Measuring delay discounting in a crowdsourced sample: An exploratory study

Amanda Rotella*

Department of Psychology, University of Guelph

Cody Fogg*

Department of Psychology, University of Guelph

Sandeep Mishra

Hill/Levene Schools of Business, University of Regina

Pat Barclay

Department of Psychology, University of Guelph

These authors contributed equally to this work

Measuring delay discounting in a crowdsourced sample: An exploratory study

ABSTRACT

Delay discounting is a measure of preferences for smaller immediate rewards over larger delayed rewards. Discounting has been assessed in many ways; these methods have variably and inconsistently involved measures of different lengths (single vs. multiple items), forced-choice methods, self-report methods, online and laboratory assessments, monetary and non-monetary compensation. The majority of these studies have been conducted in laboratory settings. However, over the past 20 years, behavioral data collection has increasingly shifted online. Usually, these experiments involve completing short tasks for small amounts of money, and are thus qualitatively different than experiments in the lab, which are typically more involved and in a strongly controlled environment. The present study aimed to determine how to best measure future discounting in a crowdsourced sample using three discounting measures (a single shot measure, the 27-item Kirby Monetary Choice Questionnaire, and a one-time Matching Task). We examined associations of these measures with theoretically related variables, and assessed influence of payment on responding. Results indicated that correlations between the discounting tasks and conceptually related measures were smaller than in prior laboratory experiments. Moreover, our results suggest providing monetary compensation may attenuate correlations between discounting measures and related variables. These findings suggest that incentivizing discounting measures changes the nature of measurement in these tasks.

INTRODUCTION

Decision-makers must consistently engage with trade-offs between certain, immediately available outcomes and the possibility of distal, but superior outcomes in the uncertain future. Depending on the parameters of the choice, this trade-off may result in delay discounting — where a smaller immediate reward is preferred over larger distal rewards (reviewed in Frederick, Loewenstein, & O'Donoghue, 2002). However, individual differences in delay discounting have been widely documented, and these individual differences have been associated with a wide array of meaningful and relevant outcomes, ranging from drug use to criminal behavior (Green & Myerson, 2004). Despite the apparent importance of delay discounting, there is some ambiguity as to the best practices for measuring discounting. Given the growing ubiquity of the use of online crowdsourcing platforms for psychological research, we explored the predictive validity of three simple-to-administer tools with two different compensation methods using a quasi-experimental design on Amazon Mechanical Turk, a widely used crowdsourcing platform.

Most simple-to-administer measures of discounting involve a relatively small number of monetary choices. For example, the widely-used Monetary-Choice Questionnaire (MCQ; Kirby, Petry, & Bickel, 1999) presents participants with a series of 27 choices between relatively smaller monetary amounts available now, and relatively larger monetary amounts available later (e.g., "Would you rather have $50 tonight, or $80 in 70 days?"). Kirby's MCQ specifically presents participants with a series of small, medium, and large choices, and from each series of choices, a "switch point" can be identified where participants' preferences shift from present rewards to future rewards. This switch point can be used to calculate a discounting parameter (k), which quantifies one's position on a hyperbola of time preference (ranging from persistent immediate-focus to persistent future-focus). One-shot discounting measures present participants

with a single choice between amounts now and later (e.g., a choice between £45 in three days or £70 in three months; Reimers, Maylor, Stewart, & Chater, 2009). Matching methods require participants to indicate the minimal amount of money they would be willing to receive to wait an additional period for compensation (e.g., a choice of $10 now, vs. $10 + X in 30 days, where participants report what minimum X they would be willing to accept; Hardisty, Thompson, Krantz, & Weber, 2013).

Regardless of the specific measure used, delay discounting instruments have been robustly associated with several theoretically consistent, impulsive real-world behaviors and outcomes, including criminal behavior, antisocial conduct, gambling, obesity, promiscuity, cigarette use, alcohol use, and drug use, among many others (e.g., Alessi & Petry, 2003; Hanoch, Rolison, & Gummerum, 2013; Kirby et al., 1999; Kirby & Petry, 2004; Mishra & Lalumière, 2017; Petry, 2001; Reimers et al., 2009; Reynolds, 2006). Behavioral discounting has also been associated with trait measures of impulsivity and self-control, consistent with suggestions that delay discounting itself represents a stable trait (Mishra & Lalumière, 2017; Odum, 2011). Evidence suggests that choice methods (as opposed to matching methods) appear to be superior at predicting real-world outcomes (Hardisty et al., 2013), although this work did not tie participant compensation to decisions made within discounting measures. It remains an open question of what simple measure is best to use on crowdsourcing platforms, and what compensation method to use.

Extant research suggests that different payment structures have meaningful influences on participants' behavior in decision-making tasks. Generally, both generous pay and performance-based payment have been associated with superior task performance and higher task completion rates in experiments (Brase, 2009; Camerer & Hogarth, 1999). Some investigations also suggest more nuanced effects of compensation on decision-making (Camerer & Hogarth, 1999; Ferrey & Mishra, 2014). For example, Ferrey and Mishra (2014) showed that in the widely-used Balloon Analogue Risk Task, participants who received session-based compensation (i.e., a consistent lump sum payment, regardless of task decisions) engaged in significantly greater risk-taking compared to those who were paid based on their actual decisions, and those who were not paid at all. For delay discounting measures specifically, evidence suggests discounting does not systematically differ when assessed using hypothetical rewards versus real rewards (Johnson & Bickel, 2002; Lagorio & Madden, 2005; Madden, Begotka, Raiff, & Kastern, 2003; Madden et al., 2004).

