Please participate in Part 2: Maximizing response rates in longitudinal MTurk designs

Elizabeth Stoycheff

Abstract
The ease and affordability of Amazon’s Mechanical Turk make it ripe for longitudinal, or panel, study designs in social science research. But the discipline has not yet investigated how incentives in this “online marketplace for work” may influence unit non-response over time. This study tests classic economic theory against social exchange theory and finds that despite Mechanical Turk’s transactional nature, expectations of reciprocity and social contracts are important determinants of participating in a study’s Part 2. Implications for future research are discussed.

Keywords
Response rate, longitudinal, panel, Mechanical Turk, surveys, experiments

A growing number of social psychological studies are reliant on Amazon’s Mechanical Turk (MTurk) worker pool to investigate human attitudes and behavior, returning tens of thousands of studies of Google Scholar search results. A wealth of cross-sectional scholarship investigating social science concepts, like framing (Berinsky et al., 2012), cultivation (Stravrositu, 2014), and credibility (Appelman and Sundar, 2016) have been tested using MTurk samples, many of which argue their work would benefit from longitudinal designs. And while MTurk is an apt platform for conducting panel studies, it remains unclear how token research incentives operate in this commercial, transactional marketplace. This study employs a methodological experiment to compare unit response rates across classic cost-benefit incentive strategies and social exchange incentive strategies and discusses how to best maximize participation in longitudinal designs.

The nature of MTurk samples
Like all volunteer convenience samples, the use of MTurk workers in social science has been greeted with mixed reviews. These participants may deviate from the general population in important ways, and studies comparing them to both representative adult samples and student participant pools have produced cautious results. Although Amazon boasts a pool of over 500,000 workers worldwide, academic laboratories estimate that populations of only about 7300 workers are available at any given time (Stewart et al., 2015). And many of these are super workers, who disproportionately participate in social science research studies (Rand et al., 2010). These subjects are susceptible to completing numerous surveys and experiments that share conceptually related goals (Chandler et al., 2014; Stewart et al., 2015), perhaps learning more about the social scientific process than many researchers would like (Hauser and Schwarz, 2016). Compared to in-person samples, they also have been shown to pay less attention to experimental stimuli and are significantly less extraverted (Clifford et al., 2015; Goodman et al., 2012; Rouse, 2015).

However, others have found MTurk workers to mirror student-sample limitations in terms of data quality (Sprouse, 2011) while providing demographic diversity more comparable to the US population at large (Berinsky et al., 2012). Many recent studies recommend MTurk workers for scholarship that seeks to understand relationships between traits,
attitudes, and behaviors based on results that show virtually indistinguishable differences across various sample types (Buhrmester et al., 2011; Casler et al., 2013; Christenson and Glick, 2013) and other indicators of high data quality, including comparable test-retest reliability (Holden et al., 2013; Huff and Tingley, 2015; Paolacci et al., 2010; Peer et al., 2014). The ease and affordability of MTurk samples have made them common practice for a variety of methodologies in social research, and studies that rely upon them are regularly published in leading journals across the social sciences (for a review, see Chandler and Shapiro, 2016).

**Maximizing longitudinal responses**

More than just a one-stop shop, the Mechanical Turk portal is also appropriately suited for online research that employs longitudinal, or panel, designs (Chandler et al., 2014; Christenson and Glick, 2013). And needing to re-contact a participant after days, weeks, months, or even years introduces MTurk to the same recruitment challenges as other modes of study administration. Although there are a myriad of factors that may contribute to unit response rate, including the study topic, presentation of online material, design of invitations, perceived credibility of the researcher, and use of study reminders (Fan and Yan, 2010; Groves et al., 2000), incentive strategies are likely to occupy a central role in MTurk’s transactional “market place for work” to re-attract subjects to participate in subsequent survey or experimental sessions. Two competing approaches—classic economic theory and social exchange theory—both provide insight into how incentives may help maximize unit response.

