Meta-Analysis of Present-Bias Estimation Using Convex Time Budgets

Taisuke Imai Tom A. Rutter Colin F. Camerer *

April 2, 2020

Abstract

We examine 220 estimates of the present-bias parameter from 28 articles using the Convex Time Budget protocol. The literature shows that people are on average present biased, but the estimates exhibit substantial heterogeneity across studies. There is evidence of modest selective reporting in the direction of overreporting present-bias. The primary source of the heterogeneity is the type of reward, either monetary or non-monetary reward, but the effect is weakened after correcting for potential selective reporting. In the studies using the monetary reward, the delay until the issue of the reward associated with the "current" time period is shown to influence the estimates of present-bias parameter.

JEL Classification codes: D90, C91

Keywords: present bias, time preferences, structural behavioral economics, meta-analysis, selective reporting

*Imai: Department of Economics, LMU Munich, taisuke.imai@econ.lmu.de. Rutter: Department of Economics, LSE, t.a.rutter@lse.ac.uk. Camerer: Division of the Humanities and Social Sciences, California Institute of Technology, camerer@hss.caltech.edu. This is a part of the project "A Large-Scale, Interdisciplinary Meta-Analysis on Behavioral Economics Parameters" supported by the Social Science Meta-Analysis and Research Transparency (SSMART) Grants from Berkeley Initiative for Transparency in the Social Sciences (BITSS). We thank Stefano DellaVigna, Tomáš Havránek, Yves Le Yaouanq, Peter Schwardmann, Charles Sprenger, and Tom Stanley for helpful comments. Imai acknowledges financial support by the Deutsche Forschungsgemeinschaft through CRC TRR 190. Rutter acknowledges the support of the 2016 SURF Fellowship from the California Institute of Technology.

1 Introduction

Most choices create benefits and costs that occur at different points in time. Domains of these intertemporal choices include health (e.g., eating and exercise), financial decision making (e.g., saving for retirement), pursuit of education, household decisions, and more. In many of these domains, introspection and experimental evidence suggest that people often exhibit *present bias*: people prefer a smaller immediate reward to a larger delayed reward in the present, but they reverse their preferences when these two alternatives are shifted to the future by the same amount of time. Understanding how and why people make such present-biased choices in many domains informs design of government policy, corporate practices, and clinical practices.

The exponentially discounted utility model (EDU; Koopmans, 1960; Samuelson, 1937) is the standard model of intertemporal choice in economics. The model assumes that an individual's intertemporal preferences are governed by a parameter δ, called the *discount factor*, and that she attaches the relative weight δ^t to the utility from consumption she receives t periods in the future. The quasi-hyperbolic discounted utility model (QHD; Laibson, 1997; Phelps and Pollak, 1968), also known as the present-biased preferences model, is a one-parameter extension of EDU. It is designed to capture dynamically inconsistent choices while retaining the tractability of EDU. In QHD, an agent (at period 0) values a consumption stream $(x_0, \ldots, x_T)$ according to

$$U(x_0, \ldots, x_T) = u(x_0) + \beta \sum_{t=1}^{T} \delta^t u(x_t), \tag{1}$$

where $\delta > 0$ is a traditional discount factor and $\beta > 0$ captures present-bias. Note that the utilities from "future" periods ($t \geq 1$) are exponentially weighted as in the standard EDU, while this stream of future utilities is also discounted by β. Note also that QHD includes EDU as a special case when $\beta = 1$ (there is no present-bias; time-consistency). QHD is the most widely used representation of present-biased preferences, although other functional forms (particularly variants of hyperbolic discounting) will exhibit present-bias too.[1]

In this paper, we assemble a dataset of empirical estimates of present-biased preferences measured with the experimental method called the Convex Time Budget (CTB; Andreoni and Sprenger, 2012) and meta-analyze those data. The meta-analysis gives tentative answers to three questions. (i) What is an average value of β? (ii) Is there selective reporting or publication bias?

[1] See, for example, DellaVigna and Malmendier (2006), Gruber and Kőszegi (2001), Heidhues and Kőszegi (2010), and O'Donoghue and Rabin (1999, 2001) for applications of (naïve) present-biased preferences and O'Donoghue and Rabin (2015) for a short overview. See Ericson and Laibson (2019) for a broad coverage of models of what they term "present-focused" preferences including, but not restricted to, QHD.

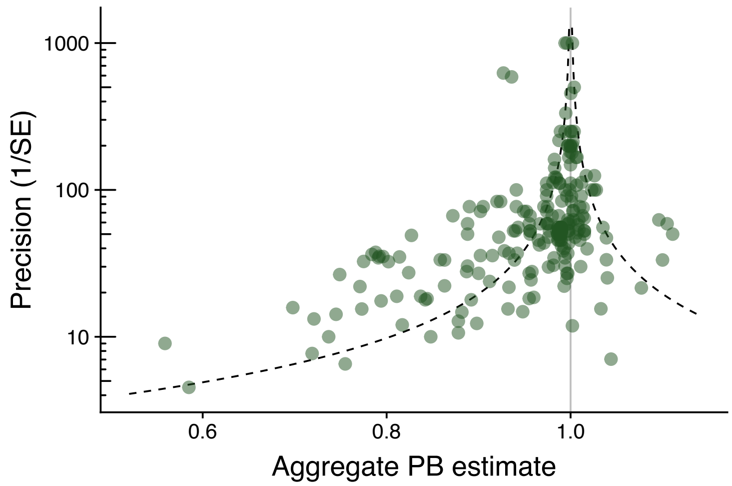


Figure 1: Funnel plot of estimates of present-bias parameter (*PB*). The y-axis (precision; inverse standard error) is presented in the log-scale. The dotted curves indicate the boundaries for rejection of the null hypothesis of no present bias ($PB = 1$; vertical grey line) for a two-sided test at the 5% significance level.

(iii) How does β vary reliably with types of rewards, subject population, estimation methods, etc.?

Our meta-analysis collects 220 estimates of the present-bias parameter in the QHD model (β in equation (1); hereafter *PB*) from 31 studies reported in 28 articles. The distribution of estimates and the relation with their associated standard errors is presented in the "funnel plot" in Figure 1. A significant proportion of estimated *PB*'s are smaller than one, indicating present bias rather than future bias. The dotted curves indicate the boundaries for rejection of the null hypothesis of no present bias ($PB = 1$) for a two-sided test at the 5% significance level; estimates outside the boundaries are rejections. The figure shows that many studies *did not* find strong evidence to reject the null of $PB = 1$, but those that do reject the null hypothesis tend to show present bias ($PB < 1$) rather than future bias ($PB > 1$).

We now provide a preview of our results. We find statistically significant evidence of present bias overall; our meta-analytic average *PB* is between 0.95 and 0.97. However, the reported estimates differ systematically by the type of reward: The values for monetary-reward studies are close to one, indicating the absence of present bias, while studies with non-monetary reward report a lower average *PB* of 0.88. We also find evidence suggesting selective reporting, in the direction of overreporting $PB < 1$ in studies using a non-monetary reward. Within the studies using monetary reward, the delay until the issue of the "current period" ($t = 0$) reward is shown to robustly influence estimated *PB*.

Our contribution is substantive because it presents the best available estimates of *PB*, and how much they vary. This evidence should be useful to many empirical economists for whom a *PB* has been applied, including in household finance (e.g., Angeletos et al., 2001; Beshears et al., 2017; Meier and Sprenger, 2010), health decisions (Fang and Wang, 2015), labor contracts (Bisin and Hyndman, 2020; Kaur et al., 2010, 2015), demand for commitment devices (Ashraf et al., 2006; Beshears et al., 2015; John, forthcoming), and others. It should also be useful for experimentalists who want to understand which aspects of the design might influence their estimates of *PB*.

Meta-analysis presumes that along with conventional "narrative" reviews, it is useful to compile studies using specific inclusion criteria, and compare numbers measured in different studies. It hardly bears mentioning that even in the presence of quantitative meta-analyses, narrative reviews will always be useful. They allow insightful commentary on which studies authors believe are particularly interesting, diagnostic, or deserving of replication and extension, in a way that meta-analysis does not easily permit.

At the same time, narrative reviews do not typically specify inclusion criteria and usually do not compare study results on one or more quantitative metrics. As a result, until a meta-analysis such as ours, it is fair to say that even the most expert scholars are not fully aware of what all existing studies have to say about the numerical size and variation in *PB*. Meta-analysis goes further by compiling accessible cross-study data (which others can re-analyze), establishing central tendency of numerical estimations, exploring cross-study moderators which affect estimates, and testing for various kinds of selective reporting.

Meta-analysis is designed to accumulate scientific knowledge, and also detect nonrandom reporting or publication of estimates that deviate from the average. Since it was first introduced by Glass (1976), meta-analysis has been playing an important role in evidence-based practices in medicine and policy (Gurevitch et al., 2018). However, meta-analysis has been less common in economics until recently (Stanley, 2001).[2] The current study is the first systematic meta-analysis on the structural estimation of present bias in QHD, focusing specifically on empirical approaches based on the CTB protocol. [3] Prominent reviews of evidence about intertemporal choices and *PB* include the classic piece by Frederick et al. (2002) and more recent coverage by Cohen et al. (forthcoming) and Ericson and Laibson (2019). These articles are narrative and do not provide systematic collection and analysis of empirical observations (they rather describe subsets of im-

[2]See a list of relevant publications indexed on RePec at: `https://ideas.repec.org/k/metaana.html`.

[3]In a companion paper, Imai et al. (2018) conduct a large-scale meta-analysis of empirical estimates of discount rates. The dataset covers estimates from both experimental and non-experimental studies in ecnonomics, psychology, neuroscience, medicine, and other fields. Matoušek et al. (2019) conduct a similar meta-analysis of discount rates, not *PB*, focusing on 34 published articles in economics.

portant contributions and themes which emerge across studies).[4]

The next section explains how we construct the dataset. Section 3 describes observable characteristics of the studies and variation in experimental design. Section 4 presents the results.

2 Data and Method

2.1 The Convex Time Budget Protocol

There is a large body of evidence on estimation of time preferences, including present-biased preferences. Many experimental methods have been proposed in the literature, but here we focus on the method called the Convex Time Budget (CTB) introduced by Andreoni and Sprenger (2012).[5]

The main goal of this method is to elicit all the parameters of the QHD model—the discount factor δ, present bias β, and instantaneous utility function u—in a single experimental instrument. Subjects in a CTB experiment are asked to choose a "bundle" of rewards (x_t, x_{t+k}) delivered at two points in time $(t, t+k)$, under an intertemporal budget constraint with a k-period gross interest rate of $1+r$. By asking a series of allocation questions with varying $(t, t+k)$ and $1+r$, one can identify parameters of the QHD model.[6] See more details in Online Appendix A.

The CTB protocol instantly became popular. The protocol has been applied not only in laboratory experiments but also in field experiments in developing countries. As we describe below, we have variation in several aspects of CTB design which we exploit in meta-regression analysis.

2.2 Identification and Selection of Relevant Studies

Every good meta-analysis starts by casting a wide net trying to identify relevant studies. In order to deliver an unbiased meta-analysis, it is important to make sure that identification and selection of papers are guided by unambiguously defined inclusion criteria. In our case, the main criterion is to "include all articles that conducted experiments or surveys with the CTB protocol." We

[4] Cohen et al. (forthcoming) document the design characteristics of 222 empirical studies identified using Google Scholar, but they do not analyze parameter estimates reported in these studies.

[5] An experimental design concept that is similar to the CTB is discussed in Cubitt and Read (2007).

[6] Roughly speaking, variation in gross interest rates $1+r$ identifies the curvature of the instantaneous utility function u, variation in the delay length k identifies the discount factor δ, and whether the sooner payment date is today ($t=0$) or not identifies present bias β. Since the key driver of the identification of β is the change in allocations between time points $(0, k)$ and $(t, t+k)$, the CTB protocol is able to recover not only present bias but also future bias. Online Appendix A illustrates optimal allocation decisions in the CTB protocol for a present-biased as well as a future-biased agent against the benchmark of the time-consistent agent (Figure A.2).