In the present study, we examined the "real-world" predictive ability of three widely used delay discounting tasks—*k*-parameter elicitation from the monetary choice questionnaire, single-shot discounting, and matching task, and examined whether different compensation methods affected predictive ability. Although this was an exploratory study, given that prior evidence suggested that discounting does not differ according to compensation, we made a weak prediction that there would be no difference in the strength of correlations between the payment and hypothetical conditions. We specifically examined these associations among a crowdsourced Amazon Mechanical Turk sample, given the growing ubiquity of such platforms for psychological research. We pre-registered our measures, analyses, and power calculations and sample size determination at https://osf.io/r7jfd/?view_only=4866441b3918496aa11868407ba889de. The dataset and analysis code is available at: https://osf.io/6x38u/?view_only=b66ce266f9c3499c87d4154e72b81914. .

METHOD

248 participants from the United States completed a 10-minute survey for $1.20 USD on Amazon Mechanical Turk (MTurk). We excluded participants who answered the survey in less than two minutes or made inconsistent choices on a questionnaire (N = 27), resulting in a final sample of 221 participants (M_{age} = 36.22; 37% female).

Delay Discounting

Monetary-choice questionnaire (MCQ)

Participants completed the 27-item Monetary-Choice Questionnaire (MCQ; Kirby et al., 1999). They viewed three blocks of 9 items, where they chose between an amount available today and an amount available at a future time (e.g. "Would you prefer $0.11 today or $0.30 in 7 days?"). Each choice was associated with a discounting parameter, which is a value of how much the future is discounted if the immediate reward is chosen. The blocks varied based on the amount of money offered (small, medium, and large), and were presented in that order (see supplementary material). Within each of the three blocks, the discounting parameter was averaged for the two values where participants 'switched' from preferring the immediate reward to the delayed reward, which was done using R code developed by Gray and colleagues (2016). Once the discounting parameter was calculated for each of the three blocks, we took the mean to obtain the total MCQ score. Because MTurk workers are used to completing short tasks for small sums of money, we adapted Kirby and colleagues' (1999) scale to an online format by dividing the dollar amount offered by 100. No changes were made to the time delay or the discounting parameters of the items. Cronbach's alpha indicated that this scale had good internal consistency, α = .88, 95%CI[.87, .89].

Single-shot discounting (SSD)

We adapted a single-shot discounting measure (Mishra & Lalumière, 2011; 2017; Reimers et al., 2009) to an online format after piloting various payment structures (see Supplementary Table S1). Participants were presented with the following question: "Would you prefer: $0.15 today or $0.23 in 21 days?" Those who chose the immediate option (i.e. those who discounted the future) were coded as 1 and those who chose the delayed option were coded as -1, such that positive scores indicated greater discounting.

Matching Task

Using Hardisty and colleagues' (2013) matching methods as a model, we adapted a single-item measure to the present study by asking participants, "If you were choosing between $10 now versus more money in one month (30 days), what is the least amount of money it would take to get you to wait 30 days for that money?" Participants could respond with values between $10 and $100.

Conceptually Related Outcome Measures

Eysenck Impulsivity Scale (EIS)

The EIS (Eysenck, Pearson, Easting, & Allsopp, 1985) involves 19 yes/no questions about impulsive behaviors (e.g. "Do you often buy things on impulse?"). The number of 'yes' responses were averaged. These items were reliable, α = .84, 95%CI[.79, .88].

Barratt Impulsiveness Scale Short-Form (BIS)

The BIS (Spinella, 2007) contains 15 items that assess the frequency of impulsive behaviors (e.g. "I say things without thinking") on a four-point scale (1 = *rarely/never*, 4 = *almost always/always*). Scale items were averaged to obtain a total score. Cronbach's alpha indicated this scale had good reliability, $\alpha = .84$, 95%CI[.81, .86].

Personal Relative Deprivation Scale (PRDS)

The PRDS (Callan, Shead, & Olson, 2011) is a 5-item scale that assesses subjective feelings of relative deprivation. Participants responded to items such as "I feel deprived when I think about what I have compared to what other people like me have" on a 6-point scale (1 = *strongly disagree*, 6 = *strongly agree*). The final PRDS score was obtained by averaging the five items. This scale had very good reliability, $\alpha = .87$, 95%CI[.83, .90].

The Problem Gambling Severity Index (PGSI)

The PGSI (Brooker, Clara, & Cox, 2009) is a 9-item measure that assesses the frequency of behaviors associated with problem and pathological gambling. Items such as "How often have you felt that you might have a problem with gambling?" were rated on a 4-point scale (1 = *never*, 4 = *almost always*). A total score was obtained by averaging the items. Cronbach's alpha demonstrated that this scale had excellent internal consistency, $\alpha = .91$, 95%CI[.88, .94].

Procrastination

Participants rated how much they agree with the statement "I procrastinate on most tasks" on a 7-point Likert scale, ranging from (1) *strongly disagree* to (7) *strongly agree*.

Alcohol Intake Frequency

Participants responded to "How often do you drink alcohol?" on a 7-point Likert scale, ranging from (1) *never* to (7) *several times a day*.

