AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2023, Vol. 152, No. 9, 2651–2665
https://doi.org/10.1037/xge0001380

How Well Do Laboratory-Derived Estimates of Time Preference Predict Real-World Behaviors? Comparisons to Four Benchmarks

Daniel M. Bartels[1], Ye Li[2], and Soaham Bharti[1]
[1] Marketing Department, University of Chicago Booth School of Business
[2] Management Area, School of Business, University of California Riverside

A large literature implicates time preference (i.e., how much an outcome retains value as it is delayed) as a predictor of a wide range of behaviors, because most behaviors involve sooner and delayed consequences. We aimed to provide the most comprehensive examination to date of how well laboratory-derived estimates of time preference relate to self-reports of 36 behaviors, ranging from retirement savings to flossing, in a test–rest design using a large sample ($N = 1{,}308$) and two waves of data collection separated by 4.5 months. Time preference is significantly—albeit modestly—associated with about half of the behaviors; this is true even when controlling for 15 other demographic variables and psychologically relevant scales. There is substantial variance in the strengths of associations that is not easily explained. Time preference's predictive validity falls in the middle of these 16 possible predictors. Finally, we ask time preference researchers ($N = 55$) to predict the variation in the relationship between time preference and behaviors, and although they are reasonably well-calibrated, these experts tend to overestimate the predictive power of time preference estimates. We discuss implications of invoking time preference as a predictor and/or determinant of behaviors with delayed consequences in light of our findings.

Public Significance Statement

Multiple scientific fields describe people's future-oriented decisions in terms of time preference—how much value an outcome retains or loses as it is delayed from the present. Time preferences are assumed to predict people's behaviors with future consequences across various domains (e.g., finance, health). To that end, several papers have reported correlations of time preference with a few focal behaviors, and a smaller subset of these papers have examined such correlations for several behaviors. The current investigation is a more comprehensive accounting of how well time preference predicts behavior—(a) examining more behaviors than existing literature, (b) involving a large sample, as well as (c) using a test-retest design. We find correlations that are mostly modest and highly heterogeneous across behaviors. In addition, experts studying time preference can predict some of this heterogeneity in correlations but tend to systematically overestimate their size. This research underscores the need for greater understanding of the moderators that influence the relationship between time preference and behaviors with future consequences.

This article was published Online First May 25, 2023.
Daniel M. Bartels https://orcid.org/0000-0003-1642-8426
Ye Li https://orcid.org/0000-0002-5170-4732
Soaham Bharti https://orcid.org/0000-0001-9688-6181
We thank Stephen Atlas, Eddy Davelaar, Berkeley Dietvorst, Lin Fei, Alex Imas, Eric Johnson, Antonia Krefeld-Schwalb, Nick Reinholtz, Jack Soll, Christoph Ungemach, Oleg Urminsky, and two anonymous reviewers for their generous and helpful comments on the manuscript. We are also grateful to Øystein Daljord, Lorien Elleman, Lin Fei, Trevor Kvaran, Justin Landy, Sanjog Misra, Chris Nosko, Bill Revelle, and Matt Taddy for their suggestions regarding our analyses. We thank Stephen Atlas, Shane Frederick, Todd Gureckis, Alex Imas, Eric Johnson, Ki Kim, Stephan Meier, Rachel Meng, Oded Netzer, Ben Newell, Daniel Read, Nick Reinholtz, Marc Scholten, Dan Shaffer, Olivier Toubia, Christoph Ungemach, Oleg Urminsky, Liad Weiss, and Gal Zauberman for their suggestions regarding Supplemental Material Section A of the manuscript. Finally, we express our gratitude to Steve Atlas, Sachin Banker, Manel Baucells, Daniel Benjamin, Sudeep Bhatia, Andy Brownback, Eugene Caruso, Gretchen Chapman, Helen Colby, Xianchi Dai, Mette Trier-Daamgard, Stefano DellaVigna, J.P. Dube, Keith Ericson, Bernd Figner, Geoff Fisher, Jeremy Fox, Shane Frederick, Dan Goldstein, Christina Gravert, Adam Harris, Lisheng He, Hal Hershfield, Günter Hitsch, Alex Imas, Pranav Jindal, Bryan Jones, Michael Kuhn, Robyn LeBoeuf, Selin Malkoc, Samuel M. McClure, Joe McGuire, Stephan Meier, Ashwani Monga, Tim Pleskac, Drazen Prelec, Daniel Read, Crystal Reeck, Stian Reimers, Sally Sadoff, Adam Sanborn, Benjamin Scheibehenne, Marc Scholten, Charlie Sprenger, Neil Stewart, Christoph Ungemach, Elke Weber, Haewon Yoon, Gal Zauberman, and six other experts who wished to remain anonymous for their participation in Study 2 in the paper.

Some of the ideas and data appearing in this manuscript have been presented at annual meetings of the Association of Consumer Research (2022), the Society for Judgment and Decision Making (2022), and the Society for the Quantitative Analyses of Behavior (2019). The data, materials, and analysis code for this paper have been made available for review and will be publicly available once the paper is published: https://researchbox.org/459&PEER_REVIEW_passcode=ZPMQZP.

Daniel M. Bartels served as lead for conceptualization and writing–original draft. Ye Li served as lead for data curation, project administration, software, contributed equally to visualization and writing–review and editing, and served in a supporting role for conceptualization, investigation, and methodology. Soaham Bharti contributed equally to project administration, validation, and served in a supporting role for writing–original draft and writing–review and editing. Ye Li and Soaham Bharti contributed equally to formal analysis.

Correspondence concerning this article should be addressed to Daniel M. Bartels, Marketing Department, The University of Chicago Booth School of Business, 5807 S. Woodlawn Ave., Chicago, IL 60637, United States. Email: bartels@uchicago.edu

Keywords: time preference, temporal discounting, intertemporal choice, time discounting, delay discounting

Supplemental materials: https://doi.org/10.1037/xge0001380.supp

Most decisions entail tradeoffs between sooner outcomes and later outcomes—from the mundane (i.e., whether to order dessert) to the consequential (i.e., how much to save for retirement). Accordingly, most choice theories either assume or estimate people's time preferences as captured by a temporal discount factor or rate—that is, how much value an outcome retains or loses as it is delayed—to describe or predict these behaviors. A large literature articulates how estimates of time preference can help us understand and predict a variety of behaviors involving delayed consequences (for reviews, see Ericson & Laibson, 2019; Frederick et al., 2002; Read & Read, 2004). Hundreds of papers have explored the association between measures of time preference and behaviors, many of which we note below (Urminsky & Zauberman, 2015).

In this paper, we build on these previous explorations and aim to provide a thorough examination of how well laboratory-derived measures of time preference relate to a wide range of behaviors—by assessing more behaviors, with more covariates, using a test–retest design, and comparing to more benchmarks. We measured time preference in an online survey and, following prior work (e.g., Bradford et al., 2017; Chabris et al., 2008; Reimers et al., 2009), examined the correlations between time preference and people's self-reports of a variety of behaviors, as well as other demographic and psychologically relevant variables. As a secondary goal, we also assessed how well experts could predict the relationships between time preference and behaviors.

Associations of Time Preference With Behaviors

One major motivation for research on time preference is that it is a potential predictor of a wide range of behaviors. Papers have found significant associations between time preference estimates and behaviors involving delayed consequences, particularly when explaining differences between substance-using and control populations (MacKillop et al., 2011; Reynolds, 2006). Studies with the general population have found that estimates of time preference significantly correlate with smoking (Bradford, 2010; Reynolds et al., 2004; Sutter et al., 2013), alcohol use (Bradford et al., 2017; Sutter et al., 2013; Vuchinich & Simpson, 1998), credit card debt (Bradford et al., 2017), mortgage choice (Atlas et al., 2017), credit scores (Li et al., 2015; Meier & Sprenger, 2012), savings (Angeletos et al., 2001; Bradford et al., 2017; Sutter et al., 2013), educational attainment (Duckworth & Seligman, 2005; Falk et al., 2018; Reed & Martens, 2011), gambling (Dixon et al., 2003; Petry, 2001), exercise (Bradford, 2010; Bradford et al., 2017), and more (Bradford et al., 2017; Chabris et al., 2008; Reimers et al., 2009).

Most behaviors have delayed consequences, but ex ante, it is not obvious which behaviors we should expect to correlate most strongly with estimates of time preference or why. With the previous significant correlations noted, researchers have also found nonsignificant correlations between measures of time preference and behaviors that one might expect to be related to time preference, including dental checkup frequency (Farrell & Fuchs, 1982), dietary behaviors, flossing, gambling, percentage of income saved, late credit card payments, wealth (Chabris et al., 2008), efficient energy use, and using sunscreen (Bradford et al., 2017). Similarly, evidence linking time preference to obesity and body mass index (BMI) is mixed (Barlow et al., 2016).

Studies assessing how well time preference predicts behaviors often do not control for other relevant and/or potentially confounding variables. Most studies that control for covariates typically incorporate demographics and a few other variables directly relevant to the target behaviors, as opposed to psychological factors frequently associated with a range of behaviors. A few notable exceptions include studies that controlled for risk and ambiguity aversion (Bradford, 2010; Bradford et al., 2017; Sutter et al., 2013) or cognitive ability (Chabris et al., 2008; Li et al., 2013, 2015).

Overview of Studies

We intend to provide a more comprehensive approach to examining the relationship between time preference and behavior on several dimensions. First, we incorporate a wider range of target behaviors with delayed outcomes, adding behaviors based on their variation across a set of 23 differentiating characteristics as assessed in a separate norming study (see Section A in the online supplemental materials). Second, we collect a large and widely varied set of additional variables to control for and compare to, including various personality scales and predictors relevant to financial decision-making. Third, with some exceptions (Li et al., 2013, 2015), studies of the correlation between time preference and behaviors do not account for attenuation due to imperfect measurement reliability.[1] We address this by collecting two waves of data, separated by 4.5 months, to assess the test–retest reliability of time preference, the 36 behaviors, and the 15 covariates. Fourth, we try to understand the considerable heterogeneity that we observe in the relationship between time preference and these 36 behaviors. Fifth, we elicit expert forecasts of these correlations to examine whether the widely varying magnitudes of correlation that we observe are, in fact, predictable.