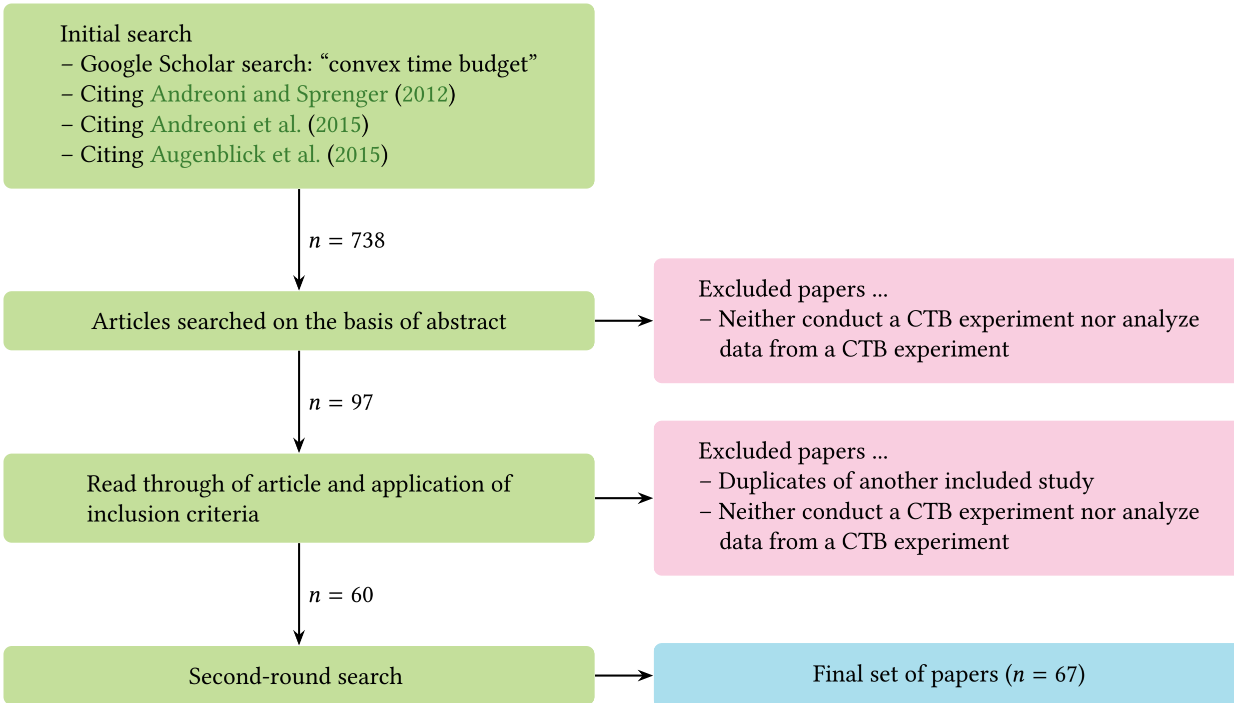


Figure 2: Paper search and data construction.

searched for both published and unpublished papers to have sufficient sample size and to be able to check indicators of publication bias and selective reporting.

We searched articles which employed the CTB protocol using Google Scholar, first by querying papers that cited Andreoni and Sprenger (2012), Andreoni et al. (2015), and Augenblick et al. (2015). We also searched for papers with the keyword 'convex time budget'. These two sets of searches, done on November 28 and December 15, 2017, returned a total of 738 results (including overlaps), which we further narrowed down by examining the titles and the abstracts.

As mentioned above, we searched for any articles, both published and unpublished, which conducted experiments or surveys involving the CTB protocol. Note that this broad inclusion criterion keeps studies even if QHD parameters are not estimated. These studies do not contribute to our main meta-analysis but still provide some additional information regarding how the CTB protocol has been used in the literature. For this reason, we kept track of these studies without estimates, too.

We performed the second-round search (using the same query) and updated the database in the Fall of 2018. The final dataset includes 67 articles.[7] Figure 2 illustrates our selection procedure.

[7]Tables B.1 and B.2 in Online Appendix list all studies (and their basic design characteristics) in the dataset, split by the existence of parameter estimates. Online Appendix D presents the full list of references.

Note that our inclusion criteria specifically exclude other studies which are informative about present bias. Narrative reviews are better equipped to weave discoveries from such papers into a coherent conclusion. For example, Augenblick (2018) varies time of delivery of initial payments, and finds a decay effect in which a few hours of delay reduces present bias substantially. There are many, many other papers in economics, psychology, and cognitive neuroscience which are important but are not included because they did not use CTB.[8]

2.3 Data Construction

After identifying relevant articles, we assembled the dataset by coding estimation results and characteristics of the experimental design. We call a collection of estimates a "study" when they are from the same experimental design. These two units of observations, an article and a study, coincide in many cases, but it allows us to distinguish two conceptually different experiments reported in a single article. For example, monetary reward and effort-cost versions of CTB in Augenblick et al. (2015) are two separate "studies."

Our primary variable of interest is the estimate of present-biasedness, but we also coded other parameters in the QHD model (such as discount factor, utility curvature, and parameter for stochastic choice, if available) as well. Studies report either aggregate-level parameter estimates (i.e., pool choice data from all subjects and estimate a set of parameters for the "representative subject") or some summary statistics, such as the mean or median of individual-level estimates. We coded these two types of estimate separately.[9] We also coded standard errors of parameter estimates from aggregate-level analysis in order to control for the quality of the study in the meta-analysis reported below.

We also coded variables describing characteristics of experimental design and econometric strategies. These variables include, among others: location of the experiments (e.g., laboratory, field, online); types of reward (e.g., real or hypothetical, money, effort); delivery method (e.g., cash, check, gift card); subject pool (e.g., children, college student, general population); and so on. Table B.4 in Online Appendix lists variables coded in the study. Some studies implemented the CTB protocol with some treatment variations, such as hunger, cognitive resource depletion,

[8]In a companion paper, we conduct a larger-scale meta-analysis using papers which estimate discounting parameters using any method, extending the scope beyond CTB (Imai et al., 2018).

[9]In our main meta-analysis below, we focus only on the aggregate-level estimates since there are not many individual-level estimates and the reporting format is not common across these studies. More precisely, we identified only 44 individual-level estimates from 10 studies. Six of these estimates are the mean of the distribution and the other 38 are the median. The former six estimates are accompanied with the standard deviation of the distribution. See Figure B.1 in Online Appendix.

financial education intervention, time pressure, and so on (Table B.3 in Online Appendix). We coded a dummy variable for treatment. We call a study "neutral" if there is no treatment variation (there is a single data set of experimental condition).

3 Features of Studies and Experimental Designs

We identified 67 articles that conducted experiments or surveys that used the CTB protocol, where 36 of them are published (or "in press") including nine articles published in one of the "Top 5" journals (as of December 31, 2018). There are 41 articles that report structurally estimated QHD parameters either at the aggregate level or at the individual level. The median number of estimates reported in an article is three. Seven studies reported more than 10 estimates, and two of them reported more than 30 (Table B.1 in Online Appendix).

Observable features of experimental design do not exhibit marked difference between studies with parameter estimates and those without (Tables 1 and 2; Figure C.3 in Online Appendix).

Roughly half of the studies report laboratory experiments. Online experiments constitute fewer than 20% of the studies in the dataset. Only one experiment studied choices made by children in a classroom. Studies were conducted in 29 different countries as shown in Figure C.2, although a third of studies analyzed data from the USA.[10]

Most of the studies recruited participants from the population of college/university students, or a general population including retirees. It is important to note that several studies in our sample estimated QHD parameters using non-monetary rewards (more precisely, using the cost of working on tedious real-effort tasks) following Augenblick et al. (2015). Studies which used monetary reward differed in how future payments were made: some used bank transfer or sent checks to the subjects, but in some other experiments subjects came back to the laboratory to pick up the payments.

These observable study characteristics exhibit some patterns of co-occurrence (Figures C.4-C.6 in Online Appendix). For example, laboratory experiments tended to have student subjects while field studies are more likely to recruit from the general population.

Experimental elicitation of time preferences requires researchers to design experiments so that the effects of potential confounding factors are minimized. As discussed in the literature, two notable examples of potential confounding factors are the uncertainty or distrust of future

[10]These 29 countries/regions are: Afghanistan; Australia; China; Colombia; Ethiopia; France; Germany; Guatemala; India; Italy; Japan; Kenya; Malawi; Mozambique; Nepal; Netherlands; Nigeria; Pakistan; Philippines; Singapore; South Africa; Spain; Taiwan; Thailand; Turkey; Uganda; UK; USA; Vietnam.

Table 1: Characteristics of CTB studies in the dataset.

	All CTB studies		Studies with estimates	
	Frequency	Proportion (%)	Frequency	Proportion (%)
Total number of studies	67	100.0	36	100.0
Content of study				
Report *PB* parameter estimates	36	53.7		
Publication status (as of 12/31/2018)				
Published	36	53.7	17	47.2
Published in “Top 5” journal	9	13.4	3	8.3
Type of study				
Lab experiment	29	43.3	15	41.7
Field experiment	27	40.3	14	38.9
Online experiment	10	14.9	6	16.7
Classroom	1	1.5	1	2.8
Geographic location				
Continent: North America	22	32.8	13	36.1
Continent: Europe	13	19.4	8	22.2
Continent: Asia	17	25.4	9	25.0
Continent: Africa	11	16.4	5	13.9
Continent: Oceania	2	3.0	0	0.0
Continent: South America	2	3.0	1	2.8
Reporting of PB parameter estimates				
Aggregate-level estimates			31	86.1
with standard errors			28	77.8
Individual-level estimates			10	27.8

Note: *“Top 5 Journal”* indicates that the paper is published (or “in press”) in one of the following journals: *American Economic Review*; *Econometrica*; *Journal of Political Economy*; *Quarterly Journal of Economics*; *Review of Economic Studies. Reporting of parameter estimates*: A paper is counted as reporting a particular type of estimate if it reports *at least one* specification reporting the given type of estimate. Five additional studies reported estimates of EDU parameters, not QHD (i.e., no *PB* parameter in the model).

payment and the differences in transaction costs between receiving outcomes at earlier and later dates (e.g., Cohen et al., forthcoming; Ericson and Laibson, 2019).[11] Andreoni and Sprenger (2012) dealt with these issues using the following strategies: (i) they gave the experimental participants the business cards of the researcher (and told them to reach out if they did not receive the pay-

[11] Our view is that both uncertainty about payment and transaction costs are minor factors which many previous experiments have controlled effectively, in the sense that they do not change estimates of *PB* by numerical amounts which would give one pause in deciding whether *PB* should be investigated in applications. See Halevy (2014) for similar skepticism.

Table 2: Characteristics of CTB studies in the dataset.

	All CTB studies		Studies with estimates	
	Frequency	Proportion (%)	Frequency	Proportion (%)
Total number of studies	67	100.0	36	100.0
Subject population				
Kids and teens	7	10.4	1	2.8
University students	28	41.8	15	41.7
General population	32	47.8	20	55.6
Reward type				
Real incentive	65	97.0	34	94.4
Certain	63	94.0	36	100.0
Gains	59	88.1	29	80.6
Money	53	79.1	29	80.6
Effort	9	13.4	8	22.2
Reward delivery method				
Bank transfer	19	28.4	11	30.6
Pickup	5	7.5	3	8.3
Check	10	14.9	6	16.7
Cash	8	11.9	7	19.4
Paypal	2	3.0	2	5.6
CTB implementation				
Corner allowed	58	86.6	30	83.3
Computer	28	41.8	19	52.8
Deal with confounding factors				
Uncertainty about future payments	46	68.7	23	63.9
Equalize transaction cost	52	77.6	28	77.8

Note: A paper is counted as offering a certain type of reward if it offers the reward to *at least one* of the samples the study analyzes.

ment) to increase trust; and (ii) they split the participation fee into two parts, one delivered together with the "sooner payment" and the other delivered with the "later payment," to reduce the difference in transaction costs of receiving rewards at two different points in time. Many of the later studies in our sample also followed these strategies.

Let us now turn to the detail of the CTB protocol. There are several variables which researchers can specify: number of budgets (i.e., questions); set of time frames (pairs (t, k) of "sooner" payment date t and delay length k); gross interest rates over k periods; and so on. Table 3 summarizes the ranges and central tendencies of these design variables.

On average, researchers asked 22 questions to recover QHD parameters. In all protocols the

Table 3: Characteristics of budgets and time frames.

	All CTB studies (60)				Studies with estimates (38)			
	Mean	Median	Min	Max	Mean	Median	Min	Max
Number of budget sets	17.69	14.50	1.00	55	21.88	20.00	4.00	55
Number of time frames	3.18	2.00	1.00	10	3.78	3.00	1.00	10
Minimum delay length (days)	34.89	28.00	1.00	365	40.88	30.00	1.00	365
Maximum delay length (days)	166.40	32.50	1.00	7,300	236.85	56.00	1.00	7,300
Mean delay length (days)	90.72	30.00	1.00	3,285	123.95	42.00	1.00	3,285

questions were asked close together in time.[12] Subjects made allocation decisions on four different (t, k) pairs on average, implying that each time frame was associated on average with five levels of gross interest rates over k periods. The length of delay between the "sooner" payment and the "later" payment varied substantially across studies. On average, the minimum waiting period is a little over one month and the maximum waiting period is six to eight months.

Finally, we look at the assumptions and econometric approaches employed to structurally estimate QHD parameters (Table 4). There are 227 estimates in the dataset, and a significant majority assume a constant relative risk aversion (CRRA) specification for the instantaneous utility function u in the model (1). The typical specification for studies using real-effort tasks is a convex effort cost function. There are five observations where the utility curvature was either fixed at some exogenous value or imputed from an additional elicitation task such as a multiple price list (Holt and Laury, 2002).