Other Measures

We asked participants if they smoked cigarettes and if they have ever been arrested (yes/no). Additionally, we presented participants with items that we predicted to be uncorrelated (rs = .00, see pre-registration) with our discounting measures (e.g., "I have a favorite pair of pants"). These analyses are presented in supplementary material.

PROCEDURE

Participants first completed the three discounting measures (as described above), and were then presented with the EIS, BIS, PRDS, and PGSI in randomized order. We varied the presentation order of the discounting measures so that participants either received the MCQ or SSD first, while the Matching Task was always presented between these two other measures (the order effect analyses are presented in Supplementary Material). The Matching Task was used for exploratory purposes, so we did not manipulate payment or order effects. All other measures were presented at the end of the survey along with demographic variables. We chose the outcome measures because delay discounting has been previously related to impulsivity (Alessi & Petry, 2003; Mishra & Lalumière, 2011; 2017), relative deprivation (Callan et al., 2011; Mishra & Novakowski, 2016), problem gambling (Alessi & Petry, 2003; Mishra & Lalumière,

2017; reviewed in Reynolds, 2006), procrastination (Schouwenburg & Groenewoud, 2001), and alcoholism (Petry, 2001; reviewed in Reynolds, 2006).

Payment incentives were also manipulated. All participants received a baseline payment for participation ($1.20 USD). Participants in the payment condition also received the outcome they chose on one randomly selected item of the MCQ and for their choice on the SSD. All participants in the payment condition received remuneration for the same item, which was chosen in advance to data collection by a random number generator; participants did not know which item was chosen. Participants who chose the immediate reward were compensated on the same day they completed the study and those who chose the delayed option received payment at the specified time delay. Participants in the hypothetical condition did not receive payment, but were asked to make decisions as if they involved real money.

ANALYSIS

Although our pre-registration specified that Pearson correlations would be used, all discounting measures violated either the assumptions of skewness or kurtosis (values > 1; see Supplementary Table S2). Thus, we analyzed the data using Spearman's *rho* correlations with 95% confidence intervals which we bootstrapped with 1,000 repetitions using the *RVAideMemoire* package (Hervé, 2018). Significance levels were obtained using the *Hmisc* (Harrell, 2018) package for R (R Core Team, 2018), and the scale Cronbach's alpha levels were obtained using the *ltm* package (Rizopoulos, 2006), with 1,000 bootstraps for the confidence intervals. Correlation magnitudes were visually compared across conditions for interpretation, and Fisher's *r*-to-*z* tests were computed to compare correlations in the payment and hypothetical conditions. Corrections for multiple comparisons were not applied because these were exploratory analyses.

To compare responses between conditions, we computed chi-square analyses using the *psych* package (Revelle, 2018) for the single-shot discounting measure, and Mann Whitney U tests on the MCQ discounting functions and Matching Task responses. These analyses were not pre-registered, but will facilitate comparison to previous studies investigating how incentivizing delay discounting measures influences its measurement.

RESULTS

Checks and Demographics

Discounting measures were not related to gender (r_ss = -.04 to .02, ps > .54) or age (r_ss = -.13 to -.07, ps > .06). We correlated the three discounting measures (Monetary Choice [MCQ], Single-Shot [SSD], and Matching Task) with the relevant continuous variables: Eysenck Impulsivity Scale (EIS), Barratt Impulsiveness Scale (BIS), Problem Gambling Severity Index (PGSI), Personal Relative Deprivation Scale (PRDS), procrastination, and alcohol intake frequency. Overall, the three discounting measures were all highly correlated with one another (r_ss = .59 to .73, ps < .001) and all outcome variables were related with each other (r_ss = .16 to .67, ps < .05), with the exception of most correlations with the alcohol frequency outcome variable (r_ss = -.002 to .075, ps > .10).

Discounting Measurement and Relationships with Theoretically Related Measures

Small to moderate correlations were found between all three discounting measures and the BIS (r_ss = .13 to .18, ps < .05). However, only the Matching Task was associated with scores on the EIS (r_s = .20, p = .002). Problem gambling tendencies as measured using the PGSI were related to scores on the MCQ and SSD (r_ss = .13 to .20, ps < .06), but not the Matching Task (r_s = .063, p = .36). None of the discounting measures were associated with personal relative deprivation (r_ss = -.048 to -.023, ps > .47) or procrastination (r_ss = .077 to .11, ps > .09). Only the Matching Task was trending towards significance with alcohol intake frequency (r_s = .13, p = .059) (see Table 1 & Supplementary Figure S1). These results suggest that all three measures have some predictive validity for theoretically related variables. Somewhat unexpectedly, however, not a single method produced correlations with all conceptually related measures, and these relationships were smaller than effects found in previous lab experiments. Moreover, these results suggest that different discounting tasks may produce stronger relationships with different outcome measures. For example, only the forced-choice measures (i.e. MCQ and SSD) were related to problem gambling, while the matching task was the only measure related to both impulsivity scales.