We aim to assess how well time preference predicts behavior, and we find that the answer to this question depends on the basis for comparison. We used four benchmarks, comparison: (a) *to zero*: how often does time preference significantly correlate with each of these 36 behaviors?, (b) *across behaviors*: which variables or domains are more correlated with time preference?, (c) *to other predictors*: how well does time preference do in predicting behavior relative to demographics and other psychologically relevant individual differences?, and (d) *to expert forecasts*: can time preference researchers predict the magnitude of time preference's association with these 36 behaviors? We believe that this benchmark is especially important since expert intuitions about these relationships directly influence which behaviors they study to gather empirical and theoretical support for the predictive validity of time preference.

Broadly, we found that although most correlations between time preference and the 36 behaviors are small to moderate (Cohen, 1988), time preference significantly predicted over half of the

[1] *Attenuation* occurs when correlations are underestimated due to measurement error in one or both underlying variables (Spearman, 1904). Correlations can be *disattenuated* by adjusting for measurement reliability: disattenuated $r_{x,y} = r_{x,y}/\mathrm{sqrt}(r_{x,x} \times r_{y,y})$.

behaviors, even when controlling for 15 demographic and psychological covariates. Also, correlations between time preference and behaviors are greater when we consider aggregated (multi-behavior) indices of financial, health, and prudential behavior, consistent with prior work (Chabris et al., 2008). At the same time, our data also suggest that aggregating by domains may not be justifiable due to considerable heterogeneity within the domain. We found large differences in these correlations across behaviors that are not well predicted by domains nor explained by ratings on 23 differentiating characteristics that formed the basis for sampling these 36 behaviors (see Section A in the online supplemental materials).

In addition, while time preference consistently predicts behavior better than some covariates, it is in the middle of the pack of 16 predictors. We observed this result even though we specifically selected these behaviors because they have delayed consequences. Finally, we found mixed evidence on the predictability of the size of these correlations. Experts' average forecasts of these relationships were positively correlated with the actual correlations in Study 1. However, even though the average expert prediction for the correlation between time preference and the 36 behaviors in our study was modest ($r = 0.11$), experts still tended to overestimate the associations, particularly for behaviors that were not correlated with time preference.

Study 1 examines the first three benchmarks (i.e., comparison to zero, across behaviors, and across predictors). In addition, Study 1 accounts for the issue of test–retest reliability for all the variables measured by incorporating data collected over two waves separated by about 4.5 months. Given the heterogeneity in predictive validity of time preference across the behaviors, we wanted to assess whether such heterogeneity was explainable or expected by experts. In addition, comparison to expert predictions potentially offers a fourth benchmark for these correlations. Study 2, therefore, examines how well forecasts of the correlation between time preference and each of 36 behaviors aligned with the heterogeneity in observed correlations across behaviors. We believe that this benchmark is especially important because expert intuitions about these relationships influence which behaviors are studied, which research questions are asked, and how researchers ultimately assess theoretical support for explanatory models involving time preference.

Study 1

In Study 1, we conducted a large, two-wave study measuring people's time preference, 36 self-reported behaviors, and various other psychologically relevant scales. Our intention was to run a large-scale examination of how well time preference predicts these behaviors while controlling for measurement error.

Method

Transparency and Openness

We report all data exclusions (if any) and all measures in the study. All data are publicly available (see author's note). Data were analyzed using R, Version 3.6.2 (R Core Team, 2019) and the following packages: *tidyverse* Version 1.3.0 (Wickham et al., 2019); *dplyr* Version 1.0.7 (Wickham et al., 2021); *afex* Version 1.0-1 (Singmann et al., 2021); *reghelper* Version 0.3.6 (Hughes, 2020); and *Hmisc* Version 4.6-0 (Harrell, 2019). This study's design and analyses were not preregistered. The protocols of Studies 1 and 2 were approved by the institutional review board.

Participants

We recruited 1,576 U.S. participants in total from Amazon Mechanical Turk ($N = 774$) and from a market research firm, PureProfile ($N = 802$), to complete two waves of the study in December 2013 and April 2014 (see Table 1 for descriptive statistics).[2] Of these 1,576 participants, 83% finished both waves for a total of 1,308 complete two-wave data points (1,308/1,576; MTurk = 604/774; commercial panel = 704/802). To test for selective attrition, we ran a logistic regression predicting attrition using the wave 1 observations of the 16 predictor variables in our study (see Table S1 in the online supplemental materials). After a Bonferroni adjustment, only panel source, parent's average education level, and numeracy/cognitive reflection (CRT) predicted participation in wave 2, so results pertaining to these variables should be interpreted with caution. For robustness, we replicate our main analyses using only wave 1 data in Table S2 in the online supplemental materials.

Procedure

Participants completed both waves of the survey online after completing a separate intake survey measuring demographics (age, gender, income, education, native/first language, and zip code). Participants completed the first wave between December 14, 2013 and January 3, 2014 (average date of December 18) and the second wave between April 22 and May 12, 2014 (average date of April 25), about 4.5 months later ($M = 128.90$ days, $SD = 3.21$, range = 110–147 days).

Time Preference Measure

We developed a 12-item battery of intertemporal choice questions posing smaller-sooner versus larger-later tradeoffs (see Table 2). The measure was developed over the course of several pretests that swapped in different alternatives to test different combinations of smaller-sooner and larger-later options, using ideas from item response theory. Our goal was to distinguish people who are more patient from those who are less patient, rather than to estimate a specific discount rate, discount factor, or estimate of present bias. (Full details about the development of the battery are available in Section D in the online supplemental materials.) We also recognize that measures like this one (and almost all others using smaller, sooner vs. larger, later monetary tasks) are affected by more than just patience, including many economic considerations, like how much money a person has now and will have in the future relative to current and future needs, their declining marginal utility for money, trust that the later payment will occur and myriad other uncertainties regarding the future, like guesses about inflation, etc. We estimated participants' time preference by simply counting the number of larger-later choice options chosen among the 12 items. Figure 1 presents a histogram of the number of larger-later options chosen.

[2] We collected data from two pools in case there were data quality issues with either pool. Notably, some concerns over MTurk data quality have arisen in recent years (e.g., Chmielewski & Kucker, 2020). Our data are of high quality by recent standards: 93.37% of our participants passed the attention checks in both waves of this study.

Table 1
Study 1 Participant Summary Statistics

Variable	Mean/%	Median	*SD*	Min	Max
Age	40.92	38	14.64	18	86
Gender	41.51% male	—	—	—	—
Source	46.18% MTurk	—	—	—	—
Pretax household income (in USD)	26.76% earning $50,000 or above	$30,000–$39,999	—	Less than $10,000	More than $250,000
Education	41.97% with at least bachelor's degree or equivalent	Associate's degree or equivalent	—	Some high school or less	Doctoral degree (e.g., PhD)

Participants chose an average of 5.25 later options ($SD = 3.01$, $Mdn = 4.5$).

Of course, no measure of time preference uses every possible monetary value and delay, and it is possible that the relationship between estimates of time preference and behaviors could be affected by these parameters.[3] For this reason and others, we also collected a second, prominent measure of time preference as a benchmark, DEEP Time (Toubia et al., 2013), an adaptive measure that estimates parameters for people's long-term discount factor (δ) and present bias parameter (β) as stipulated by the quasi-hyperbolic discounting model (Laibson, 1997). This measure uses smaller amounts of money and shorter delays than our 12-item measure. We chose to focus on the 12-item measure because (a) it makes no parametric assumptions, (b) it is easier to implement and compute for researchers, (c) we did not have hypotheses about how these behaviors would be differentially predicted by β versus δ, and (d) the number of larger, later choices on our 12-item measure had higher test–retest reliability than did parameter estimates from the DEEP method ($r_{12\text{-item}} = 0.70$ vs. $r_\beta = 0.63$ and $r_\delta = 0.63$, $zs = 3.37$, $p < .001$). Note that the 12-item measure correlates highly with the DEEP Time parameter estimates for δ ($r = 0.76$) but less well with β ($r = 0.37$).

Behaviors

We asked participants to self-report the extent to which they do 36 behaviors (or about outcomes associated with those behaviors). We report response scales for each behavior in Section C in the online supplemental materials and descriptive statistics in Table S3 in the online supplemental materials. In addition, an early goal of the project was to sample behaviors that would theoretically be highly differentiated across the 23 characteristics that a separate set of participants would rate for each behavior (see Section A in the online supplemental materials). This theoretically driven sampling procedure, which includes suggestions from several time preference researchers and some of the papers cited earlier, resulted in introducing measures for some additional behaviors. See Section C in the online supplemental materials for the wordings, scaling, and sources for all measures.

We scored behaviors so that larger values indicate what participants in a separate norming study judged to be more far-sighted, prudent, or responsible behavior (see Section E in the online supplemental materials). This required reverse-scaling responses for missing credit payments, accumulating a lot of credit card debt, accumulating a lot of educational loan debt, smoking, gaining excess weight, using recreational drugs, drinking coffee, overeating, drinking alcoholic beverages, getting tattoos, gambling or buying lottery tickets, leaving dirty dishes overnight, and driving recklessly. In the tables and figures in this paper, the word "NOT" precedes these behaviors to indicate reverse scaling. In addition, for behaviors with highly skewed distributions of responses, we log-transformed the responses for both waves and averaged these log-transformed scores for: accumulating a lot of credit card debt, having kids when older, getting tattoos, earning a large income, keeping physically active, and actively exercising.

