funding from our grant sponsor, so can offer you as much as \$1.50 for your participation today." Participants in the low-payment condition would be reminded that "you would be paid \$0.50 for your participation." As participants were not aware of the payment made to others, relative deprivation was ruled out as a possible explanatory factor. As they were debriefed in the very end, participants were paid \$1.50 regardless of their assigned conditions.

The questionnaire included two parts. The first part was the cover story – questionnaire designed for another study measuring media use and one's perceptions of pictures of nature, purported to be the actual study for which we sought their participation. The second part containing measures of interest for the current study was presented at the very end of the questionnaire, purported for participants to provide some feedback on their experience of participation, to assess participants' perceptions after cognitive reappraisal.

3.1.1. Participants

All participants were Turkers located in the United States. Originally 145 Turkers participated in the study whereas 30 of them dropped out in the middle, resulting in a total of 115 completes. Participants' age ranged from 19 to 72 ($M = 36.56$, $SD = 13.21$), with gender almost equally distributed (Male = 55, 47.8%). The majority of the participants (70.40%) were Caucasian, followed by Asian ($n = 18$, 15.70%), African American ($n = 8$, 7.00%), Hispanic/Latino ($n = 3$, 2.6%),

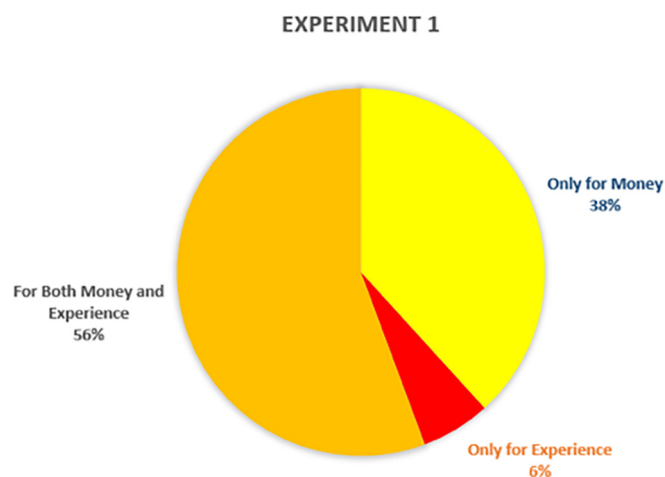


Fig. 1. Motives for working on Mturk in Experiment 1.

Mixed ($n = 3$, 2.6%), and Native American ($n = 2$, 1.7%). More than half of the participants (60%) had bachelor or higher degrees. Among all the participants, 45.2% had yearly household income lower than \$50,000 and 11.3% had higher than \$100,000. Participants' perceived financial status was moderate ($M = 4.04$, $SD = 1.66$, on a 1–7 very bad – very good scale).

Nineteen participants (16.52%) had pre-existing approval rate of 100%, whereas the rest did not have any record of approval rate by the time of their participation.

In terms of the reasons for working on MTurk, 28.70% chose for living, 84.30% for extra pocket money, 39.10% for fun, 40.00% for supporting research, 43.50% for helping people doing research, 28.70% to learn what is out there. Other reasons mentioned include self-enrichment, learning, and mind exercise. As shown in Fig. 1, 38.26% participants worked only for money and only 6.09% worked only for reasons beyond money. In other words, 93.91% worked at least partially for money.

3.1.2. Measures

Measures of Turkers' subjective experience (perceived importance of their participation, enjoyment, perceived choice, and tension) were adapted from four subscales of Intrinsic Motivation Inventory (IMI) (Deci, Eghrari, Patrick, & Leone, 1994; Ryan, Mims, & Koestner, 1983). Participants rated their agreement to each statement on a seven-point Likert-type scale from 1 “strongly disagree” to 7 “strongly agree.”

Perceived importance. The 5-item scale measured perceived importance of their participation, with items such as “I believe participating in this study could be of some value to me.” Cronbach's $\alpha = 0.94$ (Cronbach's $\alpha = 0.87$ in Experiment 2).

Enjoyment. The 8-item scale measured enjoyment during participation, with items such as “This study was fun to do.” Cronbach's $\alpha = 0.93$ (Cronbach's $\alpha = 0.93$ in Experiment 2).

Perceived choice. The 7-item scale measured perceived autonomy in participation with items such as “I believe I had some choice about participating in this study.” Cronbach's $\alpha = 0.85$ (Cronbach's $\alpha = 0.86$ in Experiment 2).

Tension. The 5-item scale measured anxiety during participation, with items such as “I felt very tense while filling out the questionnaire.” Cronbach's $\alpha = 0.77$ (Cronbach's $\alpha = 0.79$ in Experiment 2).

Deserved payment. At the end of the questionnaire, participants were asked “How much do you think you deserve to be paid?”

Dropout rate. Dropout rate was computed as the ratio of number of participants who did not complete the questionnaire to the number of all participants who had started the study.

Attention. The questionnaire included four attention check questions explicitly asking participants to check a specific point on the 7-

	Strongly disagree							Strongly agree
If I need an emergency loan of \$500, there is someone I could ask to loan it to me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My friends really try to help me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I can count on my friends when things go wrong.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have friends with whom I can share my joys and sorrows.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I can talk about my problems with my friends.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Fig. 2. Social support scale in questionnaire.

point Likert-type scale with instructions such as “Please select the second from left point on the scale for this item.” Correctly answering these questions indicates that participants paid attention during participation.