The popular econometric approach is (two-limit) Tobit regression, since researchers need to handle censoring due to corner choices. See Andreoni and Sprenger (2012) and Augenblick et al. (2015) for a detailed explanation of identification and estimation using nonlinear least squares (NLS) and Tobit approaches.

[12]It is conceivable that people are "artificially" consistent, giving the same early-late allocations under time frames $(0, k)$ and (t, k). Such a desire to appear consistent will lead to estimates of *PB* biased toward one. If this is the case, the most we can say is that *PB* estimates represent a bound (an upper bound if *PB* is less than one and a lower bound if *PB* is greater than one). A different procedure that increases elapsed time between responses might produce *PB* values closer to one. Note, however, that Imai and Camerer (2018) used an adaptively-optimal experimental design procedure that selected individually-tailored budget lines and time frames based on each subject's responses to the previous questions. In that design, questions subjects faced varied substantially from trial to trial, and $(0, k)$ and (t, k) budget lines with the same level were rarely presented together. In that design, it is more difficult to select allocations in an artificially consistent manner, yet the estimated *PB* values are similar to the standard non-adaptive design (with monetary reward) covered here. While this is just one study, it suggested that a procedural change that happened to reduce between-trial consistency did not change the value of estimated *PB* much.

Table 4: Characteristics of aggregate-level *PB* estimates.

	Frequency	Proportion (%)
Number of estimates	227	
SE reported	220	96.9
Instantaneous utility function u		
Estimated	222	97.8
Imputed	2	0.9
Fixed	3	1.3
Specification of u		
Constant relative risk aversion (CRRA)	183	80.6
Constant absolute risk aversion (CARA)	15	6.6
Other	6	2.6
Convex effort cost	22	9.7
Estimation method		
OLS or NLS	62	27.3
Tobit	107	47.1
Multinomial logit or maximum likelihood	25	11.0
Background consumption		
Fixed at zero	134	59.0
Fixed at non-zero value	70	30.8
Estimated	23	10.1

4 Results

Aggregate-level estimates of the present-bias parameter from each article in the dataset are shown in Figure 3A. About 77% of these estimates are below one, indicating present bias. It is clear from the figure that these estimates vary not only between studies but also within each study. We have 220 aggregate-level estimates with standard errors (Table 4). In this section, we first calculate the "average" present-bias parameter using the standard meta-analytic technique. We next investigate the existence or absence of selective reporting. Finally, we investigate the heterogeneity of observed estimates using the moderator variables coded in our dataset.

4.1 Meta-Analytic Synthesis of Present Bias Estimates

We start by providing a meta-analytic estimation of the "average" *PB* in the dataset. The analysis below provides a tentative answer to the question: *What is the average value of PB measured by the CTB protocol?*

We begin with setting up the simplest meta-analytic framework, the common-effect model.

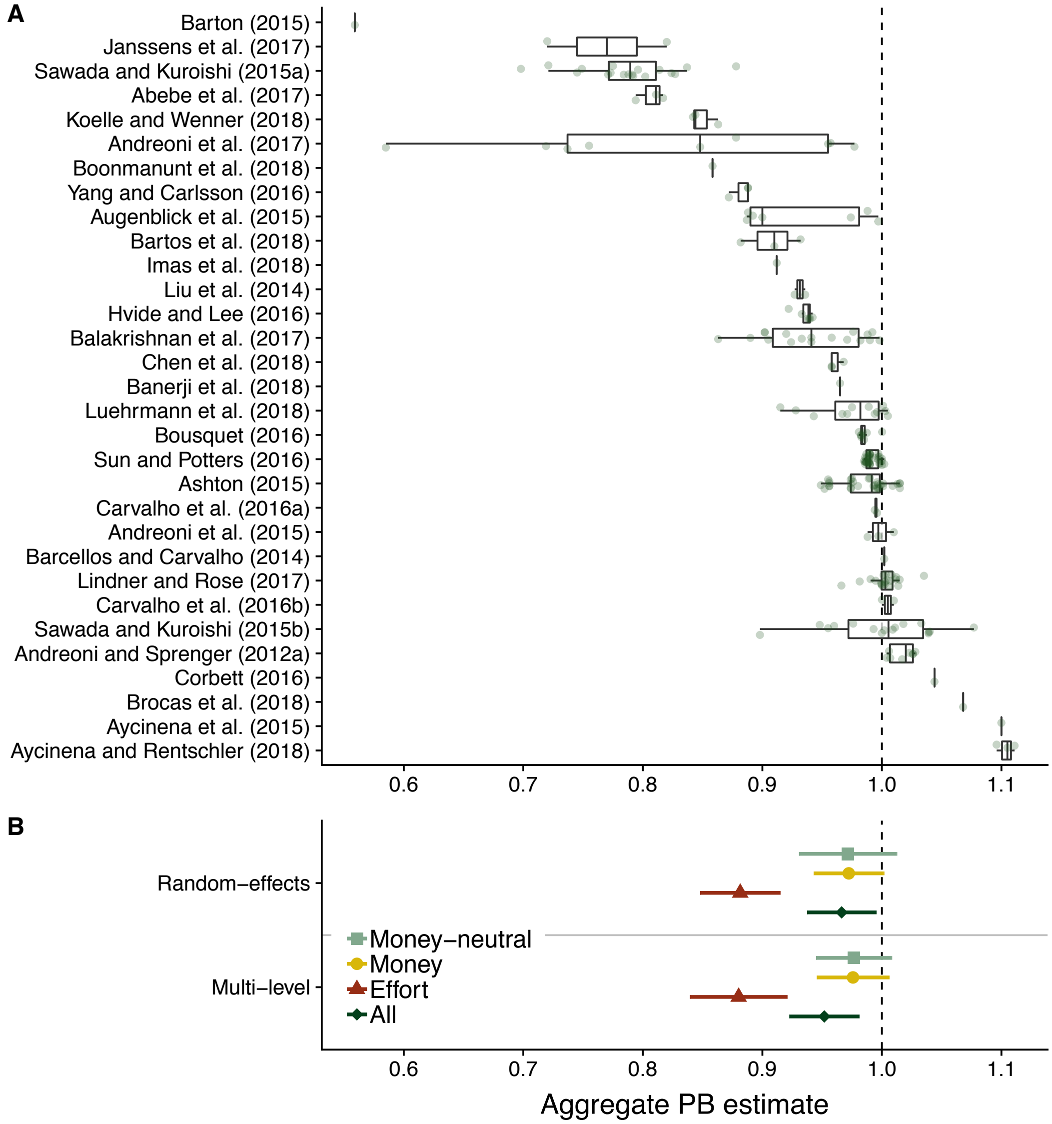


FIGURE 3: Present-bias parameter estimates. The vertical dotted line indicates no present/future bias.

It is

$$PB_j = PB_0 + \varepsilon_j, \tag{2}$$

where PB_j is the jth estimate of present-bias parameter in the dataset ($j = 1, \ldots, m$), PB_0 is the "true" present-bias parameter that is assumed to be common to all observations in the data, and ε_j is the sampling error. It is assumed that $\varepsilon_j \sim \mathcal{N}(0, v_j^2)$ and the sampling variance v_j^2 is known. We can obtain the common-effect estimate of PB_0 as the weighted average of individual estimates:

$$\overline{PB}_0^{\text{CE}} = \frac{\sum_{j=1}^{m} w_j^{\text{CE}} PB_j}{\sum_{j=1}^{m} w_j^{\text{CE}}},$$

where the weights are given by the inverse variance, $w_j^{\text{CE}} = 1/v_j^2$. In this average, estimates with higher precision (smaller standard errors) are given larger weights.

The random-effects meta-analysis (RE; DerSimonian and Laird, 1986) assumes that

$$PB_j = \mu_j + \varepsilon_j = PB_0 + \xi_j + \varepsilon_j, \tag{3}$$

where ε_j is a sampling error of PB_j as an estimate of μ_j, and the estimate-specific "true" effect μ_j is decomposed into PB_0 (the grand mean) and the sampling error ξ_j. It is further assumed that $\xi_j \sim \mathcal{N}(0, \tau^2)$, where τ^2 captures the between-observation heterogeneity, beyond the mere sampling variance, that must be estimated. Note that the random-effects model (3) reduces to the common-effect model (2) when $\tau^2 = 0$. The random-effects estimate $\overline{PB}_0^{\text{RE}}$ is again the weighted average of the individual PB_j, but now the weights are given by $w_j^{\text{RE}} = 1/(v_j^2 + \hat{\tau}^2)$ where $\hat{\tau}^2$ is the estimate of τ^2.

Our dataset includes *statistically dependent* estimates of *PB* since many studies included in our meta-analysis report multiple estimates from the same experiment (e.g., using different econometric approaches or using different subsamples). In order to account for the dependency, we use cluster-robust variance estimation to account for correlation of estimates among each study (Hedges et al., 2010).

We also address the issue of "overly influential" observations (i.e., leverage points) by calculating *DFBETAS* (Belsley et al., 1980), which measures how much the regression coefficient changes if one observation is removed, standardized by the coefficient standard error from the regression without the target observation. Following Bollen and Jackman (1985), we identify any observations to be influential if $|DFBETAS| > 1$ (i.e., the observation shifts the coefficient at least one standard error).[13] This procedure identifies three influential observations in our data: one estimate from Barcellos and Carvalho (2014) and two estimates from Liu et al. (2014). We remove these three estimates from our simple meta-analysis presented in this subsection.[14]

[13] *DFBETAS* is intended to measure the impact of removing observation m on the kth coefficient. Let $\widehat{\gamma}_k$ and $\widehat{\gamma}_k^{(m)}$ be the estimated kth coefficient with and without observation m, respectively. Then, the impact of observation m is given by $DFBETAS_m = (\widehat{\gamma}_k - \widehat{\gamma}_k^{(m)})/SE(\widehat{\gamma}_k^{(m)})$, where $SE(\widehat{\gamma}_k^{(m)})$ is the standard error of $\widehat{\gamma}_k^{(m)}$.

[14] Online Appendix C.4 presents results with these three estimates included.

Table 5: Meta-analytic average of present bias parameter.

	All studies		Monetary (all)		Monetary ("neutral")		Effort cost	
	(1) RE	(2) ML	(3) RE	(4) ML	(5) RE	(6) ML	(7) RE	(8) ML
$\overline{PB}_0$	0.9663	0.9518	0.9723	0.9758	0.9716	0.9766	0.8815	0.8802
	(0.0147)	(0.0149)	(0.0150)	(0.0154)	(0.0209)	(0.0161)	(0.0171)	(0.0208)
p-value	0.0297	0.0031	0.0805	0.1334	0.1898	0.1640	0.0001	0.0004
$\hat{\tau}^2$	0.0031		0.0029		0.0037		0.0021	
I^2	98.0824		98.1257		97.5754		45.9587	
I^2_{within}		0.7528		0.9336		0.4389		9.2236
I^2_{between}		98.1997		97.9088		97.6832		39.5324
Observations	217	217	193	193	140	140	24	24
Studies	29	29	20	20	19	19	9	9

Notes: p-values are from the two-sided test of the null hypothesis $H_0 : PB = 1$. Standard errors in parentheses are cluster-robust (Hedges et al., 2010). τ^2 in the random-effects model is estimated using the restricted maximum likelihood method. Three observations with large influence measure ($|DFBETAS| > 1$) are excluded.

We estimate the meta-analytic averages for four different subsets of the data: (i) all estimates, (ii) observations from studies using monetary reward, (iii) observations from "neutral" studies using monetary reward, and (iv) observations from studies using the real-effort version of the CTB.

Table 5 reports the results from the random-effects specification (odd-numbered columns; also presented in Figure 3B).[15] All specifications show $\overline{PB}_0^{\text{RE}} < 1$, indicating present bias. The overall $\overline{PB}_0^{\text{RE}}$ is 0.97, which is statistically significantly different from one at the 5% significance level. Two estimates from CTB studies using a monetary reward are also smaller than one, but we cannot reject the null hypothesis of no present bias. We observe a smaller $\overline{PB}_0^{\text{RE}}$ of 0.88 in the real-effort version of CTB studies compared to those using monetary reward. We explore and discuss this difference below in Section 4.3.

From the I^2 statistic (Higgins and Thompson, 2002), we observe that 98% of the total variability in estimates from monetary CTB and 46% of the total variability in estimates from real-effort CTB are due to between-observation heterogeneity rather than sampling variance.[16] Note that

[15] Anticipating the amount of between-study heterogeneity in estimated PB, we directly jump to the random-effects model. Results from the common-effect specification are reported in Table C.1 in Online Appendix.