Classic economic theory contends that human behavior is largely predictable because individuals are rational in their decision-making (Homans, 1961). They weigh potential economic benefits against potential economic costs, and if there is a high likelihood of receiving adequate benefits for minimal costs, social action is expected (Scott, 2000). For example, a meta-analysis of 30 mail experiments found that greater incentives increased response rates in all but two studies (Fox et al., 1988). And an even larger meta-analysis spanning 80 years worth of web and mail surveys revealed a similar positive association (Auszpur and Schneck, 2014). MTurk, which operates on a cash payout system, provides added leverage for subjects who view more perceived benefit from cash than other forms of compensation (Ryu et al., 2006). Thus, classic economic theory would naturally predict that

**H1.** MTurk workers who are offered a larger incentive are more likely to participate in a study’s Part 2 than MTurk workers who are offered a smaller incentive.

But ethically, institutional review boards insist survey and experimental subjects should not be “paid” for their time, in order to reduce the likelihood of coercion and undue influence, but they are rather offered a token incentive for participation (Wrench et al., 2013). On its surface, MTurk’s preponderance of low-paying tasks suggests workers may have motivations beyond merely maximizing their financial benefits. An MTurker’s average compensation was only US$1.38 per hour in 2010 (Horton and Chilton, 2010), and ranked fourth on reasons why workers report participating, behind interest, enjoyment, and boredom (Buhrmester et al., 2011).

Social exchange theory lends a more nuanced explanation for why subjects may choose to continue their participation across multiple study sessions. Rather than operating on an economic cost-benefit calculation, social exchange theory argues that individuals also view social outcomes, like approval, trust, and reciprocity, as benefits to participating (Cropanzano and Mitchell, 2005; Scott, 2000; see also Boulianne, 2013; Brawley & Pury, 2016). Agreeing to participate in a longitudinal study establishes a social contract between the subject and the researcher, whereby both parties are expected to fulfill the expectations set forth in the first interaction. Other modes of study administration have used prepaid incentives to build a relationship of trust (e.g. Singer et al., 2000), but MTurk does not allow for prepayment. Instead, the gesture of goodwill routinely employed via this channel is the use of bonus payments. Researchers can grant bonuses to workers who perform outstanding or additional work. Bonuses for Part 2 of a study should signal to the participant that it’s a continuation of an established social contract, as opposed to a new interaction. As such, social exchange theory would predict that

**H2.** MTurk workers who are offered a bonus incentive are more likely to participate in a study’s Part 2 than MTurk workers who are offered a new task for Part 2.

**Methods**

To determine the most effective incentive strategy, participants were recruited via MTurk in early 2016 for a two-part study. The initial human intelligence task (HIT) instructed interested participants that they would receive US$0.50 compensation to complete a 5-minute survey about their online behaviors and would be re-contacted in 1 week to complete a second 5-minute survey on the same subject for additional compensation. The initial survey included a series of questions that tapped participants’ attitudes about current US events, their likelihood to engage in a series of online behaviors the following week (e.g. post on social media, visit news websites, and bank online), their personality traits, and demographic information. Participants were required to meet Amazon’s qualification of residing in the United States, have a task approval rating of greater than or equal to 95%, and the final sample was demographically similar to the US population at large; 55% of the sample was female, with an average age of 37 (range=18–74, standard deviation (SD)=12.45),
and educational level of some college, but no degree (range = 1–9, mean (M) = 6.10, SD = 1.33). Sampled participants also had an average 2015 household income of slightly over US$40,000 (range = 1, less than US$10,000, to 9, over $150,000, M = 4.73, SD = 2.24), and 87% identified as White or Caucasian (6% were Black, 5.4% were Asian-American or Pacific Islander, and 1.5% identified as other). A total of 331 participants fully completed Part 1 of the study such that they were eligible for future correspondence.