Number of words. Number of words participants put in response to the open-ended question “Can you tell us more about why you enjoy or do not enjoy using social media? Details are preferred” is regarded as a proxy of the effort they put into their participation in the study. Due to the non-normality of the distribution of the raw number of words ($Skewness = 2.75$, $SE = 0.23$; $Kurtosis = 9.32$, $SE = 0.45$), square root of the raw score was calculated and used in analyses ($Skewness = 0.91$, $SE = 0.23$; $Kurtosis = 1.83$, $SE = 0.45$).

Time. Time was measured as minutes spent on the questionnaire, as recorded by Qualtrics.

Dispersion within scales. Items measuring the same construct were grouped together and displayed in the same table matrix in the questionnaire, as shown in Fig. 2. Cognitively lazy participants tend to have vertically aligned choices (i.e., responses with less variance). Higher dispersion among items within each scale indicates more attentional effort. Dispersion was computed by averaging the standardized standard deviation (SD) across all scales in the questionnaire.

Demand characteristics. We told participants that “the primary goal of this study is to see if thinking about one's friends can make one feel a greater sense of community and perceive a higher level of social support. We wonder how you feel about the social support you have right now,” followed by measures of perceived available social support. Higher score on the social support scale indicates higher demand characteristics, Cronbach's $\alpha = 0.92$ (Cronbach's $\alpha = 0.93$ in Experiment 2).

Approval concern. Participants' concern about getting approval was measured with two items that “I put effort to get a higher chance to be approved by the researcher” and “I put effort to have a better reputation on MTurk.” Cronbach's $\alpha = 0.80$ (Cronbach's $\alpha = 0.89$ in Experiment 2). Most of the participants put considerable effort in the study to get their work approved ($M = 5.42$, $Median = 6.00$, $SD = 1.72$).

3.2. Results

3.2.1. Manipulation check

Score of expectancy violation on payment was calculated by subtracting the amount of deserved payment¹ from the actual payment. Violation in low-payment condition was negative ($M = -0.71$, $SE = 0.10$) and violation in the high-payment condition was positive ($M = 0.06$, $SE = 0.07$), $t(104) = 6.66$, $p = .000$, one-tailed. Therefore, the manipulation of “underpaid” and “overpaid” was successful.

¹ Participants' perceived deserved payment ranged from 0.50 to 4.00, $M = 1.34$, $SD = 0.60$. Since the measure for deserved payment was administered at the end of the questionnaire, there was evidence of an anchoring effect—participants in the low-payment condition said that they deserved lower payment ($M = 1.21$, $SE = 0.10$) and those in the high-payment condition ($M = 1.43$, $SE = 0.07$), $t(104) = 1.91$, $p = .059$, two-tailed. Individual-difference variables had no significant effects on deserved payment (age, $r = 0.093$, $p = .342$, two-tailed; gender, $t(104) = 1.525$, $p = .13$, two-tailed; motivation, $t(104) = 0.128$, $p = .899$, two-tailed).

3.2.2. Hypotheses testing

In support of H4a, dropout rate in the low-payment condition (28.40%, 21 out of 74) was significantly higher than that in the high-payment condition (12.70%, 9 out of 71), $\chi^2(1, N = 145) = 5.45$, $p = .020$.

Among those who completed the study, independent-sample t-tests were conducted to compare the two payment conditions on all the criterion variables. H1 was not supported, as perceived importance of participation was not significantly different between low-payment ($M = 5.19$, $SE = 0.19$) and high-payment ($M = 5.54$, $SE = 0.20$) conditions, $t(113) = 1.25$, $p = .210$.

In support of H2b, higher payment led to higher perceived choice ($M = 6.41$, $SE = 0.13$) than low payment ($M = 6.01$, $SE = 0.15$), $t(113) = 2.04$, $p = .020$, one-tailed.

However, payment did not have an impact on other criterion variables measuring their subjective experience (enjoyment, tension) or effort (attention, number of words, time, dispersion within scales, and demand characteristics).

H5 and H6 were not supported as participants' concern about getting approval was not found to moderate the effect of payment on any dependent variable.

However, we found existing approval rate moderated the effect of payment on number of words. Specifically, as shown in Fig. 3, for those with 100% approval rate, low payment led to more words ($M = 5.12$, $SE = 0.74$) than high payment ($M = 3.43$, $SE = 0.70$); for those without record of approval rate, low payment led to fewer words ($M = 3.67$, $SE = 0.34$) than high payment ($M = 4.65$, $SE = 0.31$), $F(1, 111) = 5.67$, $p = .019$, partial $\eta^2 = 0.049$. If we treat the high approval rate as a proxy for caring less about getting approval, the pattern is consistent with H6b.

4. Discussion

Experiment 1 found that lower payment reduced worker's autonomy (perceived choice), and increased the dropout rate. For workers with no approval rate record, high payment motivated them to put more effort as measured by number of words in response to the open-ended question.

However, with a contrast of \$0.50 vs. \$1.50, Experiment 1 did not find a significant difference between low and high payment conditions on most of the dependent variables, suggesting that for the HIT we used, the contrast between \$0.50 vs. \$1.50 was not large enough. To better resemble the underpayment situation on MTurk, a second experiment was conducted with a larger payment differential.