[16] The I^2 statistic gives the amount of heterogeneity relative to the total amount of variance in the observed effects. Formally, the I^2 statistic is computed by

$$I^2 = \frac{\hat{\tau}^2}{\hat{\tau}^2 + s^2} \times 100\%,$$

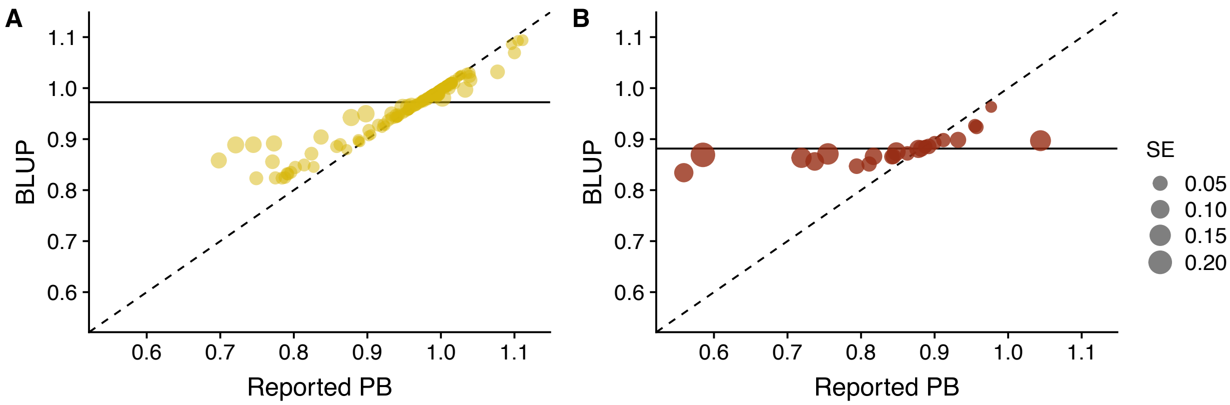


FIGURE 4: Reported *PB* and the corresponding best linear unbiased predictors (BLUPs). (A) Studies with monetary reward. (B) Studies with real-effort. Reported estimates PB_j are drawn proportionally to the associated standard errors. Solid horizontal lines correspond to random-effects estimates (0.9723 and 0.8815).

estimates from real-effort CTB are less precisely estimated (i.e., they are associated with larger standard errors) compared to those from monetary CTB.

Given the random-effects estimates $\overline{PB}_0^{\text{RE}}$ and the estimated degree of heterogeneity $\hat{\tau}^2$, we can construct the best linear unbiased predictors (BLUPs), also known as the empirical Bayes estimates:

$$\widehat{PB}_j = \omega\overline{PB}_0^{\text{RE}} + (1-\omega)PB_j = \frac{v_j^2}{v_j^2+\hat{\tau}^2}\overline{PB}_0^{\text{RE}} + \frac{\hat{\tau}^2}{v_j^2+\hat{\tau}^2}PB_j,$$

where the weight ω captures the degree to which the estimates are "pooled" together.[17] BLUPs lie between PB_j and $\overline{PB}_0^{\text{RE}}$, and the relative position depends on the size of sampling variability v_j^2 and the degree of heterogeneity $\hat{\tau}^2$. Figure 4 demonstrates how "shrinkage" of BLUPs works, especially for the less precise estimates.

As an alternative approach to handle statistically-dependent *PB* estimates within each paper, we also apply a multi-level random-effects model (ML; Konstantopoulos, 2011; Van den Noortgate et al., 2013).[18] Let PB_{ij} denote the jth estimate of *PB* parameter from study i. The first level is $PB_{ij} = \mu_{ij} + \varepsilon_{ij}$, where μ_{ij} is the "true" present-bias parameter and $\varepsilon_{ij} \sim \mathcal{N}(0, v_{ij}^2)$ for the jth

where $\hat{\tau}^2$ is the estimated value of τ^2 and

$$s^2 = \frac{(m-1)\sum w_j}{(\sum w_j)^2 + \sum w_j^2}$$

is the "typical" sampling variance of the observed effect sizes, where $w_j = 1/v_j^2$.

[17] It is called the "pooling factor" in the Bayesian hierarchical modeling (Gelman and Hill, 2007; Meager, 2019).

[18] More precisely, we assume a "three-level" model structure. The common-effect model (2) and the random-effects specification (3) described above can be seen as "two-level" models where the first level is $PB_j = \mu_j + \varepsilon_j$ and the second levels are $\mu_j = PB_0$ for the common-effect model and $\mu_j = PB_0 + \xi_j$ for the random-effects model.

estimate in study i. The second level is $\mu_{ij} = \lambda_i + \xi_{ij}^{(2)}$, where λ_i is the average present-biasedness in study i and $\xi_{ij}^{(2)} \sim \mathcal{N}(0, \tau_{(2)}^2)$. Finally, the third level is $\lambda_i = PB_0 + \xi_i^{(3)}$, where PB_0 is the population average of PB and $\xi_i^{(3)} \sim \mathcal{N}(0, \tau_{(3)}^2)$. These equations are combined into a single model:

$$PB_{ij} = PB_0 + \xi_{ij}^{(2)} + \xi_i^{(3)} + \varepsilon_{ij}.$$

A small value of $\tau_{(2)}^2$ indicates that the estimates are similar at the study level (i.e., there is little within-study variation of different estimates). A large $\tau_{(3)}^2$ suggests that the "true" present-bias parameter varies a lot across studies. Under the typical assumption of $\mathrm{Cov}(\tau_{(2)}^2, \tau_{(3)}^2) = \mathrm{Cov}(\tau_{(2)}^2, \varepsilon_{ij}) = \mathrm{Cov}(\tau_{(3)}^2, \varepsilon_{ij}) = 0$, we have $\mathbf{E}[PB_{ij}] = PB_0$.

In this multi-level specification, estimates (presented in even-numbered columns in Table 5) are close to the results from the random-effects approach discussed above. The overall average of PB is 0.95, which is statistically significantly different from one. While average PB from monetary studies is around 0.98, we cannot reject the null hypothesis of no present bias. Finally, effort-CTB produces a smaller average PB of 0.88. The heterogeneity measures I^2 adjusted to the multi-level specification indicate that, in studies with monetary reward, 98% of total variance is due to between-study heterogeneity. In the real-effort version of CTB studies, between-study heterogeneity is estimated to be less than 40%. In both cases, within-study heterogeneity is small.

Taken together, we find that the average value of PB measured by the CTB protocol is between 0.95 and 0.97. We do not observe statistically significant present bias, on average, in studies using monetary reward, but those with real-effort produce a smaller average PB of 0.88. Note that there is a genuine heterogeneity in estimates from monetary studies. Below we further explore this heterogeneity using meta-regression models with plausible moderator variables (Section 4.3).

4.2 Identifying and Correcting for Selective Reporting

This section provides a tentative answer to our second question: *Is there selective reporting or publication bias?*

Scientific cumulation of knowledge is thrown off track and slowed down by selective reporting or publication of results. The typical concern is when the sign or magnitude of a statistical relationship is strongly predicted by theory, or becomes conventionally believed after preliminary studies. Then new studies which derive an unpredicted or unconventional result may be underreported or underpublished. We will refer to this misproduction of results as "selective reporting" or "publication bias".

There are several possible sources of selective reporting. One is conscious fraud. Another is

"p-hacking", in which multiple analyses are run to get the expected effect (without accounting for multiple comparisons during the specification search). A third source is that scientists who discover a genuine contradictory effect (and do not p-hack their way out of it) may simply not report results in any form, such as a conference presentation or preprint; the contradictory effect ends up in a "file drawer". A fourth source is that even if scientists attempt to publish contradictory effects, journals may implicitly screen them out or encourage, in the review process, p-hacking.

For a single study it is very difficult to detect any of these kinds of selective reporting (except clumsy frauds). However, in a group of related studies there are ways to detect possible collective selective reporting.

The QHD model emerged to explain observed patterns of present-biased choices, including procrastination and challenges self-control. Selective reporting would therefore seem most likely to exaggerate the number of studies estimating the present-bias parameter to be significantly below one, since an estimate of the present-bias parameter below one is consistent with preferences than could generate the observed pattern of present-biased choices that the QHD model is trying to capture.

The funnel plot provides a useful first step for detecting selective reporting (and counterfactually correcting for it). Selective reporting will lead to "missing studies" which create an asymmetry in the funnel plot. Figure 1 presents suggestive evidence of selective reporting—there is an asymmetry even though the magnitude may not be huge (see also Online Appendix Figure C.1, which presents funnel plots for monetary-CTB and effort-CTB separately).

Given the relatively large standard errors of the some of the studies in our sample, it is noticeable that we don't see as many studies as we might expect (in the aggregate) with an imprecise estimate of the present-bias parameter consistent with *future bias* ($PB > 1$). Since future bias is viewed as an "unreasonable" finding (in light of voluminous evidence documenting $PB < 1$), the lack of such findings apparent from the funnel plot provides initial evidence that selective reporting may be an important factor in this literature.

A common procedure for detecting and correcting for publication selection bias is the FAT-PET procedure (Stanley and Doucouliagos, 2012, 2014).[19] In the absence of selective reporting, the reported estimates of the present-bias parameter should be uncorrelated with their standard errors. In the presence of selective reporting, on the other hand, the reported estimates are correlated with their standard errors (more imprecise estimates in the unconventional direction will

[19] This is an acronym for a combination of *Funnel Asymmetry Test* (FAT) and *Precision Effect Test* (PET).

go unreported). This motivates a simple regression model for detection of selective reporting:

$$PB_{ij} = \alpha_0 + \alpha_1 \cdot SE_{ij} + \varepsilon_{ij}, \tag{4}$$

where PB_{ij} and SE_{ij} are again the jth estimates of the present-bias parameter and their associated standard errors reported in the ith study. In this model, $\alpha_1 \neq 0$ captures the degree of selective reporting bias. The estimate of α_0 naturally serves as an estimate of the selection-corrected effect size (since it corresponds to an extrapolated effect size with zero standard error and hence perfect precision). Note that the variance of ε_{ij} in this regression will vary across estimates. Therefore, it is often suggested to use weighted least squares (WLS) with the inverse of the variance of the study's estimate ($1/SE_{ij}^2$) as the weight (Stanley and Doucouliagos, 2012). This model allows us to test the asymmetry of the funnel plot (FAT; Egger et al., 1997; Stanley, 2005, 2008) as well as whether there is a genuine effect beyond publication selection (PET). See Stanley and Doucouliagos (2012) and Stanley (2017) for in-depth discussion (especially on the limitations of these approaches).

Table 6 reports results from estimation of model (4) using the unrestricted weighted least squares (Stanley and Doucouliagos, 2015). The estimated values of α_1 are negative, indicating that less precise (i.e., larger *SE*) studies are associated with lower estimates of *PB* (i.e., more present-biased). We do not reject the null hypothesis that the coefficient on *SE* is zero in studies with monetary reward (columns (3)-(6)) while the relationship is statistically significant for studies using effort cost (columns (7)-(8)). The intercept α_0 represents an estimate of "true" underlying *PB* that has been corrected for selective reporting. The results indicate that the "bias-corrected" estimate of *PB* is statistically indistinguishable from one, due to strong relationship between reported *PB* estimates and their standard errors.

It has been argued that the performance of commonly used bias-correction methods such as the FAT-PET procedure depends on the nature of the data, and no single method dominates the other in all circumstances (Alinaghi and Reed, 2018; Carter et al., 2019; Hong and Reed, 2019). Therefore, we also report results from other bias-correction methods recently introduced in the literature.

We first apply the *latent studies* method for identification and correction for selective reporting proposed by Andrews and Kasy (2019), which models the conditional probability of publication as a function of a study's results (discussed in detail in Online Appendix C.7).[20] The results are

[20] While Andrews and Kasy (2019) model conditional *publication* probabilities, our application of the method is intend to capture conditional *reporting* probabilities.

Table 6: Funnel plot asymmetry and precision effect testing.