Discounting Measurement and Payment Incentives

To determine if payment incentives influenced correlations between discounting and theoretically relevant outcome variables, we compared the payment and hypothetical conditions, and then correlated the discounting measures with the continuous outcome measures (see Table 2B and 2C and Supplementary Figure S2). In the hypothetical condition (Table 2B) all delay discounting measures were correlated with BIS scores (r_ss = .19 to .27, ps < .05). Additionally, the MCQ was correlated with scores on the PGSI (r_s = .26, p = .005), and the Matching Task was correlated with scores on the EIS (r_s = .31, p = .001) and alcohol intake frequency (r_s = -.27, p = .005). There were no significant correlations between any of the discounting and outcome measures in the payment condition (all r_ss = -.007 to .17, ps > .076). Fisher's r-to-z tests (two-tailed) indicated that correlations were larger in the hypothetical condition compared to the payment condition for the Matching Task and alcohol intake frequency (z = 2.15, p = .032) and marginally for the Matching Task and EIS (z = 1.93, p = .054). No other effects were statistically significant, likely due to the limited sample size of the present study. Payment by order effect results are presented in Supplementary Table S3.

To address concerns about multiple comparisons, we combined all three delay discounting tasks into a single overall discounting measure by averaging participants' standardized scores across the three tasks (α = .74, 95%CI[.68, .80]). We then correlated the overall discounting measure with theoretically related variables in the payment and hypothetical conditions (see Table 2A). No correlations reached statistical significance in the payment condition (.008 < r_ss < .161, ps > .091). However, three of the six correlations reached statistical significance in the hypothetical condition (i.e., EIS, BIS, and PGSI; .214 < r_ss < .249), although Fisher's r-to-z tests did not reach statistical significance (.126 < ps < .374), likely because this study was underpowered to detect small differences in correlations (i.e., r_s differences between .12 and .17). Correlations with personal relative deprivation or with procrastination were not statistically significant, r_s = -.106, p = .269, and r_s = .007, p = .486, respectively.

Discounting Measurement Responses by Condition

For the single-shot discounting (SSD) measure, the number of participants who chose immediate or delayed rewards did not differ by payment condition, $\chi^2(1) = .063$, $p = .80$. Similarly, Mann-Whitney U-tests indicated that means of distributions did not differ according to payment for the MCQ ($W = 5692.5$, $p = .450$) or Matching Task ($W = 6516.5$, $p = .319$). These results suggest that hypothetical and real payment conditions have similar distribution shapes for all three discounting measures. Moreover, the means of the distributions did not differ according to payment for the EIS ($W = 6079.5$, $p = .96$), BIS ($W = 6093$, $p = .98$), or PGSI ($W = 5692.5$, $p = .23$), which suggests that differences in correlation magnitude as a consequence of payment condition are not due to different patterns of responding in the dependent measures.

DISCUSSION

Preliminary results suggest that payment may reduce the strength of association between delay discounting measures and theoretically-related individual differences and outcome measures (e.g., impulsivity, gambling, and alcohol intake), with the strongest correlations being observed in the hypothetical payment condition. Despite these differences, the distribution of responses for these discounting measures did not vary with payment, which is consistent with previous studies (Dixon, Lik, Green, & Myerson, 2013; Johnson & Bickel, 2002; Lagorio & Madden, 2005; Madden et al., 2003; 2004). Furthermore, the distributions of responses for our dependent measures (i.e. EIS, BIS, and PGSI) did not vary with payment in the largest sample investigating this question to-date. Together, these findings suggest that financially incentivizing delay discounting tasks does not affect the distribution of responses on the task, but provides preliminary evidence that payment may influence how discounting is associated with other instantiations of impulsivity. We do note, however, that the differences in effects were small ($.11 < r_s s < .21$). Because the obtained correlations were smaller than expected, these planned analyses were underpowered, and thus results should be considered tentative. More research is needed to confirm results.

Previous research has shown mixed influence of payment incentives on decisions in delay discounting behavioral tasks. Some studies found that payment (compared to unpaid controls) reduces the amount of money transferred in economic games (Amir, Rand, & Gal, 2012; Bühren & Kundt, 2015), but not always (Amir et al., 2012). Payment has also been found to reduce risk-taking (see Irwin, McClelland, & Schulze, 1992). As an alternative explanation, larger effects in hypothetical conditions may be due to participants overstating their preferences (List & Gallet, 2001), or due to "presentation effects" (Camerer & Hogarth, 1999). However, these interpretations may not account for the results of the present study. Rather, the present examination suggests that incentives eliminate or reduce associations between discounting and other theoretically-relevant instantiations of impulsivity.

Contrary to previous findings, delay discounting was not associated with personal relative deprivation (e.g., Callan et al., 2011; Tabri, Shead, & Wohl, 2017), smoking cigarettes (e.g., Bickel, Odum, & Madden, 1999; Reimers et al., 2009), or procrastination (Schouwenburg & Groenewoud, 2001). Additionally, delay discounting was only sometimes related to arrests (e.g., Mishra & Lalumière, 2017) and frequency of drinking alcohol. However, these differences may be due to the measurement differences in the present study; that is, we assessed frequency of alcohol intake instead of alcoholism (e.g., Petry, 2001). Alternatively, discounting responses may be characterized by high levels of random error, and therefore findings may be inconsistent.