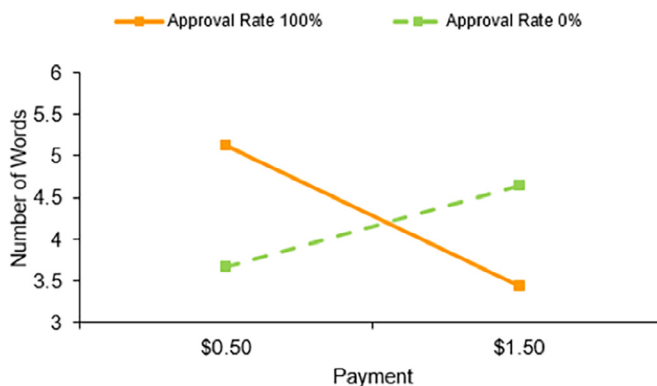


Fig. 3. Moderation effect of approval rate in Experiment 1.

EXPERIMENT 2

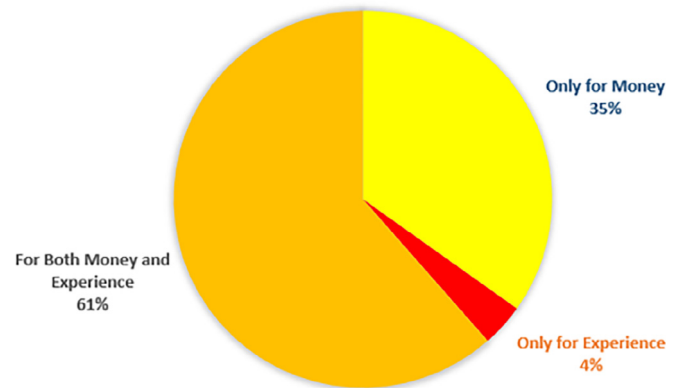


Fig. 4. Motives of working on Mturk in Experiment 2.

5. Experiment 2

5.1. Method

Experiment 2 replicated Experiment 1 with a wider gap in payment, \$0.25 vs. \$3.00. The same HIT was published on MTurk with the reward specified as \$0.25. Following the same procedure, participants were randomly assigned to one of the two conditions.

5.1.1. Participants

Participants were Turkers located in the United States. Originally, 148 Turkers participated in the study, whereas 39 of them dropped out in the middle. Among the 109 Turkers who completed the study, age ranged from 19 to 75 ($M = 36.60$, $SD = 12.38$), with gender almost equally distributed (Male = 47, 43.10%). 76.10% were Caucasian, followed by African American ($n = 12$, 11.00%), Asian ($n = 9$, 8.30%), Hispanic/Latino ($n = 4$, 3.70%), and others ($n = 1$, 0.90%). More than half of the participants (51.90%) had bachelor's or higher degrees. About half (47.70%) had yearly household income lower than \$50,000 and 12.90% had higher than \$100,000.

In terms of the purpose of working on MTurk, 33.00% chose for living, 81.70% for extra pocket money, 40.40% for fun, 45.90% for supporting research, 50.50% for helping people doing research, 32.10% to learn what is out there. As shown in Fig. 4, 34.86% participants worked only for money and only 3.67% worked only for reasons beyond money. In other words, 96.33% worked at least partially for money. Most of the participants put effort into the study to get approval ($M = 5.50$, $Med = 6.00$, $SD = 1.66$).

5.2. Results

5.2.1. Manipulation check

Expectancy violation in the low-payment condition was negative ($M = -0.58$, $SE = 0.07$) and violation in the high-payment condition was positive ($M = 1.08$, $SE = 0.14$), $t(99) = 10.29$, $p = .000$, one-tailed.² Therefore, the manipulation of “underpaid” and “overpaid” was successful. Compared with the differential between high- and low-payment conditions in Study 1 (Cohen's $d = 1.28$), we found an increase in Study 2 as expected (Cohen's $d = 1.92$).

² Perceived deserved payment in Study 2 ranged from 0.00 to 4.00, $M = 1.46$, $SD = 1.02$. There was again an anchoring effect, such that participants in the low-payment condition said that they deserved less ($M = 0.83$, $SE = 0.07$) than those in the high-payment condition ($M = 1.87$, $SE = 0.14$), $t(98) = 6.72$, $p = .000$, two-tailed. Individual difference variables had no significant impact on deserved payment (age, $r = 0.054$, $p = .591$, two-tailed; gender, $t(98) = 0.48$, $p = .632$, two-tailed; motivation, $t(98) = 0.58$, $p = .562$, two-tailed).

Table 1
Effect of payment in experiment 2.

DVs	\$0.25	\$3.00	Statistics	p
	M(SE)	M(SE)		
Perceived Importance	5.75(.14)	5.25(.15)	$t(107) = 2.42$.01
Number of Words	4.61(.33)	5.72(.32)	$t(107) = 2.28$.01
Time	3.58(.11)	3.88(.11)	$t(107) = 1.81$.04
Dropout Rate	40.60%	13.90%	$\chi^2(N = 148) = 13.48$.00

Note. p-values for t-test are one-tailed.