		All studies		Monetary (all)		Monetary ("neutral")		Effort cost	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SE of PB estimate	α_1	−1.4498	−0.3679	−1.3185	−0.2480	−1.6776	−0.1872	−2.0571	−1.8720
		(0.6187)	(0.3329)	(0.7260)	(0.3410)	(1.0459)	(0.3917)	(0.4412)	(0.1093)
Constant	α_0	1.0002		0.9998		1.0077		0.9931	
		(0.0032)		(0.0032)		(0.0056)		(0.0255)	
FAT ($H_0 : \alpha_1 = 0$)	p-value	0.0265	0.2785	0.0852	0.4759	0.1261	0.6385	0.0016	0.0000
PET ($H_0 : \alpha_0 = 1$)	p-value	0.9475		0.9393		0.1831		0.7931	
Study fixed effect		No	Yes	No	Yes	No	Yes	No	Yes
Observations		217	217	193	193	140	140	24	24
Number of studies		29	29	20	20	19	19	9	9
R^2		0.1823	0.8429	0.1400	0.8377	0.1777	0.9055	0.5100	0.9503
Adjusted R^2		0.1785	0.8186	0.1355	0.8189	0.1717	0.8906	0.4877	0.9183
Other bias-correction methods									
Latent-studies method	$\overline{PB}_0$	0.974		0.987		0.939		0.904	
		(0.040)		(0.051)		(0.064)		(0.016)	
Stem-based method	$\overline{PB}_0$	0.9910		0.9910		0.9992		0.9266	
		(0.0029)		(0.0029)		(0.0036)		(0.0253)	

Note: Estimated by weighted least squares. Standard errors are clustered at the study level. Three observations with large influence measure ($|DFBETAS| > 1$) are excluded. In the specification with study fixed effects, the constant term is dropped and all the dummy variables for the studies are included. Details of the latent-studies method and the stem-based method are presented in Online Appendices C.7 and C.8, respectively.

shown in Tables 6 (and Table C.6 in Online Appendix). Although none of the relative reporting probabilities for estimates with different intervals of Z-values ($Z < -1.96$, $-1.96 \leq Z < 0$, $0 \leq Z < 1.96$) are individually significantly different from one, joint tests show evidence that, for monetary studies, selective reporting acts to "squeeze" *PB* estimates towards one from both sides (so in this case, selective reporting acts to hide, instead of exaggerate, statistically significant findings). For effort studies, selective reporting does in fact cause overrepresentation of statistically significant estimates of present bias, consistent with our FAT-PET results. That being said, the degree of selective reporting is not drastic—the adjusted study estimates from the latent studies model are very similar to the original study estimates (shown in Figure C.26 of the Online Appendix).

Finally, we apply the *stem-based bias correction method* developed by Furukawa (2019) (adapting Stanley et al., 2010), which is discussed in more detail in Online Appendix C.8. Intuitively, this method provides a weighted average of the estimates from an optimally chosen subset of the most precise studies. The results show insignificant aggregate evidence for present bias across the most precise studies. However, when only studies in which subjects make decisions over allocations of effort are included, we find significant levels of present bias, as shown in Figure C.27.

Taken together, we view our results as demonstrating that there is evidence suggesting the existence of modest selective reporting in the direction of overreporting $PB < 1$ in studies using a real-effort task. Correcting for potential selective reporting pushes values of average PB upward toward one. They are still close to one for monetary studies. For effort studies, values are still lower than those for monetary studies, but the estimated degree of present-biasedness depends on the method used for bias-correction.

4.3 Explaining Heterogeneity

We have thus far assumed that the variability in reported estimates are mainly due to sampling errors, either at the observation level or study level, or both, and potential selective reporting. However, these estimates come from studies that use a variety of experimental designs, participants, and econometric approaches, which may result in systematic variation in reported estimates.[21] This section provides a tentative answer to our third question: *How does reported PB vary reliably with observable study characteristics?*

In order to explain heterogeneity, we now add a set of moderator variables to model (4):

$$PB_{ij} = \alpha_0 + \alpha_1 \cdot SE_{ij} + \gamma \boldsymbol{X}_{ij} + \varepsilon_{ij}, \tag{5}$$

where $\boldsymbol{X}_{ij}$ is a vector of observable characteristics of jth estimate from study i and $\boldsymbol{\gamma}$ is a coefficient vector.

Results from this meta-regression analysis report a tentative answer to the question: How does PB vary reliably with methods, subject population, and other study characteristics?

In the first set of meta-regressions presented in Table 7, we restrict samples to those using monetary reward. We consider eight basic sets of moderators as $\boldsymbol{X}_{ij}$. These variables are categorized into: treatment dummy (omitted category is *Neutral condition*), location of the experiment (omitted category is *Location: Lab*), timing of immediate reward payment (omitted category is *by the end of the experiment*), estimation method (omitted category is *Estimation: Least squares*), treatment of background (b.g.) consumption (omitted category is *Estimation: No b.g. consumption*), and interface (omitted category is *Computerized*). We also include several additional variables which are specific to experiments involving monetary reward: method of reward delivery (omitted category is *Delivery: Check*), treatment of confounding factors such as uncertainty regarding future reward and transaction costs (omitted category is *Ignored* in both variables), and

[21] Online Appendix Figures C.7-C.17 visualize the effects of some representative study characteristics on reported estimates, looking at each characteristic in isolation.

proxies for the ease of access to financial markets at the country level. We estimate the model using unrestricted weighted least squares (Stanley and Doucouliagos, 2017).

The effects of study characteristics on estimated *PB* parameter exhibit notable patterns. Regression coefficients reported in Table 7 (focusing on the first three columns for now) suggest that: (i) field experiments tend to find less present-biased preferences compared to lab studies; (ii) dealing with transaction costs makes estimated *PB* larger; (iii) we do not observe systematic effects of reward delivery method; (iv) we do not observe systematic effects of econometric approaches (e.g., Tobit or NLS); and (v) whether or not to jointly estimate background consumption has little impact on the estimates of *PB*.

Compared to studies that guaranteed to deliver the "immediate" rewards within the day of the experiment, estimated *PB* is smaller (more present-biased) when these "immediate" rewards were delivered by the end of the experiment. It is possible to reason that, if the "immediate" reward is paid at the end of the experiment, that design increases uncertainty about future payments which in turn exaggerates the behavioral *PB* measure (in the direction of the larger present-biasedness). To examine this potential confounding factor, we include the dummy "Deal uncertainty" in a regression specification in column (3) aiming to control for the confidence of future payment delivery. We find that this dummy itself has a statistically insignificant coefficient while the dummy "Immediate payment: Within day" remains to have a statistically significant effect with its magnitude virtually unchanged compared to column (2). These observations suggest that the timing of "immediate" (i.e., $t = 0$) payment appears to matter, as documented in Balakrishnan et al. (2017).

4.3.1 Comparing monetary and non-monetary rewards

Underlying models of intertemporal choices are fundamentally about utility flows at each time period, and not about the receipt of monetary payments. A large share of existing empirical studies have measured time preferences using time-dated monetary payments, but additional assumptions (such as monetary payments being "consumed" at the time of receipt) are necessary to infer individuals' discount functions from observed choices in this approach (Chabris et al., 2008; Cohen et al., forthcoming; Mulligan, 1996). More recent studies try to directly control the timing of utility flow using, for example, real-effort tasks (e.g., Augenblick et al., 2015; Augenblick, 2018; Augenblick and Rabin, 2019; Carvalho et al., 2016; Fedyk, 2018), and report evidence that non-monetary rewards provide estimates of present-bias parameter that are smaller (in the sense of conveying greater levels of present bias) than those from the standard monetary reward studies.

Building on this discussion, our next set of meta-regressions directly compares *PB* estimates

Table 7: Explaining the heterogeneity of reported estimates (studies with monetary reward).

	(1)	(2)	(3)	(4)	(5)	(6)
SE of PB estimate	-1.248^{**}	-1.951^{**}	-1.711^{*}	-0.969^{*}	-1.104^{*}	-0.600
	(0.454)	(0.636)	(0.668)	(0.461)	(0.550)	(0.512)
Non-neutral condition	-0.006	-0.003	-0.006	-0.002	-0.004	-0.001
	(0.005)	(0.005)	(0.007)	(0.006)	(0.006)	(0.008)
Location: Field	0.066^{**}	0.071^{**}	0.090^{***}	0.221^{***}	0.184^{***}	0.107^{**}
	(0.025)	(0.022)	(0.025)	(0.048)	(0.052)	(0.038)
Location: Class	0.011	0.022	0.029^{*}	0.098^{**}	0.049^{*}	-0.007
	(0.013)	(0.013)	(0.014)	(0.035)	(0.022)	(0.028)
Location: Online	-0.010	-0.031^{*}	-0.026	-0.019	-0.016	-0.016
	(0.005)	(0.014)	(0.016)	(0.012)	(0.013)	(0.014)
"Immediate" pay: Within day	0.048^{**}	0.050^{***}	0.051^{***}	0.030^{**}	0.019	0.027
	(0.017)	(0.012)	(0.014)	(0.009)	(0.014)	(0.019)
"Immediate" pay: Not reported	-0.066	-0.060	0.046	-0.080	-0.148	-0.038
	(0.056)	(0.051)	(0.065)	(0.045)	(0.085)	(0.070)
Delivery: Cash	0.029	0.017	0.024	0.009	-0.002	0.009
	(0.021)	(0.018)	(0.018)	(0.013)	(0.018)	(0.019)
Delivery: Bank	-0.004^{*}	-0.003	-0.006	0.045^{**}	0.004	-0.030
	(0.002)	(0.004)	(0.005)	(0.016)	(0.008)	(0.017)
Delivery: Other	-0.008	-0.008^{*}	-0.011^{**}	0.013	-0.001	-0.012^{*}
	(0.004)	(0.004)	(0.003)	(0.008)	(0.005)	(0.005)
Estimation: Tobit		0.018^{*}	0.016	0.005	0.002	0.010
		(0.009)	(0.009)	(0.006)	(0.009)	(0.008)
Estimation: Other		-0.002	-0.001	-0.005	-0.014	-0.002
		(0.006)	(0.006)	(0.008)	(0.010)	(0.008)
Estimation: B.g. consumption		-0.001	-0.001	-0.000	-0.003	0.007
		(0.006)	(0.007)	(0.006)	(0.006)	(0.007)
Deal uncertainty			-0.005	0.025^{**}	0.008	0.004
			(0.004)	(0.009)	(0.007)	(0.008)
Deal transaction cost			0.111^{**}	0.084^{**}	0.011	0.077
			(0.038)	(0.031)	(0.052)	(0.054)
Paper and pencil			-0.017	-0.044	-0.031	-0.025
			(0.013)	(0.031)	(0.019)	(0.017)
Credit card				0.183^{***}		
				(0.051)		
Withdrawal					0.308^{*}	
					(0.131)	
Emergency funds impossible						-0.216
						(0.123)
Constant	0.963^{***}	0.963^{***}	0.854^{***}	0.764^{***}	0.690^{***}	0.966^{***}
	(0.017)	(0.014)	(0.052)	(0.048)	(0.089)	(0.099)
Observations	193	193	193	193	193	193
R^2	0.457	0.500	0.523	0.718	0.657	0.597
Adjusted R^2	0.427	0.464	0.480	0.691	0.623	0.557

Note: Observations with large influence measure ($|DFBETAS| > 1$) are excluded. Study fixed effects are not included in the model. Standard errors are clustered at the study level. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

from studies with monetary and non-monetary rewards, correcting for selective reporting and several study characteristics, to see whether the apparent difference in present bias is evident from CTB alone. We set up a general regression model

$$PB_{ij} = \alpha_0 + \alpha_1 \cdot SE_{ij} + \alpha_2 \cdot SE_{ij}^2 + \boldsymbol{\gamma} \boldsymbol{X}_{ij} + \boldsymbol{\lambda}_1 (SE_{ij} \cdot \boldsymbol{Z}_{ij}) + \boldsymbol{\lambda}_2 (SE_{ij}^2 \cdot \boldsymbol{Z}_{ij}) + \varepsilon_{ij}, \tag{6}$$

which extends equation (5) to allow for any factors that can potentially influence selective reporting (captured by $SE_{ij} \cdot \boldsymbol{Z}_{ij}$ and $SE_{ij}^2 \cdot \boldsymbol{Z}_{ij}$). We include a dummy for monetary studies and its interaction with several study characteristics, so that the constant term (α_0) captures the average *PB* estimate from non-monetary studies.

Table 8 reports the results. The main variable of interest is the coefficient on the dummy *Reward: Money*, which captures the difference between the average *PB* from non-monetary studies and that from the "baseline" monetary studies. The definition of "baseline" studies is: "monetary studies, neutral condition" in the odd columns, and "monetary studies, neutral condition, lab, immediate rewards delivered within the day, estimation with NLS" in the even columns.[22]

As discussed in the literature, studies using non-monetary rewards estimate present-bias parameters that are generally smaller than those from the standard monetary reward studies, regardless of the definition of the baseline in monetary studies (columns (1)-(2)). The other specifications include either *SE* or SE^2, as well as its interaction with *Reward: Money*. The estimated coefficients on *Reward: Money* are not statistically significant when *SE* is included, but are significantly positive when SE^2 is used. These results suggest that the difference between average *PB* from monetary and non-monetary studies shrinks when potential selective reporting is corrected for. However, the size of this difference $PB_{\text{money}} - PB_{\text{effort}}$ depends on the assumption imposed on the relationship between reported *PB* and *SE*.

4.3.2 Access to financial markets

The literature has discussed several drawbacks with the use of dated monetary payments (Coller and Williams, 1999; Cubitt and Read, 2007; Cohen et al., forthcoming). For example, in the presence of complete financial markets (and an associated lack of borrowing constraints), an optimizing individual's rate of time preference over monetary rewards should be equal to the market interest rate at which the individual can save and borrow.