The present study had several limitations that provide directions for future research. First, our sample was a convenience sample crowdsourced on Amazon Mechanical Turk, which is a crowdsourcing website commonly used by researchers across disciplines to collect large amounts of quality data for relatively small costs (Amir et al., 2012; Bohannon, 2016; Burhmester, Kwang, & Gosling, 2011; Chandler & Shapiro, 2016; Goodman & Paolacci, 2017; Robinson, Rosenzweig, Moss, & Litman, 2019). Although our goal was to determine how to measure delay discounting in this population, results using such populations may be generalizable to other online users who do tasks for relatively small amounts of money, we cannot generalize to other populations who are accustomed to doing tasks or making decisions involving larger amounts of money. We will also note that there is a possibility that the results may be an artifact of the very low reward amounts resulting in very steep discount functions, relative to other studies with fewer subjects and larger amounts. Thus, these results are generalizable to similar contexts and methods, but not when larger rewards are used. The discounting function of large amounts may differ in some ways from that of small amounts, and perhaps change the relationships found in the present study. As such, future studies are needed to determine how the size of the payment reward influences delay discounting functions, and if the size of the reward changes the relationship between discounting and other theoretically-related measures in paid and unpaid contexts. Moreover, we cannot generalize these results to in-lab studies with special populations. Future research is needed to test the effect of incentivization on delay discounting for larger monetary amounts.

Second, despite having a much larger sample size than previous studies examining payment effects on measuring time preferences (e.g., Dixon et al., 2013; Johnson & Bickel, 2002; Lagorio & Madden, 2005; Madden et al., 2003; Madden et al., 2004), we did not have a large enough sample in each condition to test for statistically significant differences in effect sizes (i.e. would need approximately 5,000 participants). In fact, the correlations we obtained were smaller than what was reported in previous in-lab studies (e.g., Mishra & Lalumière, 2017), which were used to project required sample size for the present study. Still, all statistically significant effect sizes were found in the hypothetical payment condition, while none were found in the payment condition, indicating that there may be a broader, consistent pattern of results. Because most correlations were smaller in this study than in previous in-lab delay discounting studies, we recommend that when using delay discounting tasks online researchers use conservative effect size estimates to estimate the required number of participants in their studies.

The associations were inconsistent between delay discounting and other relevant variables, in that some but not all measures of discounting were associated with the other variables; though these findings may be due to some of our chosen measures being less psychometrically established and subject to high error variances (e.g., 1-item procrastination question, and a 1-item alcohol intake frequency question). Moreover, although personal relative deprivation is conceptually related to delay discounting, it is related indirectly (e.g., feelings of relative deprivation lead to more impulsive and discounted choices because deprived individuals feel less competitive in their respective environment). Notably, however, we found correlations between delay discounting and all other conceptually-related and well-established scales (i.e., Eysenck Impulsivity Scale, Barratt Impulsiveness Scale, and Problem Gambling Severity Index); although not with all measures of discounting. For example, the Matching Task was the only measure associated with Eysenck's impulsivity scale, but also the only measure unrelated to problem gambling. This pattern of findings suggests that various discounting measures may relate differently to conceptually related variables. In particular, the results provide preliminary

suggestions that forced-choice tasks (i.e., MCQ and SSD) may be better predictors of gambling, but the Matching task may be a better predictor of impulsivity.

Third, we did not manipulate conditions for the Matching Task, but did so for the MCQ and single-shot discounting. Interestingly, the correlations with the Matching Task were smaller in the paid condition, despite never having been incentivized in this study. This finding suggests that payment effects might extend to the next task, or that fill-in-the-blank methods may be influenced by incentivization. Given that these methods are valuable measures of discounting (e.g. Weatherly & Derenne, 2011) replications are needed to confirm this result.

Fourth, we only tested three delay discounting measures. Although these methods are commonly used, there are several other frequently used tasks. For example, the 5-trial task which presents a series of questions between some amount of a delayed commodity and half that amount available immediately, and varies the delay at which the full commodity would be available (Koffarnus & Bickel, 2014). Moreover, there are titrating amounts tasks where participants are presented six choices at seven delays, where the first choice between a delayed reward and an immediate reward calibrates the immediate value of subsequent choices. For example, if participants had chosen the immediate reward, the next immediate reward would decrease. This procedure allows researchers to converge on the subjective value of the delayed reward (Du, Green, & Myerson, 2002). We did not include these measures in the present study because investigating several measures in a single experiment can be problematic; it would introduce additional confounds, which include participant boredom, habituation to the stimuli (as for eyespots in Sparks & Barclay, 2013), and can produce demand characteristics. Future research should look at the effect of incentivization the 5-trial tasks and titrating amounts.

Last, the current study produced the unanticipated result that payment may attenuate correlations between discounting and some instantiations of impulsivity. We tentatively propose that incentivization may better predict "real" or incentivized behaviors, which were not assessed in the present study. On the other hand, hypothetical incentives may be better correlated with other non-incentivized measures. If either of these speculations is accurate, it calls into question the validity of delay discounting measures. That is, if payment changes what is being measured, what are incentivized delay discounting measures actually measuring? We speculate that it may restrict what is being measured to financial risk-taking, and no longer generalize to other manifestations of impulsivity. In fact, a recent study on self-regulation, a related construct to delay discounting, suggests that behavioral measures have lower test-retest reliabilities than survey measures (Enkavi et al., 2019). This finding may reflect the fact that behavioural measures largely assess situation-specific behaviours while self-report measures largely assess traits. Consequently, it is possible that hypothetical delay discounting measures may assess more trait-based discounting while incentivized delay discounting tasks may assess more situation-contingent behaviors. This hypothesis would offer the prediction that correlations would be stronger between theoretically-related surveys (e.g., impulsivity) and hypothetical discounting tasks than for incentivized tasks, which is what we observed. However, future research is necessary to appropriately address these hypotheses.