5.2.2. Hypotheses testing

As summarized in Table 1, in support of H4a, participants paid \$0.25 were more likely to drop out in the middle (40.60%, 28 out of 69) than those paid \$3.00 (13.90%, 11 out of 79), $\chi^2(1, N = 148) = 13.48$, $p = .000$.

For those who completed the study, in support of H1b, participants who were paid \$0.25 ($M = 5.75$, $SE = 0.14$) perceived their participation as more important than those paid \$3.00 ($M = 5.25$, $SE = 0.15$), $t(107) = 2.42$, $p = .008$, one-tailed.

In support of H4b, participants who were paid \$0.25 responded with fewer words ($M = 4.61$, $SE = 0.33$) than those who were paid \$3.00 ($M = 5.72$, $SE = 0.32$), $t(107) = 2.28$, $p = .013$, one-tailed.

In support of H4c, participants who were paid less also tended to spend less time on the study ($M = 3.58$, $SE = 0.11$) than those receiving high payment ($M = 3.88$, $SE = 0.11$), $t(107) = 1.81$, $p = .036$, one-tailed.

However, payment did not have an effect on other criterion variables (enjoyment, perceived choice, tension, attention, dispersion within scales, and demand characteristics).

To test whether perceived importance mediates the relationship between payment and effort, as predicted by H3, mediation analyses were conducted with Model 4 in PROCESS Macro (Hayes, 2013). Effect coefficients and bias-corrected 95% confidence intervals (CI) were estimated with 5000 bootstrap samples. Effect is considered statistically significant if the interval does not include zero (Shrout & Bolger, 2002).

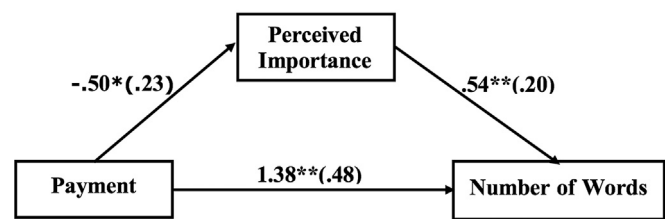
In support of H3b, in line with the prediction of cognitive dissonance theory, results, as shown in Fig. 5, revealed that payment has a negative indirect effect on number of words ($B = -0.27$, $SE = 0.17$, 95% CI [-0.73, -0.04]) through perceived importance of participation. But payment also has a positive direct effect on number of words ($B = 1.38$, $SE = 0.48$, $p < .01$). The total effect of payment on number of words was positive, $B = 1.11$, $SE = 0.49$, $p < .05$.

In support of H3e, as shown in Fig. 6, payment has a negative indirect effect on demand characteristics ($B = -0.19$, $SE = 0.11$, 95% CI [-0.50, -0.03]) through perceived importance. But payment also has a positive direct effect on demand characteristics ($B = 0.71$, $SE = 0.33$, $p < .05$), with the total effect being non-significant.

In addition, in line with the prediction of cognitive dissonance theory, we also found that through perceived importance of their participation, payment had a negative indirect effect on enjoyment ($B = -0.37$, $SE = 0.16$, 95% CI [-0.70, -0.07]) and perceived choice ($B = -0.20$, $SE = 0.10$, 95% CI [-0.46, -0.05]), and a positive indirect effect on tension ($B = 0.18$, $SE = 0.09$, 95% CI [0.04, 0.40]) as shown in Fig. 7. Direct effects of payment on enjoyment, perceived choice, and tension were not significant.

To test H5 and H6 regarding how concern about getting approved moderated the effect of payment, moderation analyses were conducted with ModProbe, a Macro in SPSS which is used for probing single-degree-of-freedom interactions in linear regression models (Hayes & Matthes, 2009). For a continuous moderator (M), such as concern for getting approval in the current study, the Johnson-Neyman technique (Bauer & Curran, 2005; Hayes & Matthes, 2009) powered by ModProbe can “derive the value along the continuum of M at which the effect of X on Y transitions between statistically significant and not significant”

Note. B (SE); * $p < .05$, ** $p < .01$



Indirect effect: $B = -.27$, $SE = .17$, 95% CI [-0.73, -0.04]

Total effect: $B = 1.11$, $SE = .49$, $p < .05$

Fig. 5. Effect of payment on number of words in Experiment 2.

(Hayes, 2012, p. 5).

With the Johnson-Neyman technique, we found that concern for getting approval moderated the effect of payment on perceived choice, tension, and number of words. The critical values of concern for getting approval where the effect of payment on the three dependent variables changed signs are summarized in Table 2. Specifically, in support of H5c and H5d, those who had low concern about getting approval tended to rationalize, as evidenced by the fact that lower payment induced more perception of choice ($B = 0.24$, $SE = 0.12$, $\Delta R^2 = 3.34\%$, $F(1, 105) = 4.03$, $p = .047$), and less tension ($B = -0.25$, $SE = 0.13$, $\Delta R^2 = 3.48\%$, $F(1, 105) = 3.98$, $p = .049$). But for those who had high concern about getting approval, lower payment did not induce more positive experience. In support of H6c, for those who were more concerned about getting approval (> 5.25), lower payment reduced their effort, in terms of the number of words, $B = 0.59$, $SE = 0.29$, $\Delta R^2 = 3.54\%$, $F(1, 105) = 4.06$, $p = .046$.