[22] In the meta-regression models presented in Table 8, we do not include dummy variables for design characteristics in non-monetary studies. This is solely due to a power issue—there are only 24 estimates from nine effort-CTB studies in our dataset. It is therefore important to revisit these meta-regression analyses after the literature accumulates more estimates from CTB studies using non-monetary rewards.

Table 8: Explaining the heterogeneity of reported estimates (monetary vs. non-monetary rewards).

	(1)	(2)	(3)	(4)	(5)	(6)
Constant (*PB* from effort-CTB)	0.907***	0.907***	0.993***	0.993***	0.932***	0.932***
	(0.023)	(0.023)	(0.024)	(0.024)	(0.024)	(0.024)
SE of PB estimates			−2.057***	−2.057***		
			(0.414)	(0.414)		
SE^2 of PB estimates					−10.918***	−10.918***
					(2.829)	(2.829)
Reward: Money	0.089***	0.093***	0.015	0.016	0.068**	0.069**
	(0.024)	(0.023)	(0.024)	(0.024)	(0.024)	(0.024)
× Non-neutral condition	−0.003	−0.012	−0.011**	−0.006	−0.007*	−0.010
	(0.004)	(0.007)	(0.004)	(0.006)	(0.003)	(0.006)
× Location: Field		0.057**		0.071***		0.064**
		(0.021)		(0.018)		(0.020)
× Location: Class		0.026		0.037***		0.031*
		(0.017)		(0.011)		(0.015)
× Location: Online		0.004		−0.026		−0.006
		(0.009)		(0.014)		(0.010)
× "Immediate": By end of exp		−0.039*		−0.042***		−0.041**
		(0.017)		(0.011)		(0.015)
× "Immediate": Not reported		−0.127*		−0.112*		−0.113*
		(0.060)		(0.051)		(0.052)
× Estimation: Tobit		0.002		0.019*		0.009
		(0.006)		(0.009)		(0.007)
× Estimation: Other		−0.005		−0.002		−0.004
		(0.005)		(0.004)		(0.004)
× *SE* of PB estimates			0.374	0.065		
			(0.854)	(0.708)		
× SE^2 of PB estimates					−36.379	−26.427*
					(22.497)	(13.157)
Observations	217	217	217	217	217	217
R^2	0.054	0.375	0.249	0.504	0.222	0.456
Adjusted R^2	0.045	0.348	0.235	0.478	0.207	0.427
$H_0 : PB_{\text{effort}} = 1$	$p = 0.0004$		$p = 0.7747$		$p = 0.0078$	

Note: Observations with large influence measure ($|DFBETAS| > 1$) are excluded. Study fixed effects are not included in the model. Standard errors are clustered at the study level. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

To explore whether access to credit markets and liquid savings may drive a wedge between utility parameters estimated in money-earlier-or-later studies and the "true" parameters associated with the individual's discounting of utility flows, we explore the extent to which lack of access to financial markets and methods of saving/borrowing is associated with the display of greater levels of present bias in studies with monetary reward. (Andreoni et al. (2018) refer to this potential effect as the arbitrage channel.)

The data we use is from the World Bank's Global Findex database (Demirgüç-Kunt et al., 2018).[23] The dataset consists of nationally representative samples of adults from 140 countries,

[23] This data can be accessed at: `https://globalfindex.worldbank.org/`.

and focuses on the ability of adults in different countries to access financial services. We focus on three variables for our analysis to proxy for the ease of access to financial markets for experimental participants (see Figures C.18-C.20 in Online Appendix):

1. *Proportion of the adult population with a credit card.* Credit cards provide a relatively cheap way for consumers to borrow against future income, generally up to 30 days in the future. We expect that having access to a relatively easy source of borrowed funds will cause individuals to display less present-biased behavior in money-later-or-earlier tasks.

2. *Proportion of the adult population who made a withdrawal from a financial institution account in the last year.*[24] Since this proxies having access to liquid savings in the current period, we expect that individuals who have made withdrawals within the last year are less likely to display present-biasedness in money-earlier-or-later tasks, since if they desired to increase their current period consumption they could withdraw from their savings accounts instead of needing the current-period reward from the experiment.

3. *Proportion of the adult population who would not be able to come up with emergency funds within the next month.*[25] We expect this to have a positive relationship with observed present-biasedness from money-earlier-or-later studies, since individuals who are unable to come up with emergency funds within the next month are more likely to have consumption closely following income in each period, and so monetary flows may more accurately proxy true utility flows for these individuals (provided other sources of income remain constant over time).

Table 7, columns (4)-(6), show the results. The coefficients have signs in the expected direction, and two of the variables *Credit card* and *Withdrawal* have statistically significant positive effects: studies conducted in countries/regions where more individuals have easier access to financial markets (through credit cards or withdrawals from liquid savings accounts) tend to exhibit less present-biasedness. These results indicate that some part of the observed heterogeneity in estimated *PB* can be attributed to the degree to which individuals have access to financial markets, and that the arbitrage channel discussed by Andreoni et al. (2018) has some effect on the estimated *PB* from money-earlier-or-later studies.

[24] A financial institution is defined by Demirgüç-Kunt et al. (2018) as "a bank or at another type of financial institution, such as a credit union, a microfinance institution, a cooperative, or the post office (if applicable), or having a debit card".

[25] "Emergency funds" are defined as 5% of gross national income per capita in the local currency.

4.3.3 Model uncertainty

The selection of variables and the order of inclusion in the first meta-regression analysis presented in Table 7 are based on prior discussion in the literature as well as co-occurence of study characteristics in the data (Figures C.5 and C.6 in Online Appendix), and thus made somewhat arbitrarily.

We augment our meta-regression analysis with the application of Bayesian model averaging (BMA) to tackle the model uncertainty resulting from the large number of explanatory variables we could have included in our meta-regression model (Hoeting et al., 1999; Moral-Benito, 2015; Steel, forthcoming). BMA runs multiple regressions with different subsets of the explanatory variables (models) and marginalizes over models to obtain the posterior density of the parameters. We provide a more detailed explanation in Online Appendix C.6.[26]

The results of our application of BMA are in line with those reported in Table 7. Figure 5 is representative of our results (the full set of results is provided in Section C.6 of the Online Appendix).

5 Conclusion

We present a quantitative meta-analysis of estimates of the present-bias parameter in the QHD model using choice data from CTB experiments. We collect 220 estimates from 28 articles and found that the meta-analytic average of the present-bias parameter is between 0.95 and 0.97, which is statistically significantly smaller than one. The values for monetary-reward studies are close to one, indicating absence of present bias, on average. On the contrary, effort-based studies report a lower meta-analytic average of 0.88, a statistically significant present bias. There is evidence suggesting selective reporting in studies using a real-effort CTB, and bias-corrected estimates of average *PB* vary from 0.90–0.93 to 0.99, depending on the method used for correction.

We also found that estimates vary greatly across studies, primarily due to their different study characteristics. Our meta-regression analysis suggests that CTB experiments with non-monetary rewards indeed found estimates that are "more present biased" than those from CTB studies with typical monetary rewards. One reason for this difference suggested in the literature relates to decision-makers arbitraging monetary payments using market interest rates, so that allocations in monetary-CTB studies are in fact not representative of underlying parameters in the decision-maker's utility function (Cohen et al., forthcoming). We found evidence that access to financial

[26]For applications of BMA in meta-analysis in economics, see Havránek et al. (2015, 2017) and Iršová and Havránek (2013).

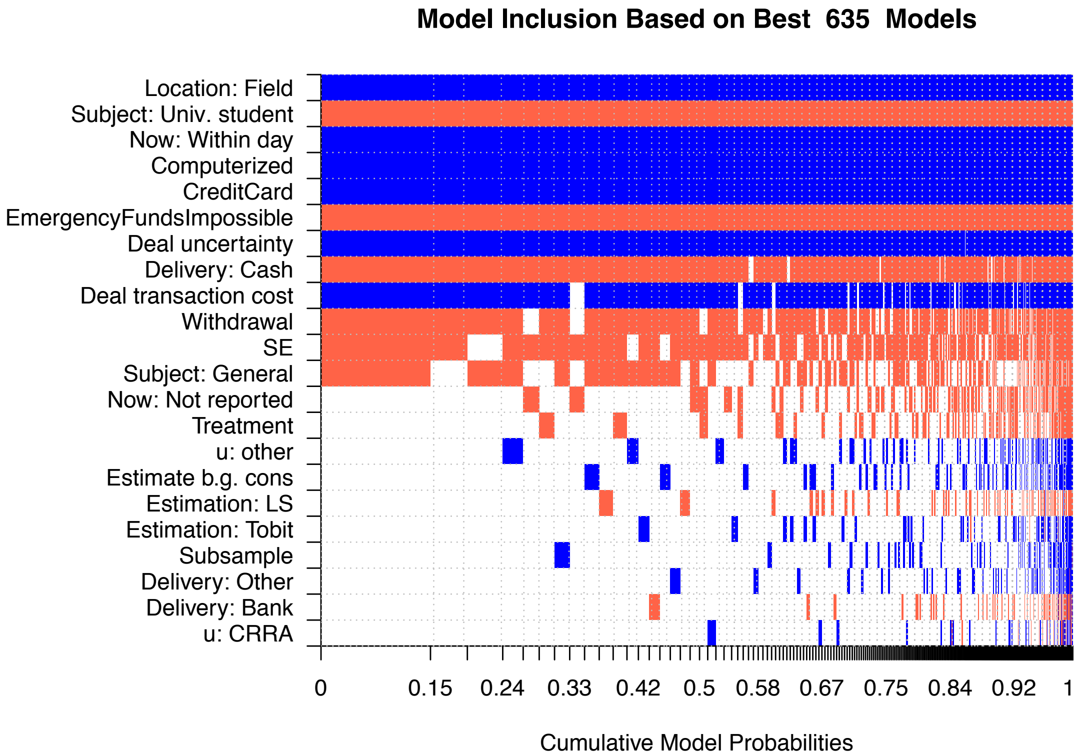


FIGURE 5: Model inclusion results from Bayesian model averaging. *Notes:* In this figure, observations are from monetary-CTB studies only. Columns denote individual models where variables are sorted by posterior model probability in a descending order. Blue cells (darker cells in grayscale) indicate that the variable is included in the model and has a positive coefficient, while red cells (lighter cells in grayscale) indicate that the variable has a negative coefficient. White cells indicate that the variable is not included in the model.

markets is in fact associated with estimates of *PB* close to one (as opposed to estimates showing present biased behavior), suggesting that this arbitrage channel does play at least some role in explaining results from monetary-CTB designs.

Furthermore, we found evidence to confirm the importance of the delay until the issue of the reward associated with the "current" time period (e.g., Balakrishnan et al., 2017; Ericson and Laibson, 2019); across a range of specifications in both our meta-regression and Bayesian model averaging approaches, studies that delivered rewards associated with the "current" period by the end of the experiment, as opposed to only by the end of the day, tended to yield lower estimates of the present-bias parameter, indicating greater levels of present bias in the behavior of subjects.

In addition, we found suggestive evidence concerning the importance of a factor on estimates of present bias that has so far not been widely discussed, the location of the study—whether it takes place in a laboratory or in the field. Both meta-regression and BMA suggest that subjects in laboratory experiments show larger present bias than subjects in field experiments. Many studies follow Andreoni and Sprenger's (2012) original econometric strategy and report estimates using both NLS and Tobit (or estimates with and without background consumption). These methods ignited significant debate in the literature (see, for example, the discussion in Andreoni et al., 2015). However, our meta-analysis showed that the econometric strategy makes little difference.

Our findings naturally lead to two follow-on questions. First, as well as being statistically different from one, are deviations of estimated *PB* from one significantly different from one in a *practical* sense? Second, given that many study characteristics have a systematic influence on the estimated degree of present bias, is there is a preferred method for eliciting present bias?

Regarding the first question, at least in the setting of effort, where we estimate *PB* to be roughly 0.90–0.93, present bias seems to be a first-order modeling concern. With a per-period *discount rate* of 4%, and a present-bias of 0.9, the effective first-period discount rate is roughly 15%. Such a discrepancy between the effective first-period discount rate and the discount rate for subsequent periods is substantial enough that it should merit consideration when analyzing, for example, individual behavior in the workplace under different contracts (Kaur et al., 2015) or common self-control problems.

Regarding the second question—an extremely challenging and important one—we do not think our meta-analysis is capable of identifying a preferred method for eliciting present bias. The meta-regression effects can only tell us which methods produce reliably different estimates than others.