Table 2
Study 1 12-Item Battery of Intertemporal Choices

Smaller-sooner option	Larger-later option
$504 now	$524 in 1 month
$600 now	$611 in 1 month
$777 now	$791 in 1 month
$1,064 now	$1,153 in 1 month
$816 now	$860 in 3 months
$457 now	$551 in 6 months
$816 now	$5,440 in 1 year
$213 now	$281 in 2 years
$816 in 6 months	$860 in 9 months
$816 in 6 months	$1,028 in 1 year
$400 in 6 months	$440 in one and a half years
$840 in 6 months	$10,125 in two and a half years
$791 today	$777 in 1 month
$621 in 6 months	$670 in 6 months

Note. The last two (gray-font) items are attention-check foils that were included in wave 2 but not evaluated as part of the measure. The order of all 12/14 items was randomized by participant.

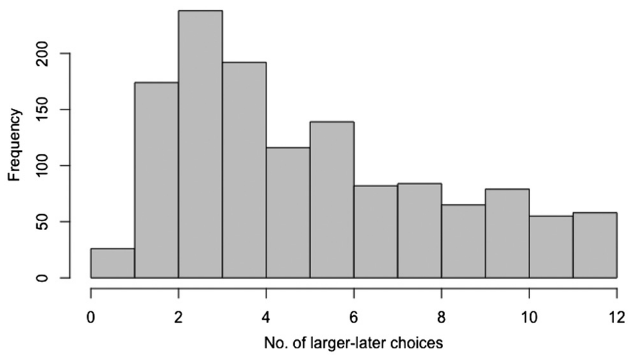
Other Variables

We included 15 covariates to serve as our third benchmark. For demographics, we measured age, gender, parents' education levels, and a sample (MTurk vs. market research panel) variable to control for other unmeasured differences. Own education level and income were two of the 36 behavioral dependent variables. We averaged parents' education levels into a single measure for ease of reporting. We measured five dimensions of personality using the 44-item Big Five Inventory (John & Srivastava, 1999), and assessed six other scales that have been related to temporal discounting and financial decision-making: (a) impulsiveness, as measured by the Barratt Impulsiveness Scale (Patton et al., 1996); (b) financial literacy (Fernandes et al., 2014); (c) a combined 8-item (Li et al., 2013) questionnaire measuring Numeracy (Lipkus et al., 2001) and Cognitive Reflection (Frederick, 2005), which were correlated 0.87 and thus collapsed into a single Numeracy/CRT scale; (d) tightwad-

[3] We thank an anonymous reviewer for raising this point.

Figure 1

Distribution of Responses on the 12-Item Time Preference Measure



spendthrift tendency—whether people typically spend more/less than they would like to (Rick et al., 2008); (e) planning propensity—people's propensity to plan out the use of their money over the next few days and next 1–2 months (Lynch et al., 2010); and (f) risk preference, as assessed by the single-item Eckel-Grossman measure (2002). Controlling for the 15 covariates also allows us to assess the unique association time preference has with behavior.

Results

Benchmark 1: Comparison of Predictive Validity of Time Preference Against Zero

Figure 2 and Table 3 show the degree of association between time preference and each of our variables using five different methods. All methods used data from the participants who completed both waves of data collection, and the associations between time preference and each of our variables are depicted graphically in Figure S1 in the online supplemental materials. The first column of Table 3 reports the Pearson correlations between our measure of time preference (higher scores = more patience/less discounting) and the 36 self-reported behaviors, with each variable averaged across two waves of data collection.

In column 2, we report the range of correlations we obtained from the four combinations of two waves of data (i.e., Wave 1/2 Time Preferences × Wave 1/2 Behaviors). In the same column, we also report, in parentheses, the number of times the relationship between time preference and the behaviors was statistically significant (at the $p < .05$ level).

The third column reports the corresponding disattenuated correlations (Spearman, 1904). This procedure is a correction for correlations that accounts for the imperfect test–retest reliability of our measures.[4] To do this, we used the reliabilities (see Table 4) for each variable and divided the correlation coefficient by the square root of the product of the reliabilities of the relevant variables.

The fourth column reports the standardized coefficients (i.e., betas) for time preference from 36 separate regressions predicting each behavior as a function of time preference while controlling for the 15 covariates (i.e., the unique associations between time preference and behavior). Column 5 presents the range of standardized coefficients, analogous to column 2.

Comparing across the five columns, whether or not a relationship reaches statistical significance is pretty consistent across the various ways we analyze the data. As depicted in Figure 2, the size of the various coefficients is similar across these different specifications. As another robustness check, we also present correlations and standardized coefficients for analyses using rank-ordered data in Table S2 in the online supplemental materials.

The correlation with time preference was significant at $p < .05$ for 20 of the 36 behaviors ($p < .01$ for 17 behaviors; $p < .001$ for 12 behaviors). Similarly, in regressions with all 15 covariates, time preference emerged as a positive and significant predictor at $p < .05$ for 18 of the 36 behaviors ($p < .01$ for eight behaviors; $p < .001$ for five behaviors). Thus, even when controlling for all other covariates, time preference was a significant predictor for over half of the behaviors. However, most of the correlation coefficients between time preference and the behaviors were moderate or small in size. The 25th, 50th, and 75th percentiles of absolute correlation coefficients were .03, .07, and .12 (.05, .10, and .16 for the corresponding absolute disattenuated correlations; .02, .06, and .07 for the median absolute standard betas).[5]

So, by the first benchmark, time preference predicts many behaviors significantly, but maybe not as strong as some might expect it to. We also found that the inference one reaches about the relative size of each correlation does not vary much across different analyses and is relatively robust to issues related to measurement error.[6]

Benchmark 2: Comparisons Across Behaviors

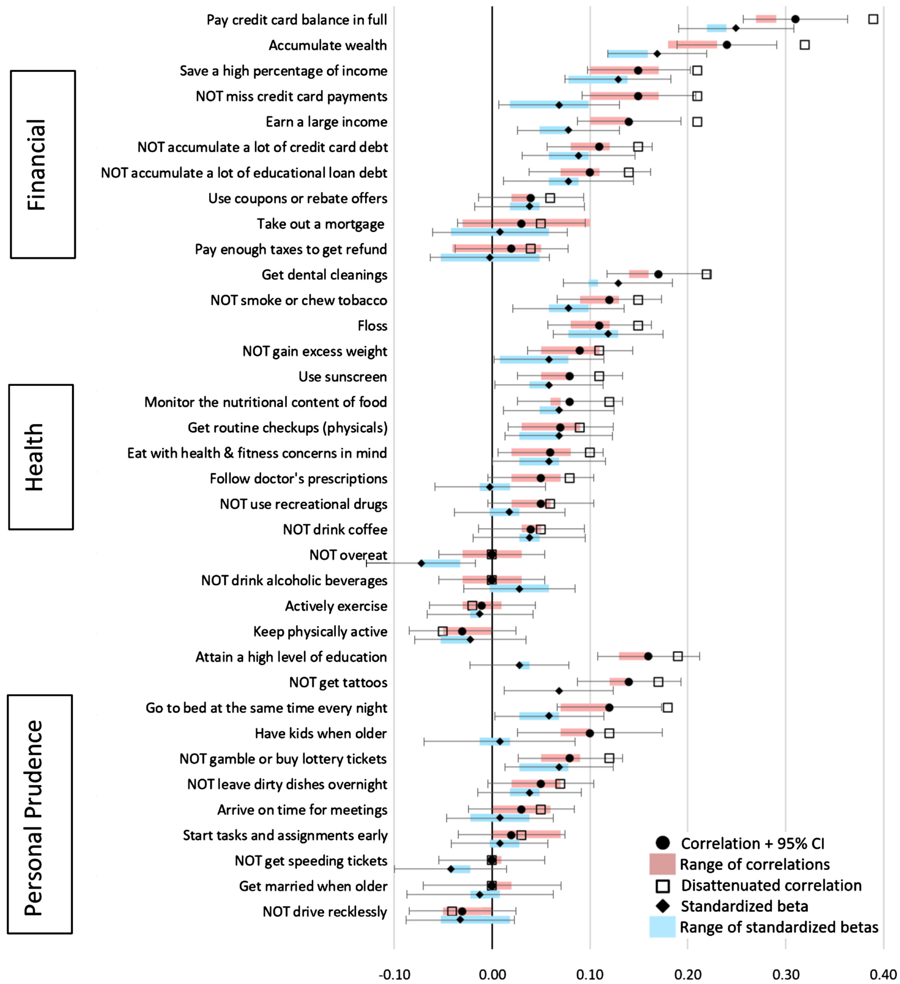
Following previous research (Bradford et al., 2017; Chabris et al., 2008), we classified the 36 behaviors into three domains: Financial, Health, and Personal Prudence (see Table 3). We created composite indices for each domain by (a) z-scoring across participants for each behavior and (b) computing each participant's average z-score within each domain. In line with previous

[4] We also calculated obviously related instrumental variables (ORIV) correlations and multiple regression coefficients, which is an econometric method for adjusting for attenuation (Gillen et al., 2019). See Table S4 in the online supplemental materials for these results. One limitation of such diattenuation-based approaches is they are unable to assess the relative influence of preference fluctuations versus measurement noise. Our analyses reveal consistent results when performed over data from the first wave only (see Table S2 in the online supplemental materials), suggesting that this limitation of the method may not be a pressing concern in our data.

[5] We thank an anonymous reviewer for suggesting an additional regression specification controlling for demographic (age, gender, and income). Table B3 in the online supplementary materials presents these analyses. While some associations were weaker when controlling for demographics (e.g., getting routine physicals, accumulating wealth, education level, having kids when older), none of these differences were significant.

[6] Table B4 in the online supplementary materials presents the association between these behaviors and the DEEP measures of time preference. Overall, the results are highly consistent between the 12-item measure and the δ parameter from DEEP. In particular, both the measures significantly predict a largely overlapping set of behaviors. On the other hand, results with the β parameter were less consistent. This is reasonable, since the 12-item measure was designed as a measure to discriminate among people with generally different time preferences (which correspond better to δ), whereas the purpose of the β parameter in quasi-hyperbolic models is to characterize a discontinuity—a change in time preferences for different (i.e., now vs. not-now) time periods.