A week later, participants who successfully completed Part 1 were randomly assigned to one of three experimental conditions and re-contacted advertising either: A new survey HIT for Part 2 for US$0.50 compensation (N = 109), a new survey HIT for Part 2 that offered US$1.00 compensation (N = 111), or a continuation of the Part 1 survey HIT for a US$0.50 bonus compensation (N = 109), a new HIT for Part 2 for US$0.50 compensation (N = 111), or a continuation of the Part 1 survey HIT for a US$1.00 bonus compensation (N = 111). Participants across all conditions were given 24 hours to complete the second survey and were sent a reminder email if they had not completed it within this time. The Part 2 survey terminated data collection after an additional 48 hours.

Results

After 72 hours of data collection, 248 of the original 331 participants had successfully completed Part 2 of the survey, amounting to an impressive 74.9% response rate, which is comparable to retention rates of other carefully conducted studies (Christenson and Glick, 2013). Of these 331 subjects, 219 participated after initial contact and 29 participated after a second follow-up email. A series of chi-square analyses was conducted on each pairing of conditions using the original N = 331 sample to determine differences in response rate between the incentive strategies, both after the initial contact for Part 2 participation and the follow-up reminder email issued 24 hours later. Results show that there was no significant difference between the US$0.50 and US$1.00 payment incentives at either initial contact $\chi^2(1, N = 220) = 0.03, p = 0.87$ or follow-up $\chi^2(1, N = 220) = 0.50, p = 0.48$, indicating that at this level of financial incentive, a 100% increase in payment did not entice greater participation, failing to support classic economic theory proposed in H1.

However, individuals who initially offered a US$1.00 bonus were significantly more likely to participate in Part 2 than those who were offered either a US$0.50 $\chi^2(1, N = 220) = 5.88, p = 0.015$ or a US$1.00 new task compensation $\chi^2(1, N = 222) = 6.76, p = 0.009$, lending support for the social exchange approach proposed in H2. But the significance of this effect wore off for those respondents who needed a second reminder email. Once those who needed a follow-up were included in the total response rate, there was no significant difference in response rate between individuals offered bonus compensation and those offered a new task incentive, either at US$0.50 $\chi^2(1, N = 220) = 2.96, p = 0.09$ or US$1.00 increments $\chi^2(1, N = 222) = 1.08, p = 0.23$. Response rates for all payment conditions are summarized in Table 1.

Discussion

Despite MTurk’s reputation as a commercial enterprise, designed to facilitate economic or market transactions between parties, higher financial incentives alone did not significantly increase response rates. These results mirror previous research that has explored how incentives influence other modes of survey response, like face-to-face, email, and telephone (Auspurg and Schneck, 2014; Boulianne, 2013; Grauenhorst et al., 2016; Singer et al., 2000), and extend to a new and increasingly popular means of data collection. Instead, this research suggests that granting bonus payments—which represent gestures of goodwill between participants and researchers—is more effective at increasing unit response, even when financial incentives are equal. This finding is particularly interesting given that the completion of a new task, as opposed to a bonus, adds to each worker’s success rate, which enhances future payment opportunities. If individuals were operating on a mere cost-benefit calculation alone, it would behoove them to participate in a new task rather than a task that is paid as a bonus. However, workers do face the very real possibility that requestors may withhold compensation, even for attentive participation, so bonuses may signal work security between the two parties, suggesting, from a utilitarian perspective, that good work is likely to be rewarded. The finding that the social exchange bonus incentives diminished after reminder emails may indicate that participants perceive the extra cost a researcher incurred to remind them to participate, compelling them to reciprocate, regardless of the compensation amount.