Approaching this question forces one to take a stand on the conceptual status of present bias. Is it a stable trait, and an ideal method would come as close to the true value as possible? Or does expressed present bias change according to elicitation method *and*, very likely, in different natural choice domains. The latter view is expressed by Frederick and Loewenstein (2008) who wrote (p. 233):

> *Like others [...] our findings suggest that respondents possess a variety of cognitive schemas, each of which can be evoked or suppressed by subtle contextual features. Thus, we believe that the major challenge for decision researchers lies not in honing parametric specifications, but in acquiring a broader understanding of the varied constituents of preferences and the problem representations that bring them to the fore.*

We suggest a middle path between the stable-trait view and the contextualist view. In psychometric language, measuring trait-like quantities well aspires to achieve two goals: Reliability and validity.

Reliability means that there is low measurement error. For example, test-retest reliability means that a person answers the same question the same way, if it is asked twice.[27] "Construct" validity is how well a measure is associated to a general construct.

We think that economic validity is best operationalized as good generalizability from one type of *PB* estimate to a different behavior which is thought (theoretically) to be correlated with *PB*. Ideal examples of this are studies in which laboratory or survey estimates are associated with natural behaviors, at the individual level. For example, (Meier and Sprenger, 2010) measured *PB* using CTB and found it was correlated with credit card use and debt level.[28]

In sum, we think the criteria for a preferred method are high reliability and good economic validity. Unfortunately, our current meta-analysis cannot measure either of these criteria well, but more ambitious studies linking estimation and natural-data observation can do so, and certainly should.

References

Alinaghi, N. and W. R. Reed (2018): "Meta-Analysis and Publication Bias: How Well Does the FAT-PET-PEESE Procedure Work?" *Research Synthesis Methods*, 9, 285–311.

Andreoni, J., C. Gravert, M. A. Kuhn, S. Saccardo, and Y. Yang (2018): "Arbitrage Or Narrow Bracketing? On Using Money to Measure Intertemporal Preferences," NBER Working Paper No. 25232.

Andreoni, J., M. A. Kuhn, and C. Sprenger (2015): "Measuring Time Preferences: A Comparison of Experimental Methods," *Journal of Economic Behavior & Organization*, 116, 451–464.

Andreoni, J. and C. Sprenger (2012): "Estimating Time Preferences from Convex Budgets," *American Economic Review*, 102, 3333–3356.

Andrews, I. and M. Kasy (2019): "Identification of and Correction for Publication Bias," *American Economic Review*, 109, 2766–2794.

Angeletos, G.-M., D. Laibson, A. Repetto, J. Tobacman, and S. Weinberg (2001): "The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation," *Journal of Economic Perspectives*, 15, 47–68.

[27] Preferably the retest is created with no ability to clearly recall the first answer, as in the footnote 12 discussion about artificial consistency.

[28] In a less related example, Chapman et al. (2018) compared a price list method and an adaptive method for measuring risk attitudes. They found that the adaptive method appeared to have lower measurement error and, as a result, high correlations with plausible demographic variables.

ASHRAF, N., D. KARLAN, AND W. YIN (2006): "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines," *Quarterly Journal of Economics*, 121, 635–672.

AUGENBLICK, N. (2018): "Short-Term Time Discounting of Unpleasant Tasks," Unpublished manuscript.

AUGENBLICK, N., M. NIEDERLE, AND C. SPRENGER (2015): "Working Over Time: Dynamic Inconsistency in Real Effort Tasks," *Quarterly Journal of Economics*, 130, 1067–1115.

AUGENBLICK, N. AND M. RABIN (2019): "An Experiment on Time Preference and Misprediction in Unpleasant Tasks," *Review of Economic Studies*, 86, 941–975.

BALAKRISHNAN, U., J. HAUSHOFER, AND P. JAKIELA (2017): "How Soon Is Now? Evidence of Present Bias from Convex Time Budget Experiments," NBER Working Paper No. 23558.

BARCELLOS, S. H. AND L. CARVALHO (2014): "Information about Self-Control and Intertemporal Choices," Unpublished manuscript.

BELSLEY, D. A., E. KUH, AND R. E. WELSCH (1980): *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Hoboken, NJ: Wiley.

BESHEARS, J., J. J. CHOI, C. CLAYTON, C. HARRIS, D. LAIBSON, AND B. C. MADRIAN (2017): "Optimal Illiquidity," Unpublished manuscript.

BESHEARS, J., J. J. CHOI, C. HARRIS, D. LAIBSON, B. C. MADRIAN, AND J. SAKONG (2015): "Self Control and Commitment: Can Decreasing the Liquidity of a Savings Account Increase Deposits?" NBER Working Paper No. 21474.

BISIN, A. AND K. HYNDMAN (2020): "Present-Bias, Procrastination and Deadlines in a Field Experiment," *Games and Economic Behavior*, 119, 339–357.

BOLLEN, K. A. AND R. W. JACKMAN (1985): "Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases," *Sociological Methods & Research*, 13, 510–542.

CARTER, E. C., F. D. SCHÖNBRODT, W. M. GERVAIS, AND J. HILGARD (2019): "Correcting for Bias in Psychology: A Comparison of Meta-Analytic Methods," *Advances in Methods and Practices in Psychological Science*, 2, 115–144.

CARVALHO, L. S., S. MEIER, AND S. W. WANG (2016): "Poverty and Economic Decision-Making: Evidence from Changes in Financial Resources at Payday," *American economic review*, 106, 260–284.

CHABRIS, C. F., D. I. LAIBSON, AND J. P. SCHULDT (2008): "Intertemporal Choice," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf and L. E. Blume, London: Palgrave Macmillan, 168–177.

Chapman, J., E. Snowberg, S. Wang, and C. Camerer (2018): "Loss Attitudes in the U.S. Population: Evidence from Dynamically Optimized Sequential Experimentation (DOSE)," NBER Working Paper No. 25072.

Cohen, J., K. Ericson, D. Laibson, and J. White (forthcoming): "Measuring Time Preferences," *Journal of Economic Literature*.

Coller, M. and M. B. Williams (1999): "Eliciting Individual Discount Rates," *Experimental Economics*, 2, 107–127.

Cubitt, R. P. and D. Read (2007): "Can Intertemporal Choice Experiments Elicit Time Preferences for Consumption?" *Experimental Economics*, 10, 369–389.

DellaVigna, S. and U. Malmendier (2006): "Paying Not to Go to the Gym," *American Economic Review*, 96, 694–719.

Demirgüç-Kunt, A., L. Klapper, D. Singer, S. Ansar, and J. Hess (2018): *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*, Washington, D.C.: World Bank Group.

DerSimonian, R. and N. Laird (1986): "Meta-Analysis in Clinical Trials," *Controlled Clinical Trials*, 7, 177–188.

Egger, M., G. D. Smith, M. Schneider, and C. Minder (1997): "Bias in Meta-Analysis Detected by a Simple, Graphical Test," *BMJ*, 315, 629–634.

Ericson, K. M. and D. Laibson (2019): "Intertemporal Choice," in *Handbook of Behavioral Economics - Foundations and Applications*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, Amsterdam: Elsevier/North-Holland, vol. 2, 1–67.

Fang, H. and Y. Wang (2015): "Estimating Dynamic Discrete Choice Models with Hyperbolic Discounting, with an Application to Mammography Decisions," *International Economic Review*, 56, 565–596.

Fedyk, A. (2018): "Asymmetric Naivete: Beliefs about Self-Control," Unpublished manuscript.

Frederick, S. and G. Loewenstein (2008): "Conflicting Motives in Evaluations of Sequences," *Journal of Risk and Uncertainty*, 37, 221–235.

Frederick, S., G. Loewenstein, and T. O'donoghue (2002): "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature*, 40, 351–401.

Furukawa, C. (2019): "Publication Bias under Aggregation Frictions: Theory, Evidence, and a New Correction Method," Unpublished manuscript.

Gelman, A. and J. Hill (2007): *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge: Cambridge University Press.

Glass, G. V. (1976): "Primary, Secondary, and Meta-Analysis of Research," *Educational Researcher*, 5, 3–8.

Gruber, J. and B. Kőszegi (2001): "Is addiction "rational"? Theory and evidence," *Quarterly Journal of Economics*, 116, 1261–1303.

Gurevitch, J., J. Koricheva, S. Nakagawa, and G. Stewart (2018): "Meta-Analysis and the Science of Research Synthesis," *Nature*, 555, 175–182.

Halevy, Y. (2014): "Some Comments on the Use of Monetary Rewards in the Measurement of Time Preferences," Unpublished manuscript.

Havránek, T., R. Horváth, Z. Iršová, and M. Rusnak (2015): "Cross-Country Heterogeneity in Intertemporal Substitution," *Journal of International Economics*, 96, 100–118.

Havránek, T., M. Rusnak, and A. Sokolova (2017): "Habit Formation in Consumption: A Meta-Analysis," *European Economic Review*, 95, 142–167.

Hedges, L. V., E. Tipton, and M. C. Johnson (2010): "Robust Variance Estimation in Meta-Regression with Dependent Effect Size Estimates," *Research Synthesis Methods*, 1, 39–65.

Heidhues, P. and B. Kőszegi (2010): "Exploiting Naïvete about Self-Control in the Credit Market," *American Economic Review*, 100, 2279–2303.

Higgins, J. P. T. and S. G. Thompson (2002): "Quantifying Heterogeneity in a Meta-Analysis," *Statistics in Medicine*, 21, 1539–1558.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999): "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14, 382–401.

Holt, C. A. and S. K. Laury (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644–1655.

Hong, S. and W. R. Reed (2019): "A Performance Analysis of Some New Meta-Analysis Estimators Designed to Correct Publication Bias," Unpublished manuscript.

Imai, T. and C. F. Camerer (2018): "Estimating Time Preferences from Budget Set Choices Using Optimal Adaptive Design," Unpublished manuscript.

Imai, T., T. Rutter, and C. F. Camerer (2018): "Meta-Analysis of Estimation of Time Discounting of Rewards," Unpublished manuscript.

Iršová, Z. and T. Havránek (2013): "Determinants of Horizontal Spillovers from FDI: Evidence from a Large Meta-Analysis," *World Development*, 42, 1–15.

John, A. (forthcoming): "When Commitment Fails – Evidence from a Field Experiment," *Man-*

agement Science.

Kaur, S., M. Kremer, and S. Mullainathan (2010): "Self-Control and the Development of Work Arrangements," *American Economic Review: Papers & Proceedings*, 100, 624–28.

——— (2015): "Self-Control at Work," *Journal of Political Economy*, 123, 1227–1277.

Konstantopoulos, S. (2011): "Fixed Effects and Variance Components Estimation in Three-level Meta-Analysis," *Research Synthesis Methods*, 2, 61–76.

Koopmans, T. C. (1960): "Stationary Ordinal Utility and Impatience," *Econometrica*, 28, 287–309.

Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112, 443–478.

Liu, E. M., J. Meng, and J. T.-y. Wang (2014): "Confucianism and Preferences: Evidence from Lab Experiments in Taiwan and China," *Journal of Economic Behavior & Organization*, 104, 106–122.

Matoušek, J., T. Havránek, and Z. Iršová (2019): "Individual Discount Rates: A Meta-Analysis of Experimental Evidence," Unpublished manuscript.

Meager, R. (2019): "Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments," *American Economic Journal: Applied Economics*, 11, 57–91.

Meier, S. and C. Sprenger (2010): "Present-Biased Preferences and Credit Card Borrowing," *American Economic Journal: Applied Economics*, 2, 193–210.

Moral-Benito, E. (2015): "Model Averaging in Economics: An Overview," *Journal of Economic Surveys*, 29, 46–75.

Mulligan, C. (1996): "A Logical Economist's Argument against Hyperbolic Discounting," Unpublished manuscript.

O'Donoghue, T. and M. Rabin (1999): "Doing It Now or Later," *American Economic Review*, 89, 103–124.

——— (2001): "Choice and Procrastination," *Quarterly Journal of Economics*, 116, 121–160.

——— (2015): "Present Bias: Lessons Learned and to be Learned," *American Economic Review*, 105, 273–279.

Phelps, E. S. and R. A. Pollak (1968): "On Second-Best National Saving and Game-Equilibrium Growth," *Review of Economic Studies*, 35, 185–199.

Samuelson, P. A. (1937): "A Note on Measurement of Utility," *Review of Economic Studies*, 4, 155–161.

Stanley, T. D. (2001): "Wheat from Chaff: Meta-analysis as Quantitative Literature Review," *Journal of Economic Perspectives*, 15, 131–150.

——— (2005): "Beyond Publication Bias," *Journal of Economic Surveys*, 19, 309–345.

——— (2008): "Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection," *Oxford Bulletin of Economics and Statistics*, 70, 103–127.

——— (2017): "Limitations of PET-PEESE and Other Meta-Analysis Methods," *Social Psychological and Personality Science*, 8, 581–591.

Stanley, T. D. and H. Doucouliagos (2012): *Meta-Regression Analysis in Economics and Business*, London: Routledge.