Figure 2
Graphical Depiction of Associations



Note. Correlations are plotted as circles with 95% confidence intervals, while disattenuated correlations are plotted as squares. Standardized betas are plotted as diamonds with 95% confidence intervals. The pink and blue shaded bars represent the range of four possible correlations/standardized betas calculable using our two observations of each variable (wave 1-wave 1, 1–2, 2–1, and 2–2). See also Table 3.

research, we found that time preference was a better predictor of these aggregate indices of behavior than it was for most individual behaviors—financial: $r(1{,}306) = 0.27$, health: $r(1{,}306) = 0.16$, personal: $r(1{,}306) = 0.15$, $ps < .001$; the corresponding standardized betas from regressions on these indices, using time preference and the other 15 covariates as predictors were 0.17, 0.10, and 0.06, $ps < .05$. Time preference was a better predictor of the financial index than the health and personal prudence indices ($ts > 2.80$, $ps < .005$), while being similarly predictive of those latter two ($t = -0.12$, $p = .90$). This finding is perhaps unsurprising because, following the prior literature, we measured time preference by offering people sooner and later monetary rewards. See Section B in the online supplemental materials for a discussion on the extent to which collapsing these behaviors into domains for the purpose of assessing their relationship with time preference might or might not be informative.

To examine differences in the predictive validity of time preference across domains, we ran three analyses of variance to compare each domain's (a) correlation coefficients, (b) disattenuated correlation coefficients, and (c) standardized regression coefficients, finding significant differences in all three specifications, $Fs(2, 35) = 3.73$, 4.30, and 4.48; $ps < .05$. In particular, the associations with financial behaviors were greater than those for both health-related and personal behaviors. At the same time, these results were driven almost entirely by the large correlation for "pay credit card balance in full," which is an outlier (it is 2.05 interquartile ranges above the 75th percentile and 3.12 standard deviations above the average correlation; Tukey, 1977). However, there was a large variation in predictive validity of time preference in each domain and participants' responses for behaviors within a given domain were only modestly correlated (average correlations for financial, health, and prudential behaviors were 0.09, 0.07, and 0.08).

Table 3
Study 1 Measures of Association Between Time Preference and 36 Self-Reported Behaviors

Domain	Behaviors	Test–retest reliability	Correlation	Across 4 correlation combinations	Disattenuated correlation	Standardized betas (controlling for 15 covariates)	Across 4 regression combinations
Financial	Pay credit card balance in full	0.86	0.31***	0.27 to 0.29 (4)	0.39	0.25***	0.22 to 0.24 (4)
	Accumulate wealth	0.77	0.24***	0.18 to 0.23 (4)	0.32	0.17***	0.12 to 0.16 (4)
	Save a high percentage of income	0.76	0.15***	0.10 to 0.17 (4)	0.21	0.13***	0.08 to 0.14 (4)
	NOT miss credit card payments	0.71	0.15***	0.10 to 0.17 (4)	0.21	0.07*	0.02 to 0.10 (2)
	Earn a large income	0.65	0.14***	0.10 to 0.14 (4)	0.21	0.08**	0.05 to 0.08 (3)
	NOT accumulate a lot of credit card debt	0.78	0.11***	0.08 to 0.12 (4)	0.15	0.09**	0.06 to 0.10 (3)
	NOT accumulate a lot of educational loan debt	0.78	0.10**	0.07 to 0.11 (3)	0.14	0.08*	0.04 to 0.09 (1)
	Use coupons or rebate offers	0.75	0.04	0.02 to 0.04 (0)	0.06	0.04	0.02 to 0.05 (0)
	Take out a mortgage	0.70	0.03	−0.03 to 0.10 (1)	0.05	0.01	−0.04 to 0.06 (0)
	Pay enough taxes to get refund	0.51	0.02	−0.04 to 0.05 (0)	0.04	0	−0.05 to 0.05 (0)
	Get dental cleanings	0.84	0.17***	0.14 to 0.16 (4)	0.22	0.13***	0.10 to 0.11 (4)
Health	NOT smoke or chew tobacco	0.90	0.12***	0.09 to 0.13 (4)	0.15	0.08**	0.06 to 0.10 (4)
	Floss	0.82	0.11***	0.08 to 0.12 (4)	0.15	0.12***	0.08 to 0.13 (4)
	NOT gain excess weight	0.96	0.09**	0.05 to 0.11 (2)	0.11	0.06*	0.01 to 0.08 (2)
	Use sunscreen	0.75	0.08**	0.06 to 0.08 (4)	0.11	0.06*	0.04 to 0.06 (2)
	Monitor the nutritional content of food	0.62	0.08**	0.06 to 0.07 (4)	0.12	0.07*	0.05 to 0.07 (1)
	Get routine checkups (physicals)	0.78	0.07*	0.03 to 0.09 (3)	0.09	0.07*	0.03 to 0.07 (2)
	Eat with health & fitness concerns in mind	0.52	0.06*	0.02 to 0.08 (2)	0.1	0.06*	0.03 to 0.07 (2)
	Follow doctor's prescriptions	0.58	0.05	0.02 to 0.07 (1)	0.08	0	−0.01 to 0.02 (0)
	NOT use recreational drugs	0.87	0.05	0.02 to 0.06 (1)	0.06	0.02	0.00 to 0.03 (0)
	NOT drink coffee	0.89	0.04	0.03 to 0.05 (0)	0.05	0.04	0.03 to 0.05 (0)
	NOT overeat	0.63	0	−0.03 to 0.03 (0)	0.00	−0.07*	−0.07 to −0.03 (2)
	NOT drink alcoholic beverages	0.84	0	−0.03 to 0.03 (0)	0.00	0.03	0.00 to 0.06 (1)
	Actively exercise	0.71	−0.01	−0.03 to 0.01 (0)	−0.02	−0.01	−0.02 to −0.01 (0)
	Keep physically active	0.59	−0.03	−0.05 to 0.00 (0)	−0.05	−0.02	−0.05 to 0.02 (0)
	Attain a high level of education	0.96	0.16***	0.13 to 0.16 (4)	0.19	0.03	0.03 to 0.04 (0)
Personal	NOT get tattoos	0.97	0.14***	0.12 to 0.14 (4)	0.17	0.07*	0.06 to 0.06 (4)
	Go to bed at the same time every night	0.65	0.12***	0.07 to 0.12 (4)	0.18	0.06*	0.03 to 0.07 (2)
	Have kids when older	0.97	0.10*	0.07 to 0.10 (3)	0.12	0.01	−0.01 to 0.02 (0)
	NOT gamble or buy lottery tickets	0.71	0.08**	0.05 to 0.09 (3)	0.12	0.07*	0.03 to 0.08 (2)
	NOT leave dirty dishes overnight	0.77	0.05	0.02 to 0.07 (2)	0.07	0.04	0.02 to 0.05 (0)
	Arrive on time for meetings	0.66	0.03	0.00 to 0.06 (1)	0.05	0.01	−0.02 to 0.04 (0)
	Start tasks and assignments early	0.61	0.02	−0.01 to 0.06 (1)	0.03	0.01	0.00 to 0.03 (0)
	NOT get speeding tickets	0.71	0	0.00 to 0.01 (0)	0	−0.04	−0.04 to −0.02 (0)
	Get married when older	0.93	0	0.00 to 0.02 (0)	0	−0.01	−0.02 to 0.01 (0)
	NOT drive recklessly	0.64	−0.03	−0.05 to 0.01 (0)	−0.04	−0.03	−0.05 to 0.02 (0)

Note. (1) Pearson correlations between our measure of time preference (higher scores = more patience/less discounting) and the 36 self-reported behaviors, with each variable averaged across two waves of data collection. (2) Range of correlations obtained from the four combinations of two waves of data (i.e., Wave 1/2 Time Preferences × Wave 1/2 Behaviors) and, in parentheses, the number of times the relationship between time preference and the behaviors was statistically significant (at the $p < .05$ level). (3) Disattenuated correlations (Spearman, 1904), which account for the imperfect test–retest reliability of our measures (from Table 4), equal to the correlation coefficient divided by the product of the square root of the reliabilities of the relevant variables. (4) The standardized coefficients (i.e., betas) for time preference from 36 separate regressions predicting each behavior as a function of time preference while controlling for the 15 covariates. (5) The range of standardized coefficients, analogous to column 2.
* $p < .05$. ** $p < .01$. *** $p < .001$.

And finally, an early goal of this project was to explain heterogeneity in these 36 correlations through moderation. In a separate norming study (reported in Section A in the online supplemental materials), we had participants rate each of these 36 behaviors on 23 differentiating characteristics. We then used the mean facet ratings for each behavior to predict the measures of association

Table 4
Study 1 Test–Retest Reliabilities and Correlations Between Predictors (Each Predictor Averaged Across two Waves)

	Time preference	Age	Parent education	Gender (1 = Male)	Source (1 = MTurk)	Extraversion	Conscientiousness	Openness	Agreeableness	Neuroticism	Barratt Impulsiveness Scale	Financial literacy	Numeracy-CRT	Tightwad-spendthrift	Propensity to plan	Risk preference
Test–retest reliability	0.70***	0.99***	0.94***	1.00***	1.00***	0.90***	0.86***	0.88***	0.87***	0.89***	0.83***	0.82***	0.81***	0.80***	0.75***	0.37***
Age	0.09**															
Parent Education	0.10***	−0.29***														
Gender (1 = Male)	0.07**	0.02	0.02													
Source (1 = MTurk)	−0.03	−0.44***	0.13***	0.13***												
Extraversion	−0.05	0.14***	−0.02	−0.06*	−0.27***											
Conscientiousness	0.04	0.25***	−0.10***	−0.04	−0.17***	0.31***										
Openness	−0.02	−0.07*	0.10***	0.04	0.08**	0.31***	0.19***									
Agreeableness	−0.02	0.20***	−0.09**	−0.11***	−0.19***	0.32***	0.45***	0.12***								
Neuroticism	−0.06*	−0.16***	0	−0.20***	0.03	−0.39***	−0.50***	−0.16***	−0.48***							
Barratt Impulsiveness Scale	−0.16***	−0.12***	−0.03	−0.08**	−0.01	−0.16***	−0.68***	−0.21***	−0.35***	0.49***						
Financial Literacy	0.25***	0.17***	0.15***	0.30***	0.21***	−0.05	0.05	0.05*	−0.12***	−0.13***	−0.21***					
Numeracy-CRT	0.21***	−0.19***	0.24***	0.30***	0.39***	−0.17***	−0.13***	0.06*	−0.22***	−0.03	−0.09**	0.57***				
Tightwad-Spendthrift	−0.20***	0	−0.02	−0.09**	−0.16***	0.18***	−0.17***	0	0	0.10***	0.38***	−0.14***	−0.14***			
Propensity to Plan	0.02	−0.04	0.01	−0.06*	0.02	0.19***	0.32***	0.22***	0.21***	−0.18***	−0.37***	−0.01	−0.08**	−0.19***		
Risk Preference	−0.07*	0	0.04	0.07**	−0.05	0.08**	0.01	0.04	0.05	−0.06*	0.01	−0.12***	−0.14***	0.08**	0.02	

* $p < .05$. ** $p < .01$. *** $p < .001$.

presented in Figure 2. Although this initial analysis yielded a few significant results, indicating *some* ability to predict what characteristics of behaviors made them more associated with time preference, further analyses of these relationships, unfortunately, reveal moderations that are difficult to interpret theoretically.