As is customary in longitudinal designs, this study notified all participants at the onset that there would be subsequent study waves, which may have contributed to an elevated expectation of reciprocity across all conditions. Research that fails to inform subjects of future interactions may experience greater attrition because an initial social

| HIT Type                  | N completed Part 1 | N completed Part 2 after 24 hours | N completed Part 2 after 72 hours |
|---------------------------|--------------------|-----------------------------------|----------------------------------|
| New HIT US$0.50 compensation | 109                | 67                                | 80                               |
| New HIT US$1.00 compensation | 111                | 67                                | 77                               |
| Bonus HIT US$1.00 compensation | 111                | 85                                | 91                               |

HIT: human intelligence task.
contract has not been established. This research also only investigated two waves of data collection, and future research is needed to determine whether bonus payments operate similarly for Part 3 and beyond; however, one would expect an even greater level of trust and reciprocity after further interactions, suggesting bonus payments may be even more beneficial for subsequent invitation requests. Self-report measures that capture workers’ feelings of trust and reciprocity toward MTurk requestors would further shed light on the consciousness of this process. The US$0.50 and US$1.00 incentives offered in this study are small, but comparable to those of social scientific studies regularly recruited on MTurk’s platform, and are reasonable given the modest, 5-minute time commitment required of participants. Although higher incentives do not appear to increase unit response, and previous research has shown they do not significantly affect data quality (Buhrmester et al., 2011) or the population willing to participate (Stewart et al., 2015), the scientific community should be ethically mindful of not economically exploiting MTurk workers (Fort et al., 2011). Rather, researchers should strive to provide compensation to online workers, which is a fair token of the time required, while also considering non-financial incentives that build mutually beneficial relationships with workers via learning experiences and social benefits in order to maximize high-quality responses and a positive worker experience.

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) received no financial support for the research, authorship, and/or publication of this article.

References
Appelman A and Sundar SS (2016) Measuring message credibility: Construction and validation of an exclusive scale. Journalism & Mass Communication Quarterly 93: 59–79.
Auspurg K and Schneck A (2014) What difference makes a difference? A meta-regression approach on the effectiveness conditions of incentives in self-administered surveys. In: Proceedings of the MAER-Net 2014 Athens colloquium (MAER-Net), Athens. Available at: http://metaanalysis2014.econ.uoa.gr/fileadmin/metaanalysis2014.econ.uoa.gr/uploads/Schneck_Andreas.pdf
Berinsky AL, Huber GA and Lenz GS (2012) Evaluating online labor markets for experimental researcher: Amazon.com’s Mechanical Turk. Political Analysis 20: 351–368.
Boulianne S (2013) Examining the gender effects of different incentive amounts in a web survey. Field Methods 25(1): 91–104.
Brawley AM and Pury CLS (2016) Work experiences on MTurk: Job satisfaction, turnover, and information sharing. Computers in Human Behavior 54: 531–546.
Buhrmester M, Kwang T and Gosling SD (2011) Amazon’s Mechanical Turk: A new source of inexpensive, yet high-quality data? Perspectives on Psychological Science 6(1): 3–5.
Casler K, Bickel L and 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 J and Shapiro D (2016) Conducting clinical research using crowdsourced convenience samples. Annual Review of Clinical Psychology 12: 53–81.
Chandler J, Mueller P and Paolacci G (2014) Nonnaivete among Amazon mechanical Turk workers: Consequences and solutions for behavioral researchers. Behavior Research Methods 46(1): 112–130.
Christenson DP and Glick DM (2013) Crowdsourcing panel studies and real-time experiments in MTurk. The Political Methodologist 20(2): 27–32.
Clifford S, Jewell RM and Waggoner PD (2015) Are samples drawn from Mechanical Turk valid for research on political ideology? Research & Politics. Epub ahead of print December. DOI: 10.1177/2053168015622072.
Cohen EL and Lancaster AL (2014) Individual differences in in-person and social media television coviewing: The role of emotional contagion, need to belong, and coviewing orientation. Cyberpsychology, Behavior, and Social Networking 17(8): 512–518.
Cropanzano R and Mitchell MS (2005) Social exchange theory: An interdisciplinary review. Journal of Management 31: 874–900.
Fan W and Yan Z (2010) Factors affecting response rates of the web survey: A systematic review. Computers in Human Behavior 26: 132–139.
Fort K, Adda G and Cohen KB (2011) Amazon mechanical Turk: Gold mine or coal mine? Computational Linguistics 37: 413–420.
Fox RJ, Crask MR and Kim J (1988) Mail survey response rate: A meta-analysis of selected techniques for inducing response. Public Opinion Quarterly 52(4): 467–491.
Goodman JK, Cryder CE and Cheema A (2012) Data collection in a flat world: The strengths and weaknesses of mechanical Turk samples. Journal of Behavioral Decision Making 26(3): 213–224.
Grauenhorst T, Blohm M and Koch A (2016) Respondent incentives in a national face-to-face survey: Do they affect response quality? Field Methods 28: 266–283.
Grovés RM, Singer E and Corning A (2000) Leverage-saliency theory of survey participation: Description and an illustration. Public Opinion Quarterly 64(3): 299–308.
Hauser DJ and Schwarz N (2016) Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. Behavior Research Methods 48: 400–407.
Holden CJ, Dennie T and Hicks AD (2013) Assessing the reliability of the M5-120 on Amazon’s mechanical Turk. Computers in Human Behavior 29(4): 1749–1754.
Homans G (1961) Social Behavior: Its Elementary Forms. London: Routledge.
Horton JJ and Chilton LB (2010). The labor economics of paid crowdsourcing. In: Proceedings of the 11th ACM conference on electronic commerce, Cambridge, MA, 7–11 June, pp. 209–218. New York: ACM.
Huff C and Tingley D (2015) “Who are these people?” Evaluating the demographic characteristics and political preferences of MTurk survey respondents. Research & Politics. Epub ahead of print September. DOI: 10.1177/2053168015604648.