6. Analysis across both experiments

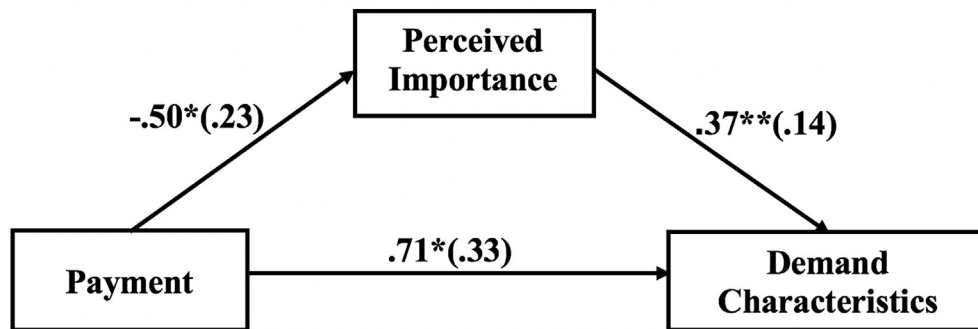
By comparing the two studies, we found higher payment attracted more high-quality workers to participate. With the announced compensation rate as \$0.50, in Experiment 1, 19 participants (out of 115) had an approval rating of 100%. However, in Experiment 2, with the announced compensation rate as \$0.25, only two participants (1.83%) had a pre-existing approval rating of 100%, whereas the rest did not have any record of approval. The percentage of high-quality participants was significantly lower than that in Experiment 1 (16.52%), $\chi^2(1, N = 224) = 14.21$, $p = .000$.

To avoid the problem of self-selection and to ensure random assignment of participants into different conditions, both Experiment 1 and 2 implemented the high-payment manipulation within the questionnaire, which might have introduced the element of elation arising from the pleasant surprise of a multi-fold increase in their compensation. To test if the effect found still held when there was no such pleasant surprise, we compared the \$0.50 condition in Study 1 and \$0.25 condition in Study 2, i.e., the low-payment conditions of both studies wherein no surprise was introduced. Perceived importance among participants paid \$0.25 ($M = 5.75$, $SE = 0.18$) was significantly higher than that for those paid \$0.50 ($M = 5.19$, $SE = 0.18$), $t(92) = 2.36$, $p = .01$, one-tailed, thus showing support for H1.

7. Overall discussion

Although in Experiment 1 we did not observe effect of payment on most of the criterion variables, in Experiment 2, with a larger differential between the payment levels, we found evidence for the cognitive dissonance induced by low payment and that Turkers either rationalized their behavior or adjusted their actual effort to avoid or resolve cognitive dissonance.

Note. B (SE); * $p < .05$, ** $p < .01$



Indirect effect: $B = -.19$, $SE = .11$, 95% CI $[-.50, -.03]$

Total effect: $B = .52$, $SE = .33$, n.s.

Fig. 6. Effect of payment on demand characteristics in Experiment 2.

On the one hand, findings in Experiment 2 suggest that low payment (\$0.25) drives Turkers to change their private attitudes accordingly by construing their participation as more important. Although it appears to benefit researchers that participants think their participation as more important, which is also associated with more positive perceptions and more effort, inducing cognitive dissonance by low payment is both unethical and threatens scientific validity of research. Cognitive dissonance is characterized as a state of unpleasant arousal with the pressure of aligning inconsistent items of information so as to maintain a coherent self-concept (Cooper, 2007; Festinger, 1962; Festinger & Carlsmith, 1959; Hinojosa et al., 2017). For most social science research with online users, cognitive dissonance is not a relevant component. Making participants go through an unpleasant experience that is irrelevant to the research is unjustified and unethical. Moreover, unexpected arousal induced by cognitive dissonance could distort participants' psychological response to experimental manipulations in research, which threatens the validity of the study, especially when arousal and affect are important in the processes being investigated.

Although an alternative explanation may be that the low payment filtered out those who were not intrinsically interested in the study, with only those who deemed the study as important staying, resulting in the higher perceived importance among Turkers in the low-payment condition, the correlation pattern between the perceptual variables and behavioral data refute this possibility. For the alternative explanation to hold, intrinsic belief about the importance of the study should predict both the perceived importance of their participation and their effort, that is, like perceived importance of participation should correlate with the effort measures. However, within the low-payment condition, perceived importance did not correlate significantly with objective effort (demand characteristics, $r = 0.18$, $p = .252$; number of words, $r = 0.03$, $p = .832$; scale dispersion, $r = 0.16$, $p = .318$; time, $r = 0.23$, $p = .156$) as it did in the high-payment condition (demand

Table 2

Moderation effect of concern for getting approval in experiment 2.

DVs	Approval Concern	Payment Effect
Perceived Choice	Low [1, 2.77]	–
	High [2.77, 7]	n.s.
Tension	Low [1, 4.13]	+
	High [4.13, 7]	n.s.
Number of Words	Low [1, 5.25]	n.s.
	High [5.25, 7]	+

Note. n.s. stands for non-significant effect; + stands for positive effect of payment on the corresponding variables; – stands for negative effect of payment on the corresponding variables.

characteristics, $r = 0.30$, $p = .014$; number of words, $r = 0.32$, $p = .007$). However, it was significantly correlated with psychological variables (enjoyment, $r = 0.56$, $p = .000$; perceived choice, $r = .30$, $p = .054$; tension, $r = -0.32$, $p = .042$), suggesting that self-reported perceived importance in low-payment condition is not completely what they “truly” believe about the value of their participation but (at least partially) distorted to be consistent with other cognitions. Therefore, a more likely explanation is that the observed higher perceived importance among Turkers in the low-payment condition is a reflection of their rationalization, in order to resolve cognitive dissonance.