Conclusion

Our results demonstrate the importance of testing different discounting measurements. At present, there is little consistency in how discounting tasks are administered, which may have critical consequences on study outcomes. It is often thought that offering incentives for one or

more delay discounting choices yields more attentive responding from participants and is more ecologically valid. The present study finds that providing incentives may reduce relationships between delay discounting and measures that are theoretically related, which brings into question the ecological validity of the incentivized measure. In other words, what do incentivized delay discounting measures assess, if not time preference? This preliminary result suggests that experimental costs may be reduced by eliminating measurement incentives in online contexts. However, more research is needed for a stronger conclusion. Moreover, we found that correlations between delay discounting and theoretically related measures were smaller than reported in other studies, suggesting that online studies using delay discounting tasks should use conservative estimates to calculate power. We look forward to further work in this area.

ELECTRONIC SUPPLEMENATRY MATERIAL

- ESM 1. Tables S1 – S6; Figures S1-S2. Supplementary Material.

Contains pilot results, descriptive statistics, corrplots, order effects analyses, and payment condition by MCQ reward sizes analyses.

REFERENCES

Alessi, S. M., & Petry, N. M. (2003). Pathological gambling severity is associated with impulsivity in a delay discounting procedure. *Behavioural Processes, 64*, 345-354. doi:10.1016/S0376-6357(03)00150-5

Amir, O., Rand, D. G., & Gal, Y. K. (2012). Economic games on the internet: The effect of $1 stakes. *PLoS ONE, 7*(2), e31461. doi://10.1371/journal.pone.0031461

Bickel, W. K., Odum, A. L., & Madden, G. J. (1999). Impulsivity and cigarette smoking: Delay discounting in current, never, and ex-smokers. *Psychopharmacology, 146*, 447-454. doi:10.1007/PL00005490

Bohannon, J. (2016). Mechanical Turk upends social sciences. *Science, 352*, 1263-1264. doi:10.1126/science.352.6291.1263

Brase, G. L. (2009). How different types of participant payments alter task performance. *Judgment and Decision Making, 4*, 419-428.

Brooker, I. S., Clara, I. P., & Cox, B. J. (2009). The Canadian Problem Gambling Index: Factor structure and associations with psychopathology in a nationally representative sample. *Canadian Journal of Behavioural Science, 41*, 109-114. doi:10.1037/a0014841

Bühren, C., & Kundt, T. C. (2015). Imagine being a nice guy: A note on hypothetical vs. incentivized social preferences. *Judgment and Decision Making, 10*, 185-190.

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. doi:10.1177/1745691610393980

Callan, M. J., Shead, N. W., & Olson, J. M. (2011). Personal relative deprivation, delay discounting, and gambling. *Journal of Personality and Social Psychology, 101*, 955–973. doi:10.1037/a0024778

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty, 19*, 7-42. doi:10.1023/A:1007850605129

Chandler, J., & Shapiro, D. (2016). Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology, 12*, 53-81. doi:10.1146/annurev-clinpsy-021815-093623

Dixon, M. R., Lik, N. M. K., Green, L., & Myerson, J. (2013). Delay discounting of hypothetical and real money: The effect of holding reinforcement rate constant. *Journal of Applied Behavior Analysis, 46*, 512-517. doi:10.1002/jaba.42

Du, W., Green, L., & Myerson, J. (2002). Cross-cultural comparisons of discounting delayed and probabilistic rewards. *The Psychological Record, 52*, 479-492. doi:10.1007/BF03395199

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences, 116*, 5472-5477. doi: 10.1073/pnas.1818430116

Eysenck, S. B. G., Pearson, P. R., Easting, G., & Allsopp, J. F. (1985). Age norms for impulsiveness, venturesomeness and empathy in adults. *Personality and Individual Differences, 6*, 613-619. doi:10.1016/0191-8869(85)90011-X

Ferrey, A. E., & Mishra, S. (2014). Compensation method affects risk-taking in the Balloon Analogue Risk Task. *Personality and Individual Differences, 64*, 111-114. doi:10.1016/j.paid.2014.02.008

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature, 40*, 351-401. doi:10.1257/002205102320161311

Goodman, J. K., & Paolacci, G. (2017). Crowdsourcing consumer research. *Journal of Consumer Research, 44*, 196-210. doi:10.1093/jcr/ucx047

Gray, J. C., Amlung, M. T., Palmer, A. A., & MacKillop, J. (2016). Syntax for calculation of discounting indices from the monetary choice questionnaire and probability discounting questionnaire. *Journal of the Experimental Analysis of Behavior, 106*, 156-163. doi:10.1002/jeab.221

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin, 130*, 769-792. doi:10.1037/0033-2909.130.5.769

Hanoch, Y., Rolison, J., & Gummerum, M. (2013). Good things come to those who wait: Time discounting differences between adult offenders and nonoffenders. *Personality and Individual Differences, 54*, 128-132. doi:10.1016/j.paid.2012.08.025

Hardisty, D. J., Thompson, K. F., Krantz, D. H., & Weber, E. U. (2013). How to measure time preferences: An experimental comparison of three methods. *Judgment and Decision Making, 8*, 236-249.

Harrell, F. E., Jr. (2018). Hmisc: Harrell miscellaneous. R package version 4.1-1. [Computer software]. R package version 4.1-1. Available from http://cran.r-project.org/package=Hmisc.