Thus, the predictive validity of time preference varies widely across the 36 behaviors in ways that we did not predict and, we suspect, might be difficult to predict. The estimates between time preference and behaviors were somewhat larger on average for the financial domain, than for the other two domains. This domain difference may be attributable to the fact that our measure of time preference, like most measures in the field, involves monetary outcomes. Aggregating behaviors into indices based on domains improved time preference's predictive validity, replicating previous research (Bradford et al., 2017; Chabris et al., 2008). That said, the intercorrelations of behaviors within a domain are small, suggesting that categorizing these behaviors by domains might not be justifiable. So, on our second benchmark, there is a lot of heterogeneity in correlation: It seems like time preference is a reasonably strong predictor of a few of the behaviors, but the reasons why it predicts well in some cases and not others have proven elusive.

Benchmark 3: Comparison of Time Preference with Other Covariates

We next assessed how well time preference and 15 covariates each predicted behaviors in two ways: (a) We counted how many times each variable was significant (at $p < .05$ level) across the 36 regressions reported in column 7 of Table 3 and (b) We calculated the median absolute standardized betas across these regressions (see Table S5 in the online supplemental materials). By both metrics, time preference ranks near the middle of our 16 predictor variables in terms of predictive validity. Table S5 in the online supplemental materials reports results from different multiple regression specifications for robustness, with similar results.

Note that we sampled these behaviors because they entail clear delayed consequences. So, we expected time preference to outperform other predictors on theoretical grounds (in aggregate), but its performance places it in the middle of the pack of predictors. Based on the number of times each variable was a significant predictor of behavior, time preference was outperformed by age, parent education, extraversion, and Barratt Impulsiveness scale. Based on median absolute standardized coefficients, age, extraversion, conscientiousness, Barratt Impulsiveness Scale, financial literacy, and numeracy/CRT were more strongly associated with the 36 behaviors than time preference. If our only goal was to predict these 36 behaviors, we would have been better served to measure Barratt Impulsiveness, Extraversion, or age rather than to measure time preference.

Discussion

Based on Study 1's findings, we set out to examine whether the magnitudes of these correlations between time preference and these 36 behaviors were, in fact, predictable. In particular, we sought to elicit expert intuitions about these relationships because: (a) some of the correlations were smaller than we expected and (b) a clear explanation for the large heterogeneity in time preference's predictive validity across behaviors proved elusive to us. To that end, we ask researchers with expertise on time preference to forecast the size of these correlations in Study 2. Obtaining expert's intuitions about these associations also allow us to compare the observed correlations to a fourth benchmark which might be more reasonable, for some purposes, than a null hypothesis of zero correlation.

Study 2

For our fourth benchmark, we asked a group of experts to forecast the 36 correlations between time preference and self-reported behaviors from Study 1. Our motivation behind eliciting expert predictions is as follows: Many papers on time preference (including those we have authored) operate on the assumption that time preference is likely implicated in behaviors with delayed consequences and should therefore predict such behaviors. If we find that experts forecast these correlations to be larger than we found or poorly predict the relative magnitudes of these correlations across behaviors, we will have uncovered gaps in our understanding. Obtaining expert forecasts also helps us calibrate on whether our own surprise at many of the small correlations observed in Study 1 was idiosyncratic. More broadly, Study 2 was intended to capture expert intuitions about the relationship between time preference and behaviors, as these researchers have been exposed to a wealth of relevant data. In addition, expert intuitions guide which research questions and domains receive attention in time preference research.

Method

Transparency and Openness

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. All data are publicly available (see author's note). Data were analyzed using *R* Version 3.6.2 (R Core Team, 2019) and the following packages: *tidyverse* Version 1.3.0 (Wickham et al., 2019); *dplyr* Version 1.0.7 (Wickham et al., 2021); and *data.table* Version 1.14.2 (Dowle & Srinivasan, 2021). This study's design and analyses were not preregistered.

Participants

We aimed to recruit at least 40 academic researchers with expertise on time preference. We initially sent email invitations to 46 experts and then used snowball sampling (common in anthropology, criminology, and other fields), where we asked those experts to nominate other experts. We invited these additional 68 nominated experts in the same fashion as the original 46. Each expert received up to two reminders, at approximately 10 days and 20 days after the initial invitation. We stopped data collection after 90 days, when completion rates had slowed to a trickle, resulting in a total of 55 complete responses. Each expert received $100 for their participation. As an incentive for accuracy, we also donated $500 to the chosen charity of the expert who provided the best estimates (using the scoring criteria outlined below).

Procedure

After consenting to participate, these experts forecasted a series of correlations—between our measure of time preference and responses on each of 36 self-reported behaviors. Assessing the association between variables by directly asking experts for correlations is simpler, more accurate, and more consistent than asking for other measures of association (Clemen et al., 2000). We informed them that we ran a survey assessing these correlations and presented demographic information. We introduced them to the measure of time preference and explained that we scored it by simply counting the number of larger, later choices. We then presented them with a list of all 36 behaviors. We also reminded them of the scoring criteria for determining the forecasting competition winner (which had been specified to them in the invitation email). We would rank participants on two criteria: (a) the correlation between their predictions and the observed correlations in Study 1 in descending order and (b) the mean absolute deviation (MAD) between their predictions and the actual correlations in ascending order. The winner would be the participant with the smallest sum of these two ranks.

For the main task, experts sequentially viewed the original wording and possible responses for each of the 36 behavior questions in Study 1 in a randomized order. We also informed them of how each response was scored, whether the scores were log-scaled, and whether and how multiple questions were aggregated into a single measure. We asked experts to provide their best estimate of the correlation between our time preference measure and each of these 36 questions using a slider ranging from −1 to 1 (with 0.01 precision). On each screen, we reminded them that a higher score on the measure means larger, later choices. Next, they answered questions about themselves: department(s) they held appointments in, knowledge of topics related to time preference, number of years spent studying related topics, and number of projects on related topics. We then asked, "How much confidence do you have in the estimates you made in this survey?" on a 5-point scale (1 = *none*, 5 = *a lot*). Finally, we requested suggestions for other experts in time preference who would be good candidates for participating in the study.

Consensus Analysis

To ascertain whether there was enough consensus among experts to analyze their data in terms of one aggregate group (vs. separate subgroups of experts), we performed a cultural consensus analysis (Romney et al., 1987). To do so, we ran a principal component analysis (PCA) across experts, using each expert's 36 forecasts. The PCA produced first and second eigenvalues of 15.58 and 4.48, with a ratio of 3.48, which exceeds the threshold of three recommended to establish consensus (Weller, 2007). Hence, we proceeded with our aggregate analyses.

Results

Benchmark 4: Comparison to Expert Predictions

The 55 experts held academic appointments in departments including marketing (21), psychology (15), economics (11), decision science (eight), organizational behavior/management (four), behavioral science (two), and a few others (four). More than half (29) have studied intertemporal choice and related topics for over 10 years and all but one have studied it for at least 3–5 years. The median expert had 3–5 projects (published and unpublished) in the area, with 12 reporting 5–10 projects and 12 reporting more than 10 projects. Despite their experience in the topic, nearly all experts reported "some" (26) or "a little" (25) confidence in their