In addition to the rationalization, with Experiment 1 and Experiment 2, we found that when underpaid, Turkers also reduced their actual effort, in order to avoid or resolve cognitive dissonance. The more workers get paid, the less likely they are to drop out in the middle of the study, and the more willing they are to put effort into their participation (more words responding to open-ended questions, more time spent on the study). These findings are at odds with past studies on MTurk, which did not find impact of pay rate on workers' performance (Buhrmester et al., 2011; Mason & Watts, 2010), but it

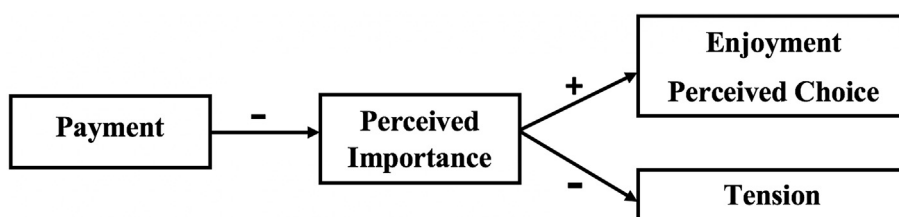


Fig. 7. Effect of payment on workers' perceptions in Experiment 2. Underpaying microworkers for participating in research induces cognitive dissonance.

seems that payment manipulations in previous studies tended to be too subtle and lacked sufficient differential across conditions (e.g., from \$0.02 to \$0.50 in Buhrmester et al., 2011; \$0.01 vs. \$0.1 in Mason & Watts, 2010; \$0.10 vs. \$0.30 in Rouse, 2015). On the other hand, studies with higher payment that is fair and with a larger differential (\$4, \$8 and \$10 per hour) have found higher payment to yield higher quality results (Aker et al., 2012).

One caveat when paying microworkers at a higher rate is its potential for increasing demand characteristics. High payment, while not compromising perceived importance of the study, increased demand characteristics in our study. Therefore, higher payment and other factors indicating the importance of the study should be accompanied by better masking of the true purpose of the study, in order to avoid incentivizing microworkers to provide demand responses.

As we hypothesized, not all Turkers will take the cognitive route to resolve cognitive dissonance through rationalization of their acceptance of the unfair payment by regarding their participation as more important. Those without prior existing approval rate (Experiment 1) and those who cared about getting approved (Experiment 2) were more likely to take the behavioral route by reducing actual effort when underpaid. In the current study, although the majority of the participants (62% in Experiment 1; 65% in Experiment 2) were motivated for reasons other than money, which seems to corroborate the notion that microworkers are intrinsically motivated (Paolacci & Chandler, 2014; Paolacci et al., 2010; Pittman & Sheehan, 2016; Ross et al., 2010), they were at least partially motivated by gaining the material reward such that among participants of the two experiments, 36.60% worked on MTurk only for money, and only 6.70% solely worked for reasons beyond money. In other words, 93.30% were at least partially motivated by the monetary reward. In addition, considering that more and more Turkers are counting on doing HITs to make ends meet and arguably care about getting researchers' approval, underpaying them might lead to poorer performance, thereby producing lower quality data for researchers.

We also found higher payment attracted more high-quality workers to participate. Comparing Experiment 1 and Experiment 2, which announced the compensation rate as \$0.50 and \$0.25 respectively, Experiment 1 had more participants with prior existing approval rate of 100% than Experiment 2. High-end workers were not even attracted to a study offering such low compensation. The effect of low pay rate in excluding a certain segment of Turkers should lead researchers and reviewers to interpret their findings in light of the reward and performance record of the sample.

8. Limitations

The current study has several limitations. The first limitation is the lack of a control condition. In classic studies of incentive effects and cognitive dissonance, there is often a control condition in which participants receive no reward at all. Without a control condition, we are not able to tell for sure whether the effect is due to the lower payment being too low or the higher payment being too high. However, it is impossible to include such a no-payment control condition with Turkers.

As discussed earlier, payment might influence the sample's demographic composition. As the announced pay rate in the current study was either \$0.25 or \$0.50, workers who found them as too low might not be included in the sample. However, considering the current norm of low payment on MTurk, samples in the current study should be representative of the vast majority of workers who routinely accept HITs with such low payment. Due to lack of knowledge about Turkers' demographics, we do not know how representative our sample is, for example, whether having nineteen Turkers with 100% approval rate is representative or not on MTurk.

In the current study, concern for getting approval was used as a proxy for extrinsic motivation. However, both intrinsically- and

extrinsically-motivated participants may have concern for their approval because approval allows both groups to participate in future studies, regardless of their motivation. In future studies, a better measure of intrinsic vs. extrinsic interest as a moderating variable should be used for investigating individual differences in Turkers' psychology.