Paolacci G, Chandler G and Ipeirotis PG (2010) Running experiments on Amazon mechanical Turk. *Judgment and Decision Making* 5(5): 411–418.

Peer E, Vosgerau J and Acquisti A (2014) Reputation as a sufficient condition for data quality. *Behavior Research Methods* 46(4): 1023–1031.

Rand DG, Peysakhovich A, Kraft-Todd GT, et al. (2014) Social heuristics shape intuitive cooperation. *Nature Communication* 5: 3677.

Rouse SV (2015) A reliability analysis of mechanical Turk data. *Computers in Human Behavior* 43: 304–307.

Ryu E, Couper MP and Marans RW (2006) Survey incentives: Cash vs. in-kind; face-to-face vs. mail; response rate vs. nonresponse error. *International Journal of Public Opinion Research* 18: 89–106.

Scott J (2000) Rational choice theory. In: Browning G, Halcli A and Webster F (eds) *Understanding Contemporary Society: Theories of the Present*. Thousand Oaks, CA: SAGE, pp. 126–138.

Singer E, Van Hoewyk J and Maher MP (2000) Experiments with incentives in telephone surveys. *Public Opinion Quarterly* 64(2): 171–188.

Sprouse J (2011) A validation of Amazon mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* 43(1): 155–167.

Stewart N, Ungemach C, Harris AJ, et al. (2015) The average laboratory samples a population of 7,300 Amazon mechanical Turk workers. *Judgment and Decision Making* 10(5): 479–491.

Stravrositu CD (2014) Does TV viewing cultivate meritocratic beliefs? Implications for life satisfaction. *Mass Communication and Society* 17(1): 148–171.

Wrench JS, Thomas-Maddox C, Peck Richmond V, et al. (2013) *Quantitative Research Methods* (2nd edn). New York: Oxford University Press.

**Author biography**

Elizabeth Stoycheff is an Assistant Professor in the Department of Communication at Wayne State University. Her research uses online sampling techniques to investigate the effects of new media on democratic development and opinion formation.