Table 5
Expert Predictions Versus Observed Correlations in Study 1

Domain	#	Behavior	Mean Prediction	Median Prediction	Observed Correlation	Mean Difference	t-statistic	p-value	Mean |Difference|(MAD)
Financial	1	Pay credit card balance in full	0.24	0.20	0.31	−0.07	**−2.59***	0.012	0.17
	2	Accumulate wealth	0.19	0.16	0.24	−0.04	−1.62	0.110	0.16
	3	Save a high percentage of income	0.23	0.21	0.15	0.07	**2.05***	0.045	0.18
	4	NOT miss credit card payments	0.17	0.18	0.15	0.02	0.87	0.389	0.15
	5	Earn a large income	0.22	0.20	0.14	0.08	**2.41***	0.020	0.17
	6	NOT accumulate a lot of credit card debt	0.13	0.15	0.11	0.02	0.54	0.594	0.20
	7	NOT accumulate a lot of educational loan debt	0.01	0	0.10	−0.09	**−3.19****	0.250	0.19
	8	Use coupons or rebate offers	0.01	0.02	0.04	−0.03	−1.16	0.717	0.13
	9	Take out a mortgage	0.04	0.02	0.03	0.01	0.36	0.003	0.14
	10	Pay enough taxes to get refund	0.11	0.07	0.02	0.09	**3.08****	0.002	0.14
Health	11	Get dental cleanings	0.11	0.1	0.17	−0.06	**−2.74****	0.008	0.13
	12	NOT consume nicotine	0.16	0.15	0.12	0.04	1.40	0.168	0.16
	13	Floss	0.13	0.1	0.11	0.01	0.69	0.496	0.09
	14	NOT gain excess weight	0.07	0.08	0.09	−0.02	−0.79	0.435	0.12
	15	Use sunscreen	0.11	0.09	0.08	0.03	1.88	0.066	0.1
	16	Monitor the nutritional content of food	0.12	0.1	0.08	0.04	1.94	0.057	0.12
	17	Get routine checkups (physicals)	0.12	0.09	0.07	0.05	**2.37***	0.021	0.11
	18	Follow a diet plan	−0.04	−0.02	0.06	−0.10	**−5.84*****	<0.001	0.12
	19	Follow doctor's prescriptions	0.14	0.1	0.05	0.08	**4.26*****	<0.001	0.11
	20	NOT use recreational drugs	0.12	0.14	0.05	0.07	**2.93****	0.005	0.15
	21	NOT drink coffee	−0.02	0	0.04	−0.06	**−4.53*****	<0.001	0.08
	22	NOT overeat	0.08	0.08	0	0.08	**4.21*****	<0.001	0.12
	23	NOT drink alcoholic beverages	0.09	0.09	0	0.09	**4.12*****	<0.001	0.14
	24	Actively exercise	0.14	0.11	−0.01	0.15	**8.33*****	<0.001	0.17
	25	Keep physically active	0.08	0.08	−0.03	0.11	**6.88*****	<0.001	0.14
Personal prudence	26	Attain a high level of education	0.27	0.3	0.16	0.11	**3.47****	0.001	0.20
	27	NOT get tattoos	0.12	0.1	0.14	−0.02	−0.80	0.427	0.13
	28	Go to bed at the same time every night	0.09	0.06	0.12	−0.03	**−2.28***	0.026	0.09
	29	Have kids when older	0.14	0.10	0.10	0.04	1.48	0.145	0.15
	30	NOT gamble or buy lottery tickets	0.12	0.10	0.08	0.03	1.10	0.278	0.16
	31	NOT leave dirty dishes overnight	0.07	0.04	0.05	0.02	0.96	0.340	0.10
	32	Arrive on time for meetings	0.10	0.10	0.03	0.07	**2.91****	0.005	0.14
	33	Start tasks and assignments early	0.12	0.11	0.02	0.10	**4.57*****	<0.001	0.15
	34	NOT get speeding tickets	0.07	0.05	0	0.07	**3.47****	0.001	0.12
	35	Get married when older	0.08	0.05	0	0.08	**3.40****	0.001	0.13
	36	NOT drive recklessly	0.08	0.07	−0.03	0.11	**6.00*****	<0.001	0.14
		Overall mean	0.11	0.10	0.08	0.03	1.39	0.170	0.14

Note. MAD = mean absolute deviation. Bold values indicate $p < .05$.
* $p < .05$. ** $p < .01$. *** $p < .001$.

estimates, while four reported "a fair amount" and none reported "a lot" of confidence.

Aggregated Experts

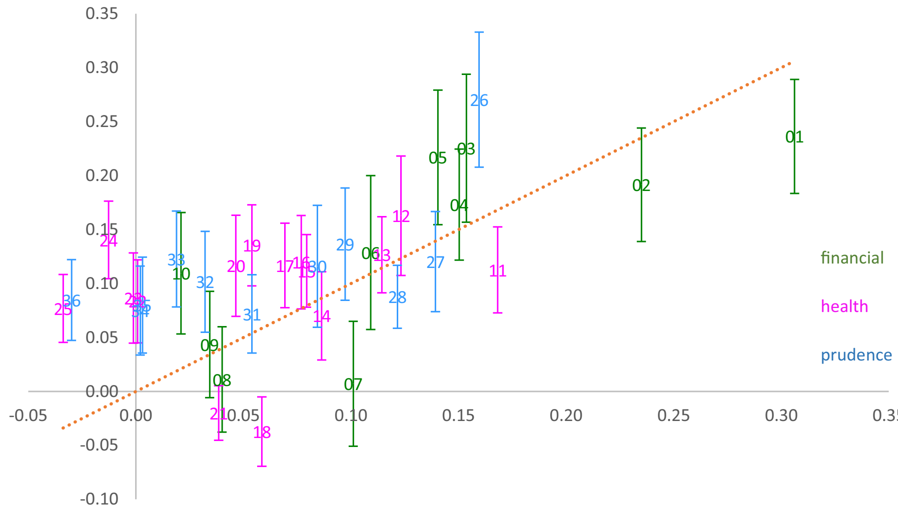
Table 5 presents summary statistics for the expert predictions, their differences from the observed correlations in Study 1, and tests comparing expert predictions with the observed correlations. (Table S6 in the online supplemental materials reports analyses using the disattenuated correlations, with similar results.) The average expert prediction for the correlation between time preference and the 36 behaviors in our study was 0.11 (range from −0.04 to 0.27), although the observed average correlation was even smaller (0.08, range from −0.03 to 0.31).

In general, experts overestimated the correlations between time preference and the behaviors. The average expert forecast significantly (at the $p < .05$ level) overestimated the correlations between time preference and behaviors for 16 of the behaviors and underestimated for 6. The average forecast was not significantly different from the observed correlations for the remaining 14 behaviors (although 10 of those were directionally overestimated). Figure 3 plots the relationship between the average forecasts and observed correlations. And at the individual level, 19 of 55 experts made forecasts that on average significantly overestimated the correlations, while 11 significantly underestimated.[7]

[7] Based on an anonymous reviewer's suggestion, we also examined whether there were any differences in the predictive accuracy of experts across the three different domains of behavior. The degree of correlation between the average expert forecasts and observed correlations was indeed highest for the 10 financial behaviors ($r = 0.75$), was reasonably high for the 11 prudential behaviors ($r = 0.64$), but was much lower for the 15 health behaviors ($r = 0.23$). However, due to the small sample sizes, no differences were significant ($p = .12$ for the difference between financial and health behaviors). In terms of mean absolute differences between predicted and observed correlations (MAD), the experts did not make better forecasts for the financial-domain behaviors than for other domains ($MAD_{financial} = 0.16$, $MAD_{health} = 0.12$, $MAD_{prudential} = 0.14$).

Figure 3

Scatterplot of Expert Forecasts (on the y*-Axis) Versus the Correlations Observed in Study 1 (on the* x*-Axis)*



Note. Each point corresponds to one of 36 behaviors (numbers for each behavior refer to the order in which they are listed in Table 5) and error bars represent 95% confidence intervals. The 95% confidence intervals that fall above (below) the dashed line indicated that experts forecasted a higher (lower) degree of positive correlation than was observed in Study 1. Financial behaviors are presented in green, health behaviors in pink, and personal prudence behaviors in blue.

Along with some evidence for overprediction, we also found that the aggregate predictions were well-calibrated—the aggregated experts can predict which correlations are larger and smaller. Figure 3 depicts the high overall degree of correlation between the average expert forecasts and observed correlations across the 36 behaviors, $r = 0.60$, $t(34) = 4.36$, $p < .001$. Results using the median expert forecasts were similar ($r = 0.58$). However, it is important to note that the fact that average predictions were well-calibrated benefits from averaging over experts who did not always agree with each other (i.e., wisdom of the crowds).

Individual Experts

We next consider expert forecasts at the individual level. Figure 4 presents the distribution of correlations between each expert's forecasts and the observed correlations. The experts were mostly positively correlated between their forecasts and the observed correlations in Study 1 ($M = 0.28$, $SD = 0.22$, range $= -0.42$ to 0.69).[8] Twenty-two experts (40%) made forecasts that were significantly correlated with the observed correlations (threshold for significance at $r = 0.33$). Consistent with wisdom of the crowds, only one individual expert made predictions that were more accurate than the aggregated expert predictions.

We also examined the MAD between each expert's forecasts and the observed correlations (see Figure 5 and the last column of Table 1). The average MAD was 0.14 ($SD = 0.08$), which we interpret as small in an absolute sense but large in a relative sense. For comparison, 29 of the 36 behaviors had absolute observed correlations smaller than 0.14.

Finally, we ran a simple regression predicting expert performance as a function of their background and expertise, but no variable (including department, self-assessed knowledge, years studying time preference, number of projects, and forecast confidence) predicted either the correlation between their forecasts and observed forecasts or their MAD.

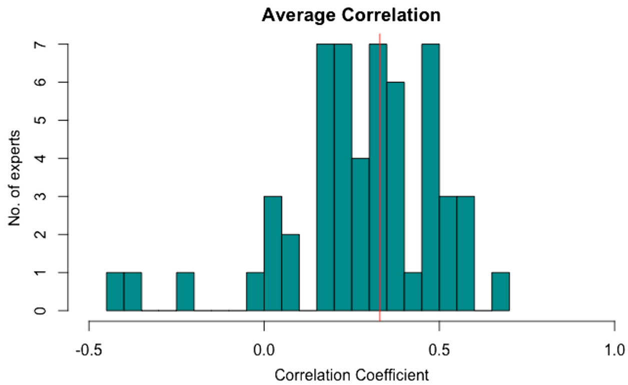
Discussion

In Study 2, we aimed to assess the predictive validity of time preference against our fourth benchmark: expert forecasts. We asked whether experts could predict the large heterogeneity in the association between time preference and behaviors that we observed in Study 1. The results paint a mixed picture of whether time preference researchers can determine which real-world intertemporal choice behaviors can be predicted using time preference. On the one hand, the average of experts' predictions tracked quite well which behaviors were more and less strongly correlated with time preference. Furthermore, 22 of 55 experts' predictions were significantly correlated with the observed correlations in Study 1 at $r > 0.33$.

On the other hand, the average expert significantly overpredicted nearly half of the correlations, suggesting that time preference researchers often believe that time preference is more predictive than we found in Study 1. Experts expected time preference to predict behaviors more consistently than it actually did, missing many of the behaviors that were not predicted well. Overpredictions tended to occur for behaviors that were not highly correlated with time preference—the 10 lowest observed correlations were all significantly overpredicted.