In addition, measures of conation, which were not used in the current study, may be useful to consider in the future. For example, although underpaid participants contributed a decent amount of effort as a result of cognitive dissonance, they might not want to have the same experience again. Future research could include measures such as whether they would like to do a follow-up study.

9. Conclusion

In conclusion, too low a payment may elicit cognitive dissonance among online microworkers who serve as study participants in social science research. Although our data show that they perceive participation as more valuable when paid less, inducing unnecessary cognitive dissonance is unethical and threatens the validity of the research, especially when arousal and affect are crucial to the processes or outcomes of interest, as is often the case in social science research.

Compared with fair payment, low payment increases dropout rate, discourages high-quality workers from participating, makes them feel less self-determined, and therefore expend less effort in their participation. On the other hand, higher payment motivates them to put more effort, especially by those who have higher concerns of getting approval and those who do not yet have a high approval rating. One caveat is that increasing payment had better be accompanied by study procedures to counter demand characteristics.

In sum, it is imperative to pay microworkers their due for ensuring ethically correct and technically sound practice of scientific research.

References

- Adams, J. S., & Rosenbaum, W. B. (1962). The relationship of worker productivity to cognitive dissonance about wage inequities. *Journal of Applied Psychology*, 46(3), 161–164. <https://doi.org/10.1037/h0047751>.
- Aker, A., El-Haj, M., Albakour, M. D., & Kruschwitz, U. (2012). Assessing crowdsourcing quality through objective tasks. *Proceedings of the eight international conference on language resources and evaluation (LREC'12)* (pp. 1456–1461). Istanbul, Turkey: European Language Resources Association (ELRA) ISBN 9782951740877.
- Amazon Mechanical Turk. (2017). Retrieved from <https://www.mturk.com/mturk/welcome>.
- Arnold, H. J. (1976). Effects of performance feedback and extrinsic reward upon high intrinsic motivation. *Organizational Behavior & Human Performance*, 17(2), 275–288. [https://doi.org/10.1016/0030-5073\(76\)90067-2](https://doi.org/10.1016/0030-5073(76)90067-2).
- Aronson, E., & Mills, J. (1959). The effect of severity of initiation on liking for a group. *Journal of Abnormal and Social Psychology*, 59(2), 177–181. <https://doi.org/10.1037/h0047195>.
- Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research*, 40(3), 373–400. https://doi.org/10.1207/s15327906mbr4003_5.
- Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. *Journal of Abnormal and Social Psychology*, 52(3), 384–389. <https://doi.org/10.1037/h0041006>.
- 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(1), 3–5. <https://doi.org/10.1177/1745691610393980>.
- Busarovs, A. (2013). Ethical aspects of crowdsourcing, or is it a modern form of exploitation. *International Journal of Economics & Business Administration*, 1(1), 3–14.
- Cameron, J., Banko, K. M., & Pierce, W. D. (2001). Pervasive negative effects of rewards on intrinsic motivation: The myth continues. *The Behavior Analyst*, 24(1), 1–44. <https://doi.org/10.1007/BF03392017>.
- Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research*, 64(3), 363–423. <https://doi.org/10.3102/00346543064003363>.
- 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(6), 2156–2160. <https://doi.org/10.1016/j.chb.2013.05.009>.
- Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, 140(4), 980–1008.
- Cherrington, D. J., Reitz, H. J., & Scott, W. E. (1971). Effects of contingent and non-contingent reward on the relationship between satisfaction and task performance. *Journal of Applied Psychology*, 55(6), 531–536. <https://doi.org/10.1037/h0032041>.