Hervé, M. (2018). RVAideMemoire: Testing and plotting procedures for biostatistics. [Computer software]. R package version 0.9–69-3. Available from https://cran.r-project.org/packages=RVAideMemoire

Irwin, J. R., McClelland, G. H., & Schulze, W. D. (1992). Hypothetical and real consequences in experimental auctions for insurance against low-probability risks. *Journal of Behavioral Decision Making, 5*, 107-116. doi:10.1002/bdm.3960050203

Johnson, M. W., & Bickel, W. K. (2002). Within-subject comparison of real and hypothetical money rewards in delay discounting. *Journal of the Experimental Analysis of Behavior, 77*, 129-146. doi:10.1901/jeab.2002.77-129

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General, 128*, 78-87. doi:10.1037/0096-3445.128.1.78

Kirby, K. N., & Petry, N. M. (2004). Heroin and cocaine abusers have higher discounting rates for delayed rewards than alcoholics or non-drug-using controls. *Addiction, 99*, 461-471. doi:10.1111/j.1360-0443.2004.00669.x

Koffarnus, M. N., & Bickel, W. K. (2014). A 5-trial adjusting delay discounting task: Accurate discount rates in less than one minute. *Experimental and Clinical Psychopharmacology, 22*, 222-228. doi:10.1037/a0035973

Lagorio, C. H., & Madden, G. J. (2005). Delay discounting of real and hypothetical rewards III: Steady-state assessments, forced-choice trials, and all real rewards. *Behavioural Processes, 69*, 173-187. doi:10.1016/j.beproc.2005.02.003

List, J. A., & Gallet, C. A. (2001). What experimental protocol influence disparities between actual and hypothetical stated values? *Environmental and Resource Economics, 20*, 241-254. doi:10.1023/A:1012791822804

Madden, G. J., Begotka, A. M., Raiff, B. R., & Kastern, L. L. (2003). Delay discounting of real and hypothetical rewards. *Experimental and Clinical Psychopharmacology*, *11*, 139-145. doi:10.1037/1064-1297.11.2.139

Madden, G. J., Raiff, B. R., Lagorio, C. H., Begotka, A. M., Mueller, A. M., Hehli, D. J., Wegener, A. A. (2004). Delay discounting of real and hypothetical rewards: II. Between- and within-subject comparisons. *Experimental and Clinical Psychopharmacology, 12,* 251-261. doi:10.1037/1064-1297.12.4.251

Mishra, S., & Lalumière, M. L. (2011). Individual differences in risk-propensity: Associations between personality and behavioral measures of risk. *Personality and Individual Differences, 50,* 869–873. doi:10.1016/j.paid.2010.11.037

Mishra, S., & Lalumière, M. L. (2017). Associations between delay discounting and risk-related behaviors, traits, attitudes, and outcomes. *Journal of Behavioral Decision Making, 30,* 769-781. doi:10.1002/bdm.2000

Mishra, S., & Novakowski, D. (2016). Personal relative deprivation and risk: An examination of individual differences in personality, attitudes, and behavioral outcomes. *Personality and Individual Differences, 90,* 22-26. doi:10.1016/j.paid.2015.10.031

Odum, A. L. (2011). Delay discounting: Trait variable? *Behavioural Processes, 87,* 1-9. doi:10.1016/j.beproc.2011.02.007

Petry, N. M. (2001). Delay discounting of money and alcohol in actively using alcoholics, currently abstinent alcoholics, and controls. *Psychopharmacology*, *154,* 243-250. doi:10.1007/s002130000638

R Core Team (2018). R: A language and environment for statistical computing. [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.R-project.org/

Reimers., S., Maylor, E. A., Stewart, N., & Chater, N. (2009). Associations between a one-shot delay discounting measure and age, income, education and real-world impulsive behavior. *Personality and Individual Differences, 47,* 973-978. doi:10.1016/j.paid.2009.07.026

Revelle, W. (2018) psych: Procedures for personality and psychological research. [Computer software]. R package version 1.8.4. Available from https://CRAN.R-project.org/package=psychVersion = 1.8.4.

Reynolds, B. (2006). A review of delay-discounting research with humans: Relations to drug use and gambling. *Behavioural Pharmacology, 17,* 651-667. doi:10.1097/FBP.0b013e3280115f99

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analysis. *Journal of Statistical Software*, *17*, 1-25. doi:10.18637/jss.v017.i05

Robinson, J., Rosenzweig, C., Moss, A. J., & Litman, L. (2019). *Tapped out or barely tapped? Recommendations for how to harness the vast and largely unused potential of the Mechanical Turk participant pool*. Preprint. doi:10.31324/osf.io/jq589

Schouwenburg, H. C., & Groenewoud, J. (2001). Study motivation under social temptation; effects of trait procrastination. *Personality and Individual Differences, 30*, 229-240. doi:10.1016/S0191-8869(00)00034-9

Spinella, M. (2007). Normative data and a short form of the Barratt Impulsiveness Scale. *International Journal of Neuroscience, 117,* 359-368. doi:10.1080/00207450600588881

Tabri, N., Shead, N. W., & Wohl, M. J. A. (2017). Me, myself, and money II: Relative deprivation predicts disordered gambling severity via delay discounting, especially

among gamblers who have a financially focused self-concept. *Journal of Gambling Studies, 33,* 1201-1211. doi:10.1007/s10899-017-9673-7

Weatherly, J. N., & Derenne, A. (2011). Comparing delay discounting rates when using the fill-in-the-blank and multiple-choice methods. *The Journal of General Psychology, 138,* 300-318. doi:10.1080/00221309.2011.606442

Table 1

Overall (N = 221) Spearman's Rho correlations [95% CIs] between the three delay discounting measures (MCQ, SSD, and Matching Task; light grey), the continuous outcome variables (EIS, BIS, PGSI, PRDS, and Procrastination; dark grey), and between discounting and outcome measures (white).