[8] Upon examination, we found that two of our 55 experts reported negative correlations for over half the behaviors. This raises the possibility that they may have miscomprehended the direction of our time preference measure. If we exclude these two experts, we find an average correlation of 0.30 between expert forecasts and absolute correlations, and a MAD of 0.13.

Figure 4
Expert Study Distribution of Correlations Between Forecasted Estimates and Observed Estimates for the 55 Experts



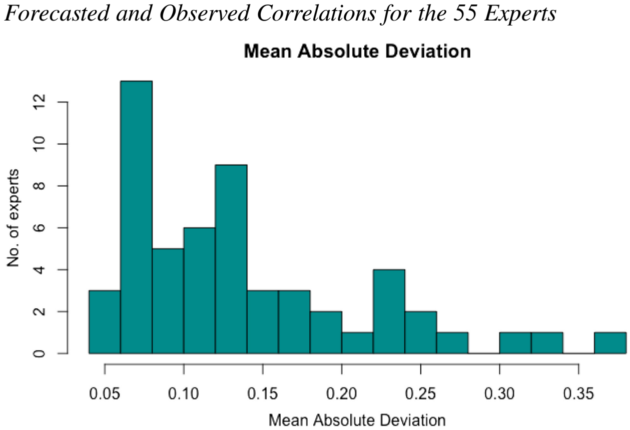
Note. Correlations to the right of the line are experts whose forecasts are significantly correlated with observed estimates. See the online article for the color version of this figure.

Despite their tendency to overpredict, our experts were not very optimistic about time preference's predictive validity, with an average predicted correlation of only 0.11 (which made the degree of overprediction small, on average). So, experts expressed a lack of optimism regarding time preference's predictive validity along with an explicit lack of confidence in their forecasts. Some of this pessimism could be due to unfamiliarity with our novel time preference measure.

Our results suggest that time preference researchers have more to learn about whether, when, and why time preference is relevant for a given behavior. That is, for a randomly selected behavior, we are not especially likely to know how well time preference will relate to that behavior. We, therefore, suggest exercising caution in generalizing our findings to other domains on the assumption that they will generalize without actually studying those domains.

On the whole, our findings suggest that: (a) Experts are good, in the aggregate, and most are individually, at predicting the relative size of these relationships. (b) When experts err, their forecasts are more often overpredictions—predicting a higher degree of positive association between time preference and behavior than the observed correlation. (c) There was no detectable relationship between variables that might indicate level of expertise and good forecasting (as in DellaVigna & Pope, 2018). We also did not find a strong relationship between accuracy and confidence (as in Tsai et al., 2008).

Figure 5
Study 2 Distribution of the Mean Absolute Deviations Between Forecasted and Observed Correlations for the 55 Experts



Note. See the online article for the color version of this figure.

General Discussion

Prior literature has implicated time preference in a wide range of behaviors of interest to social scientists, from the mundane to some of the most important decisions people make. In this paper, we present the most comprehensive examination to date of how well laboratory-derived estimates of time preference predict self-reported real-world behaviors. We approached this investigation by comparing time preference to four benchmarks: (a) zero, (b) across behaviors, (c) against other predictors, and (d) expert forecasts. We find that how promising time preference is as a predictor depends on which of our four benchmarks we use.

Study 1 used the first three benchmarks. On the first benchmark, time preference is a significant predictor for about half the behaviors, even when controlling for 15 other relevant variables. However, we sampled these 36 behaviors precisely because we expected them to be related to time preference on theoretical grounds. So, although it predicts many of the behaviors, some may find it surprising that it did not predict more of these behaviors. Such surprise may be especially warranted considering that a correlation as small as 0.06 is significant in our study because of our large sample.

On the second benchmark, there was considerable heterogeneity in the association between time preference and the behaviors that was neither accounted for by domain (e.g., financial vs. health) nor easily explained by ratings of the behaviors on their psychological facets (see Section A in the online supplemental materials). It is possible that people do not possess one singular time preference. For example, Chapman (1996) found that time preferences for health and time were not correlated. Also, some have suggested that other important preferences, like risk sensitivity and loss aversion (Lejarraga & Hertwig, 2021; Stephens, 1981) may be dynamic and state-dependent, varying as a function of an organism's environment and metabolic needs. Time preferences might similarly differ within an individual across contexts (Krefeld-Schwalb et al., 2023) and mental states, and this intra-individual variability could undermine the predictive validity of time preference. Whatever the cause of the unexplainable heterogeneity of the predictive validity of time preference that we observe across behaviors, we think this reveals gaps in our knowledge of the moderators of these relationships. Time preference appears to be relevant for some behaviors and not very relevant for a host of others, and we do not know why.

On the third benchmark, time preference outperformed a number of relevant variables but also consistently underperformed compared to some others. On the one hand, we included relevant, widely used variables from personality psychology and financial decision-making, and so it is promising that time preference performed better than many of them. On the other hand, given these behaviors were specifically chosen to be related to time preference, one might expect time preference to perform better than most if not all other variables.

Study 2 introduced the final benchmark. At the individual level, and even more so at the group level, expert forecasts broadly aligned with observed correlations. However, experts' forecasted correlations were overall fairly small and offered with muted confidence; despite that, their forecasts systematically overestimated the degree of positive association between time preference and the behaviors. There was no detectable relationship between variables that might indicate level of expertise and more accurate forecasting (as in DellaVigna & Pope, 2018).

Implications

So, what should we conclude about how well measures of time preference predict behaviors? One's answer likely depends on (a) prior beliefs about the importance of time preference, (b) implementational concerns, and (c) whether one thinks studies like ours can meaningfully address how well time preference predicts behavior.

First, do we start with the assumption that time preference is an important and/or useful predictor of behavior? Most of us who do research on intertemporal choice believe that it is an important research domain, for many reasons. Hence, we are inclined to look for evidence of prediction "successes" as support for the importance of the research endeavor. Others likely start with different expectations and reach different conclusions of how good or bad our mixed bag of evidence is for the general predictive validity of time preference. Regardless of one's expectations, the associations we found between time preference and our wide range of behaviors were mostly small in magnitude and notably smaller than our experts' already moderate predictions.

Second, why are time preference's relationships with behavior important and what other predictors are available? Predicting behavior is an important endeavor, and time preference has been associated with decisions of major import, like mortgage choice (Atlas et al., 2017), retirement savings decisions (Angeletos et al., 2001), and smoking (e.g., Reynolds et al., 2004). However, we found several covariates that were better predictors of the behaviors we measured, even though we specifically chose behaviors we expected to be related to time preference. At the same time, time preference presumably outperforms many more covariates that we did not measure (and some we did, like height—measured for BMI—which was a worse predictor for 29 of the 36 behaviors). And measures of time preference are more easily administered than measures of many other predictors—they can be collected remotely, using a few simple questions, and are fairly reliable. We are certainly not advocating that people stop collecting data on time preference as one of a set of predictors of important behaviors.

Finally, can studies like ours meaningfully address how well time preference predicts behavior? A handful of caveats apply to our investigation and to the majority of the literature we cited as precedent for our investigation. First, our data are correlational, with the usual caveats about inferring causality (e.g., reverse causation and omitted variables). Following prior literature and for ease of communication, we use variants of the word "predict" to describe the relationship between time preference and behaviors. However, the coefficients we reported are degrees of association; we do not (and should not) make strong causal claims about these relationships. Second, our measures were self-reports, which raises the usual concerns about how memory errors and social desirability influence participant responses. Third, our measure of time preference was not incentive-compatible, which introduces the possibility that participants might not have responded according to their true time preference (Harrison & Rutström, 2008). Then again, past research suggests that incentivized elicitation measures do not generally differ significantly from non-incentivized measures (Camerer & Hogarth, 1999; Starmer & Sugden, 1991). Fourth, following the prior literature, we sampled our list of behaviors and covariates to have the most relevance for the current investigation. In particular, the behaviors we chose all have delayed consequences, and this method of stimulus sampling influenced our results. However, our stimulus sampling was theoretically motivated: We curated our list of behaviors either by adapting items from previous studies (Chabris et al., 2008; Reimers et al., 2009), or created new items such that they varied along 23 potentially informative characteristics (see Section A in the online supplemental materials).

With those caveats in mind, our goal was to provide the most comprehensive investigation to date of how well laboratory-derived measures of time preference predict real-world behaviors. We did so by sampling more behaviors, including more relevant variables as covariates, employing a test–retest design with a large sample. We also elicited forecasts of the size of these correlations by the group of people most well-positioned to make these predictions —researchers who think about and publish on the topics of intertemporal choice and time preference. We think the predictions of these experts are important as a benchmark against which the observed correlations can be compared. Perhaps more importantly, it is the predictions of experts like these that determine where we look for evidence on the role of time preference in behavior. We think it is telling that nearly all of our experts expressed low levels of confidence in their predictions. We hope that the mixed bag of results produced by our comprehensive investigation of the topic help guide whether and where these and other researchers use time preference as a metaphor for and/or predictor of behaviors in future research.