- Cooper, J. (2007). *Cognitive dissonance: 50 years of a classic theory*. Sage.
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18(1), 105–115. <https://doi.org/10.1037/h0030644>.
- Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. R. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality*, 62(1), 119–142. <http://dx.doi.org/10.1111/j.1467-6494.1994.tb00797.x>.
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627–668.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum.
- Dholakia, U. (2015). *My experiences as an Amazon mechanical Turk (MTurk) worker*. Blogpost. Retrieved from <https://www.linkedin.com/pulse/my-experience-amazon-mechanical-turk-mturk-worker-utpal-dholakia>.
- Dickerson, C. A., Thibodeau, R., Aronson, E., & Miller, D. (1992). Using cognitive dissonance to encourage water conservation. *Journal of Applied Social Psychology*, 22(11), 841–854. <http://dx.doi.org/10.1111/j.1559-1816.1992.tb00928.x>.
- Eisenberger, R., & Cameron, J. (1996). Detrimental effects of reward: Reality or myth? *American Psychologist*, 51(11), 1153–1166.
- Eisenberger, R., Pierce, W. D., & Cameron, J. (1999). Effects of reward on intrinsic motivation—negative, neutral, and positive: Comment on Deci, Koestner, and ryan (1999). *Psychological Bulletin*, 125(6), 677–691. <https://doi.org/10.1037/0033-2909.125.6.677>.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16(1), 143–149.
- Feng, D., Besana, S., & Zajac, R. (2009 August). Acquiring high quality non-expert knowledge from on-demand workforce. *Proceedings of the 2009 workshop on the people's web meets NLP: Collaboratively constructed semantic resources* (pp. 51–56). Association for Computational Linguistics.
- Festinger, L. (1962). *A theory of cognitive dissonance*, Vol. 2. Stanford University Press.
- Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58(2), 203–210. <https://doi.org/10.1037/h0041593>.
- Harris, M. (2014 December 3). *Amazon's mechanical Turk workers protest: 'I am a human being not an algorithm'*. The Guardian. Retrieved from <https://www.theguardian.com/technology/2014/dec/03/amazon-mechanical-turk-workers-protest-jeff-bezos>.
- 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(1), 400–407.
- Hayes, A. F. (2012). *Process: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling [white paper]*. Retrieved from <http://www.afhayes.com/public/process2012.pdf>.
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Publications.
- Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behavior Research Methods*, 41(3), 924–936. <https://doi.org/10.3758/BRM.41.3.924>.
- Heyman, J., & Ariely, D. (2004). Effort for payment: A tale of two markets. *Psychological Science*, 15(11), 787–793. <https://doi.org/10.1111/j.0956-7976.2004.00757.x>.
- Hinojosa, A. S., Gardner, W. L., Walker, H. J., Coglisier, C., & Gullifor, D. (2017). A review of cognitive dissonance theory in management research: Opportunities for further development. *Journal of Management*, 43(1), 170–199. <https://doi.org/10.1177/0149206316668236>.
- Holden, C. J., Dennie, T., & Hicks, A. D. (2013). Assessing the reliability of the M5-120 on Amazon's mechanical turk. *Computers in Human Behavior*, 29(4), 1749–1754. <https://doi.org/10.1016/j.chb.2013.02.020>.
- Hübner, R., & Schlösser, J. (2010). Monetary reward increases attentional effort in the flanker task. *Psychonomic Bulletin & Review*, 17(6), 821–826. DOI: <https://doi.org/10.3758/PBR.17.6.821>.
- Lazear, E. P. (2000). Performance pay and productivity. *The American Economic Review*, 90(5), 1346–1361. <http://dx.doi.org/10.1257/aer.90.5.1346>.
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's mechanical turk. *Behavior Research Methods*, 44(1), 1–23. <https://doi.org/10.3758/s13428-011-0124-6>.
- Mason, W., & Watts, D. J. (2010). Financial incentives and the performance of crowds. *ACM SigKDD Explorations Newsletter*, 11(2), 100–108. <http://dx.doi.org/10.1145/1809400.1809422>.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17(11), 776–783. <https://doi.org/10.1037/h0043424>.
- Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding mechanical turk as a participant pool. *Current Directions in Psychological Science*, 23(3), 184–188. <https://doi.org/10.1177/0963721414531598>.
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon mechanical turk. *Judgment and Decision Making*, 5(5), 411–419.
- Pittman, M., & Sheehan, K. (2016). Amazon's Mechanical Turk a digital sweatshop? Transparency and accountability in crowdsourced online research. *Journal of Media Ethics*, 31(4), 260–262. <https://doi.org/10.1080/23736992.2016.1228811>.
- Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, 37(1), 7–63. <http://dx.doi.org/10.1257/jel.37.1.7>.
- Proaps, A. B., Landers, R. N., Reddock, C. M., Cavanaugh, K. J., & Kantowitz, T. M. (2014 April). Mobile and computer-based talent assessments: Implications of workload and usability. *CHI'14 extended abstracts on human factors in computing systems* (pp. 2299–2304). ACM. <http://dx.doi.org/10.1145/2559206.2581136>.
- Ross, J., Irani, L., Silberman, M., Zaldivar, A., & Tomlinson, B. (2010 April). Who are the crowdworkers?: shifting demographics in mechanical turk. *CHI'10 extended abstracts on human factors in computing systems* (pp. 2863–2872). ACM. <http://doi.acm.org/10.1145/1753846.1753873>.
- Rouse, S. V. (2015). A reliability analysis of Mechanical Turk data. *Computers in Human Behavior*, 43, 304–307. <https://doi.org/10.1016/j.chb.2014.11.004>.
- Ryan, R. M., Mims, V., & Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology*, 45(4), 736–750. <https://doi.org/10.1037/0022-3514.45.4.736>.
- Schneider, C., & Cheung, T. (2013). The power of the crowd: Performing usability testing using an on-demand workforce. *Information systems development* (pp. 551–560). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4614-4951-5_44.
- Scholz, T. (2012). *Digital labor: The Internet as playground and factory*. Routledge.
- Scott, J. (2000). Rational choice theory. In G. Browning, A. Halcli, & F. Webster (Eds.). *Understanding contemporary society: Theories of the present* (pp. 126–138). Thousand Oaks, CA: Sage.
- Shay, R., Komanduri, S., Durity, A. L., Huh, P. S., Mazurek, M. L., Segreti, S. M., & Cranor, L. F. (2014 April). Can long passwords be secure and usable? *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2927–2936). ACM. <http://doi.acm.org/10.1145/2556288.2557377>.
- Shay, R., Komanduri, S., Durity, A. L., Huh, P. S., Mazurek, M. L., Segreti, S. M., & Cranor, L. F. (2016). Designing password policies for strength and usability. *ACM Transactions on Information and System Security*, 18(4), article no. 13 <http://doi.acm.org/10.1145/2891411>.
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7(4), 422–445.
- United States Department of Labor (2018). *Minimum wage*. Retrieved from <https://www.dol.gov/general/topic/wages/minimumwage>.