	1. MCQ	2. SSD	3. Matching Task	4. EIS	5. BIS	6. PGSI	7. PRDS	8. Procrasti-nation
1. MCQ	-							
2. SSD	**.729*** ** [.658, .788]	-						
3. Matching Task	**.679*** ** [.583, .754]	**.587*** ** [.490, .673]	-					
4. EIS	.094 [-.035, .223]	.073 [-.055, .206]	**.204** ** [.062, .341]	-				
5. BIS	**.140*** [.006 .274]	**.133*** [.010, .257]	**.176** ** [.054, .300]	**.462*** ** [.342, .566]	-			
6. PGSI	**.194** ** [.057, .328]	**.126**† [.004, .239]	.063 [-.080, .203]	**.309*** ** [.185, .418]	**.288*** ** [.164, .403]	-		
7. PRDS	-.030 [-.155, .097]	-.023 [-.149, .104]	-.048 [-.181, .088]	**.162*** [.036, .285]	**.415*** ** [.293, .526]	**.140*** [.023, .259]	-	
8. Procrastination	.077 [-0.05, .212]	.111 [-.017, .238]	.110 [-.010, .241]	**.327*** ** [.208, .440]	**.668*** ** [.581, .744]	**.190** ** [.062, .304]	**.390*** ** [.258, .502]	
9. Alcohol	-.009 [-.126, .113]	-.018 [-.139, .113]	**-.127**† [-.249, -.002]	.044 [-.078, .183]	-.002 [-.137, .130]	**.262*** ** [.142, .382]	.061 [-.086, .180]	.075 [-.070, .210]

Notes: All tests are two-tailed Spearman's Rho. 95% confidence intervals were obtained through bootstrapping with 1000 repetitions. Significant or near significant correlations ($p < .07$) are in bold, with *** denoting correlations significant at the .001 level, ** the .01 level, * the .05 level, and †. $p < .07$. MCQ = Monetary Choice Questionnaire, SSD = Single Shot Discounting, EIS = Eysenck Impulsivity Scale, BIS = Barratt Impulsiveness Scale, PGSI = Problem Gambling Severity Index, PRDS = Personal Relative Deprivation Scale.

Table 2

Spearman's Rho correlations (+/- 95% CIs) for (A) aggregated discounting tasks by payment conditions, and separated by delay discounting task in the (B) payment condition (N = 111), and (C) hypothetical condition (N = 110) between the three delay discounting measures and continuous outcome variables.

Table 2A: Aggregated delay discounting measures

	EIS	BIS	PGSI	PRDS	Procrastination	Alcohol
Payment Condition	.083 [-.101, .276]	.120 [-.079, .307]	.008 [-.187, .218]	.008 [-.184, .199]	.161 [-.079, .338]	.031 [-.133, .203]
Hypothetical Condition	**.249**** [.064, .424]	**.237*** [.047, .410]	**.214*** [-.001, .398]	-.106 [-.302, .079]	.067 [-.103, .246]	**-.174†** [-.358, .004]

Table 2B: Payment Condition

	EIS	BIS	PGSI	PRDS	Procrastination	Alcohol
MCQ	.061 [-.122, .230]	.079 [-.100, .284]	.128 [-.073, .330]	.036 [-.167, .226]	.104 [-.082, .279]	-.052 [-.122, .221]
SSD	.021 [-.165, .201]	.078 [-.102, .254]	.142 [-.023, .323]	-.002 [-.182, .185]	.169 [-009, .340]	-.097 [-.094, .283]
Matching Task	.096 [-.109, .285]	.067 [-.124, .231]	.013 [-.167, .181]	-.007 [-.197, .196]	.135 [-.047, .292]	-.019 [-.151, .193]

Table 2C: Hypothetical Condition

	EIS	BIS	PGSI	PRDS	Procrastination	Alcohol
MCQ	.125 [-.060, .313]	**.195*** [.003, .367]	**.264**** [.083, .428]	-.098 [-.279, .076]	.053 [-.136, .234]	.068 [-.267, .113]
SSD	.128 [-.062, .308]	**.188*** [-.000, .362]	.108 [-.080, .284]	-.044 [-.218, .136]	.047 [-.144, .233]	.143 [-.314, .045]
Matching Task	**.313***** [.113, .474]	**.279**** [.099, .449]	.132 [-.090, .350]	-.076 [-.259, .113]	.088 [-.108, .257]	**.267**** [-.433, -.080]

Notes: Significant correlations are in bold, with *** denoting correlations significant at the .001 level, ** the .01 level, and * the .05 level. MCQ = Monetary Choice Questionnaire, SSD = Single Shot Discounting, EIS = Eysenck Impulsivity Scale, BIS = Barratt Impulsiveness Scale, PGSI = Problem Gambling Severity Index, PRDS = Personal Relative Deprivation Scale.