References

Angeletos, G. M., Laibson, D., Repetto, A., Tobacman, J., & Weinberg, S. (2001). The hyperbolic consumption model: Calibration, simulation, and empirical evaluation. *Journal of Economic Perspectives*, *15*(3), 47–68. https://doi.org/10.1257/jep.15.3.47

Atlas, S. A., Johnson, E. J., & Payne, J. W. (2017). Time preferences and mortgage choice. *Journal of Marketing Research*, *54*(3), 415–429. https://doi.org/10.1509/jmr.14.0481

Barlow, P., Reeves, A., McKee, M., Galea, G., & Stuckler, D. (2016). Unhealthy diets, obesity and time discounting: A systematic literature review and network analysis. *Obesity Reviews*, *17*(9), 810–819. https://doi.org/10.1111/obr.12431

Bradford, W. D. (2010). The association between individual time preferences and health maintenance habits. *Medical Decision Making*, *30*(1), 99–112. https://doi.org/10.1177/0272989X09342276

Bradford, W. D., Courtemanche, C., Heutel, G., McAlvanah, P., & Ruhm, C. (2017). Time preferences and consumer behavior. *Journal of Risk and Uncertainty*, *55*(2–3), 119–145. https://doi.org/10.1007/s11166-018-9272-8

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*(1–3), 7–42. https://doi.org/10.1023/A:1007850605129

Chabris, C. F., Laibson, D., Morris, C. L., Schuldt, J. P., & Taubinsky, D. (2008). Individual laboratory-measured discount rates predict field behavior. *Journal of Risk and Uncertainty*, *37*(2–3), 237–269. https://doi.org/10.1007/s11166-008-9053-x

Chapman, G. B. (1996). Temporal discounting and utility for health and money. *Journal of Experimental Psychology: Learning, Memory,* *and Cognition*, *22*(3), 771–791. https://doi.org/10.1037/0278-7393.22.3.771

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, *11*(4), 464–473. https://doi.org/10.1177/1948550619875149

Clemen, R. T., Fischer, G. W., & Winkler, R. L. (2000). Assessing dependence: Some experimental results. *Management Science*, *46*(8), 1100–1115. https://doi.org/10.1287/mnsc.46.8.1100.12023

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

DellaVigna, S., & Pope, D. (2018). What motivates effort? Evidence and expert forecasts. *The Review of Economic Studies*, *85*(2), 1029–1069. https://doi.org/10.1093/restud/rdx033

Dixon, M. R., Marley, J., & Jacobs, E. A. (2003). Delay discounting by pathological gamblers. *Journal of Applied Behavior Analysis*, *36*(4), 449–458. https://doi.org/10.1901/jaba.2003.36-449

Dowle, M., & Srinivasan, A. (2021). *data.table: Extension of "data.frame"* (Version 1.14.2). https://cran.r-project.org/package=data.table

Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, *16*(12), 939–944. https://doi.org/10.1111/j.1467-9280.2005.01641.x

Eckel, C. C., & Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, *23*(4), 281–295. https://doi.org/10.1016/S1090-5138(02)00097-1

Ericson, K. M., & Laibson, D. (2019). Intertemporal choice. In B. D. Bernheim, S. DellaVigna, & D. Laibson (Eds.). *Handbook of behavioral economics: Applications and foundations 1* (Vol. 2, pp. 1–67). North-Holland.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, *133*(4), 1645–1692. https://doi.org/10.1093/qje/qjy013

Farrell, P., & Fuchs, V. R. (1982). Schooling and health: The cigarette connection. *Journal of Health Economics*, *1*(3), 217–230. https://doi.org/10.1016/0167-6296(82)90001-7

Fernandes, D., Lynch, J. G., & Netemeyer, R. G. (2014). Financial literacy, financial education, and downstream financial behaviors. *Management Science*, *60*(8), 1861–1883. https://doi.org/10.1287/mnsc.2013.1849

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, *40*(2), 351–401. https://doi.org/10.1257/jel.40.2.351

Gillen, B., Snowberg, E., & Yariv, L. (2019). Experimenting with measurement error: Techniques with applications to the Caltech cohort study. *Journal of Political Economy*, *127*(4), 1826–1863. https://doi.org/10.1086/701681

Harrell, F. E. J. (2019). *Hmisc: Harrell miscellaneous* (Version 4.3-0). https://cran.r-project.org/package=Hmisc

Harrison, G. W., & Rutström, E. E. (2008). Risk aversion in the laboratory. *Research in Experimental Economics*, *12*(8), 41–196. https://doi.org/10.1016/S0193-2306(08)00003-3

Hughes, J. (2020). *reghelper: Helper functions for regression analysis* (Version 0.3.6). https://cran.r-project.org/package=reghelper

John, O. P., & Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin, & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102–138). Guilford Press.

Krefeld-Schwalb, A., Bartels, D. M., & Johnson, E. J. (2023). *More than money over time: Consumers' intertemporal preferences for money, health and getting things done* [Manuscript submitted for publication]. Rotterdam School of Management, Erasmus University.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, *112*(2), 443–478. https://doi.org/10.1162/003355397555253

Lejarraga, T., & Hertwig, R. (2021). Three theories of choice and their psychology of losses. *Perspectives on Psychological Science*, *17*(2), 334–345. https://doi.org/10.1177/17456916211001332

Li, Y., Baldassi, M., Johnson, E. J., & Weber, E. U. (2013). Complementary cognitive capabilities, economic decision making, and aging. *Psychology and Aging*, *28*(3), 595–613. https://doi.org/10.1037/a0034172

Li, Y., Gao, J., Enkavi, A. Z., Zaval, L., Weber, E. U., & Johnson, E. J. (2015). Sound credit scores and financial decisions despite cognitive aging. *Proceedings of the National Academy of Sciences*, *112*(1), 65–69. https://doi.org/10.1073/pnas.1413570112

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*(1), 37–44. https://doi.org/10.1177/0272989X0102100105

Lynch, J. G. J., Netemeyer, R. G., Spiller, S. A., & Zammit, A. (2010). A generalizable scale of propensity to plan: The long and the short of planning for time and for money. *Journal of Consumer Research*, *37*(1), 108–128. https://doi.org/10.1086/649907

MacKillop, J., Amlung, M. T., Few, L. R., Ray, L. A., Sweet, L. H., & Munafò, M. R. (2011). Delayed reward discounting and addictive behavior: A meta-analysis. *Psychopharmacology*, *216*(3), 305–321. https://doi.org/10.1007/s00213-011-2229-0

Meier, S., & Sprenger, C. D. (2012). Time discounting predicts creditworthiness. *Psychological Science*, *23*(1), 56–58. https://doi.org/10.1177/0956797611425931

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1996). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology*, *51*(6), 768–774. https://doi.org/10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1

Petry, N. M. (2001). Pathological gamblers, with and without substance abuse disorders, discount delayed rewards at high rates. *Journal of Abnormal Psychology*, *110*(3), 482–487. https://doi.org/10.1037/0021-843X.110.3.482

R Core Team. (2019). *R: A language and environment for statistical computing* (Version 3.6.2). R Foundation for Statistical Computing. https://www.r-project.org/

Read, D., & Read, N. L. (2004). Time discounting over the lifespan. *Organizational Behavior and Human Decision Processes*, *94*(1), 22–32. https://doi.org/10.1016/j.obhdp.2004.01.002

Reed, D. D., & Martens, B. K. (2011). Temporal discounting predicts student responsiveness to exchange delays in a classroom token system. *Journal of Applied Behavior Analysis*, *44*(1), 1–18. https://doi.org/10.1901/jaba.2011.44-1

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*(8), 973–978. https://doi.org/10.1016/j.paid.2009.07.026

Reynolds, B. (2006). A review of delay-discounting research with humans: Relations to drug use and gambling. *Behavioural Pharmacology*, *17*(8), 651–667. https://doi.org/10.1097/FBP.0b013e3280115f99

Reynolds, B., Richards, J. B., Horn, K., & Karraker, K. (2004). Delay discounting and probability discounting as related to cigarette smoking status in adults. *Behavioural Processes*, *65*(1), 35–42. https://doi.org/10.1016/S0376-6357(03)00109-8

Rick, S. I., Cryder, C. E., & Loewenstein, G. (2008). Tightwads and spendthrifts. *Journal of Consumer Research*, *34*(6), 767–782. https://doi.org/10.1086/523285

Romney, A. K., Batchelder, W. H., & Weller, S. C. (1987). Recent applications of cultural consensus theory. *American Behavioral Scientist*, *31*(2), 163–177. https://doi.org/10.1177/000276487031002003

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021). *afex: Analysis of factorial experiments* (Version 1.0.1). https://cran.r-project.org/package=afex

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72–101. https://doi.org/10.2307/1412159

Starmer, C., & Sugden, R. (1991). Does the random-lottery incentive system elicit true preferences? An experimental investigation. *The American Economic Review*, *81*(4), 971–978.

Stephens, D. W. (1981). The logic of risk-sensitive foraging preferences. *Animal Behaviour*, *29*(2), 628–629. https://doi.org/10.1016/S0003-3472(81)80128-5

Sutter, M., Kocher, M. G., Glätzle-Rützler, D., & Trautmann, S. T. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *American Economic Review*, *103*(1), 510–531. https://doi.org/10.1257/aer.103.1.510

Toubia, O., Johnson, E. J., Evgeniou, T., & Delquié, P. (2013). Dynamic experiments for estimating preferences: An adaptive method of eliciting time and risk parameters. *Management Science*, *59*(3), 613–640. https://doi.org/10.1287/mnsc.1120.1570

Tsai, C. I., Klayman, J., & Hastie, R. (2008). Effects of amount of information on judgment accuracy and confidence. *Organizational Behavior and Human Decision Processes*, *107*(2), 97–105. https://doi.org/10.1016/j.obhdp.2008.01.005

Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2, pp. 131–160). Addison-Wesley.

Urminsky, O., & Zauberman, G. (2015). The psychology of intertemporal preferences. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (Vol. 1, pp. 141–181). John Wiley & Sons. https://doi.org/10.1002/9781118468333.ch5

Vuchinich, R. E., & Simpson, C. A. (1998). Hyperbolic temporal discounting in social drinkers and problem drinkers. *Experimental and Clinical Psychopharmacology*, *6*(3), 292–305. https://doi.org/10.1037/1064-1297.6.3.292

Weller, S. C. (2007). Cultural consensus theory: Applications and frequently asked questions. *Field Methods*, *19*(4), 339–368. https://doi.org/10.1177/1525822X07303502

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), Article 1686. https://doi.org/10.21105/joss.01686

Wickham, H., Romain, F., Lionel, H., & Kirill, M. (2021). *dplyr: A grammar of data manipulation* (Version 1.0.7). https://cran.r-project.org/package=dplyr

Received November 15, 2021
Revision received November 18, 2022
Accepted January 13, 2023 ■