——— (2014): "Meta-Regression Approximations to Reduce Publication Selection Bias," *Research Synthesis Methods*, 5, 60–78.

——— (2015): "Neither Fixed nor Random: Weighted Least Squares Meta-Analysis," *Statistics in Medicine*, 34, 2116–2127.

——— (2017): "Neither Fixed nor Random: Weighted Least Squares Meta-Regression," *Research Synthesis Methods*, 8, 19–42.

Stanley, T. D., S. B. Jarrell, and H. Doucouliagos (2010): "Could It Be Better to Discard 90% of the Data? A Statistical Paradox," *The American Statistician*, 64, 70–77.

Steel, M. F. J. (forthcoming): "Model Averaging and its Use in Economics," *Journal of Economic Literature*.

Van den Noortgate, W., J. A. López-López, F. Marín-Martínez, and J. Sánchez-Meca (2013): "Three-Level Meta-Analysis of Dependent Effect Sizes," *Behavior Research Methods*, 45, 576–594.

Supplementary Online Material

Meta-Analysis of Present-Bias Estimation Using Convex Time Budgets

Taisuke Imai Tom A. Rutter Colin F. Camerer

Contents

A The Convex Time Budget Protocol

Idea. Consider two time points t ("sooner") and $t+k$ ("later"). A linear budget set of allocations of monetary rewards to be received at those two times is a line connecting two points $(\bar{x}_t, 0)$ and $(0, \bar{x}_{t+k})$ on a two-dimensional plane. The first point corresponds to an agent receiving a certain amount $\bar{x}_t$ of reward at time t and nothing at $t+k$. The second point corresponds to receiving a certain amount $\bar{x}_{t+k}$ at time $t+k$ and nothing at t. Any points on the interior of a budget set represent allocations in which she receives positive rewards on both dates.

Figure A.1 illustrates two such budgets and choices from those budgets, marked as B^i and x^i, $i = a, b$. The *slopes* of budget lines represent intertemporal tradeoffs between rewards at two time points (reflecting an implicit interest rate). This kind of budget-line figure appears in every microeconomics textbook, typically showing a budget line in two-good space and a family of continuous iso-utility indifference curves for bundles of goods in that space.

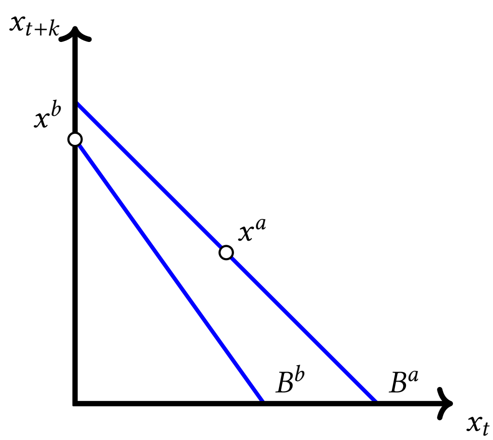


FIGURE A.1: An illustration of linear budget sets which ask allocations of monetary rewards to be received at dates t and $t+k$. A hypothetical subject chose allocation x^a from budget B^a, from which the subject receives positive amount on both dates t and $t+k$. On the other hand, the subject receives positive amount only on date $t+k$ (and nothing on date t) from allocation x^b.

In order to identify and estimate parameters of different kinds of time preferences, an experimenter needs to vary the time points $(t, t+k)$, the slopes of the budget lines, and the level of the budget lines. Each budget line can be expressed as a set of these numbers.

Implementation. There are two main approaches to implement the CTB protocol. In the first approach, subjects make allocation decisions. For example, in the original Andreoni and Sprenger (2012) experiment, subjects are endowed with 100 tokens which they allocate to "sooner" and "later" tokens. Each account is associated with an exchange rate, which converts tokens into monetary amounts. When the exchange rates are (e_t, e_{t+k}), allocating (a_t, a_{t+k}) tokens to two accounts implies monetary rewards of $(a_t \times e_t, a_{t+k} \times e_{t+k})$. The ratio of exchange rates e_{t+k}/e_t is the k-period gross interest rate. Many computerized experiments in the laboratory follow this approach. In the second approach, used first in Andreoni et al. (2015), subjects select a reward schedule (x_t, x_{t+k}) from a set of options (typically less than 10) that are evenly spaced on the budget line.

Econometric Strategy. Consider quasi-hyperbolic discounting with a constant relative risk aversion (CRRA) utility function of the form:

$$U(x_t, x_{t+k}) = \frac{1}{\alpha}(x_t + \omega_t)^\alpha + \beta^{\mathbf{1}\{t=0\}}\delta^k \frac{1}{\alpha}(x_{t+k} + \omega_{t+k})^\alpha, \tag{A.1}$$

where δ is the per-period discount factor, β is the present bias, α is the curvature parameter, and ω_t and ω_{t+k} are background consumption parameters. Maximizing (A.1) subject to an intertemporal budget constraint

$$(1 + r)x_t + x_{t+k} = I,$$

where $1 + r$ is the gross interest rate (over k days) and I is the budget, yields an intertemporal Euler equation

$$\frac{x_t + \omega_t}{x_{t+k} + \omega_{t+k}} = \left(\beta^{\mathbf{1}\{t=0\}}\delta^k(1 + r)\right)^{\frac{1}{\alpha-1}}.$$

Andreoni and Sprenger (2012) propose two methods for estimating parameters (α, β, δ). The first one estimates the parameters in the log-linearized version of the Euler equation

$$\log\left(\frac{x_t + \omega_t}{x_{t+k} + \omega_{t+k}}\right) = \frac{\log\beta}{\alpha - 1} \cdot \mathbf{1}\{t = 0\} + \frac{\log\delta}{\alpha - 1} \cdot k + \frac{1}{\alpha - 1} \cdot \log(1 + r), \tag{A.2}$$

using two-limit Tobit regression in order to handle corner solutions under an additive error structure. The second one estimates the parameters in the optimal demand for sooner consumption

$$\begin{aligned} x_t = {} & \left(\frac{1}{1 + (1 + r)(\beta^{\mathbf{1}\{t=0\}}\delta^k(1 + r))^{1/(\alpha-1)}}\right)\omega_t \\ & + \left(\frac{(\beta^{\mathbf{1}\{t=0\}}\delta^k(1 + r))^{1/(\alpha-1)}}{1 + (1 + r)(\beta^{\mathbf{1}\{t=0\}}\delta^k(1 + r))^{1/(\alpha-1)}}\right)(I + \omega_{t+k}), \end{aligned} \tag{A.3}$$

using Nonlinear Least Squares (NLS). In either case, parameters (α, β, δ) are recovered via a nonlinear combination of the estimated coefficients.

Econometric strategies used in effort CTB experiments follow a similar idea and are discussed in detail in Augenblick et al. (2015).

QHD prediction. What kind of behavior do we expect to see in the CTB protocol if an agent has present-biased preferences? Consider a QHD-CRRA utility function (A.1). Given a preference parameter combination (α, β, δ), we can calculate the optimal allocation (x_t^*, x_{t+k}^*) for each budget $((t,k), 1+r)$ using equation (A.3). We assume that background consumption parameters (ω_t, ω_{t+k}) are zeroes for all (t,k) for simplicity.

Figure A.2 presents predicted allocations for five budgets under time frame $(t,k) = (0, 35)$. We fix utility curvature and discount factor at $(\alpha, \delta) = (0.897, 0.999)$, a pair of the aggregate estimates reported in Andreoni and Sprenger (2012), while varying present-biasedness $\beta \in \{0.9, 1.0, 1.1\}$. We can first observe in panel B that the agent allocates fewer tokens to the sooner payment date as the gross interest rate (over k period) increases (i.e., the price of the sooner reward becomes higher).

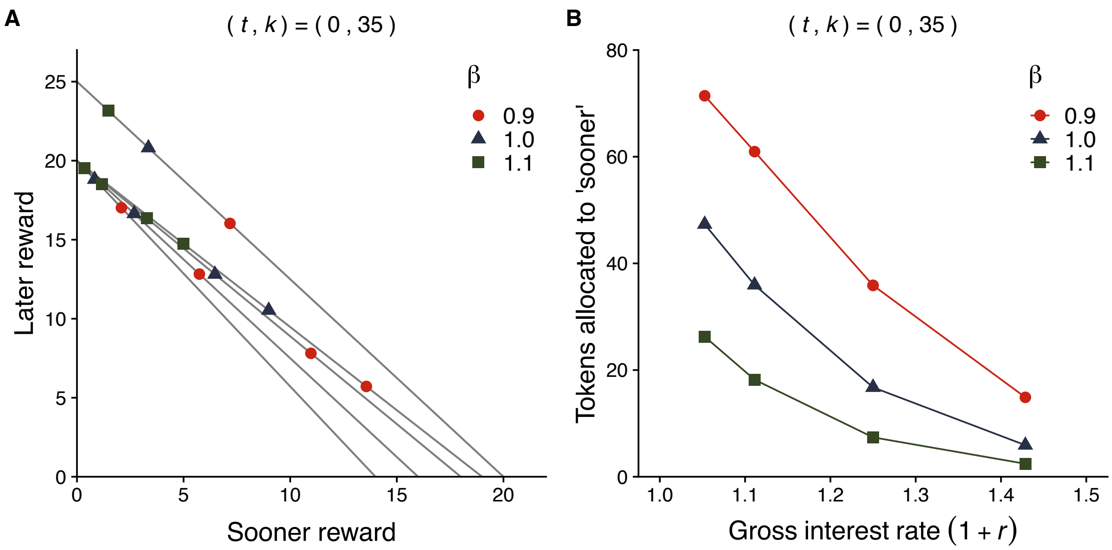


FIGURE A.2: Predicted allocation decisions.

Note that β does not play a role when the sooner payment date is also in the future (i.e., $t > 0$). It implies that predicted allocations for a $\beta \neq 1$ agent coincide with those for the time-consistent, $\beta = 1$, agent. Holding everything else constant, a present-biased ($\beta < 1$) agent allocates more to the sooner payment date when the sooner payment date is the current period ($t = 0$) than when both payment dates are in the future ($t > 0$). A future-biased ($\beta > 1$) agent exhibits the opposite pattern.

B Data

B.1 Identification and Selection Procedure

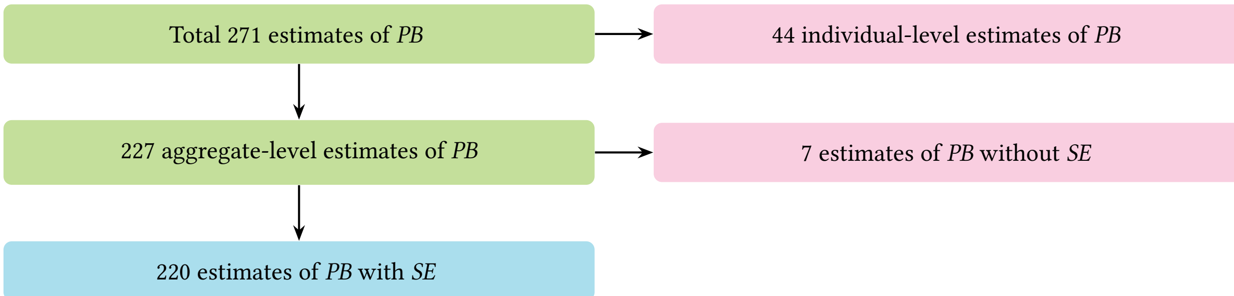


FIGURE B.1: Types of *PB* estimates in the dataset.

B.2 Summary of Included Papers

Table B.1: List of articles using the CTB protocol (with QHD parameter estimates).

#	Article	Country	Location	Subject	Reward	Delivery	Interface	# budgets	# options	# frame
1	Abebe et al. (2017)	Ethiopia	Field	General	Effort	Other	Paper and pencil	10	51	2
2	Andreoni and Sprenger (2012)	USA	Lab	Students	Money	Bank	Computer	45	101	9
3	Andreoni et al. (2015)	USA	Lab	Students	Money	Bank	Paper and pencil	24	6	4
4	Andreoni et al. (2017)	Pakistan	Field	General	Effort		Computer	8	277	2
5	Ashton (2015)	USA	Lab	Students	Money	Paypal	Computer	55	101	9
6	Augenblick et al. (2015)	USA	Lab	Students	Money	Cash	Computer	20		3
7	Aycinena and Rentschler (2018)	Guatemala	Field	General	Money	Mix	Paper and pencil	24	6	4
8	Aycinena et al. (2015)	Guatemala	Field	General	Money	Gift card	Computer	24	6	4
9	Balakrishnan et al. (2017)	Kenya	Lab	General	Money	Gift card	Computer	48		6
10	Banerji et al. (2018)	India	Field	General	Money	Check	Paper and pencil	24	5	4
11	Barcellos and Carvalho (2014)	USA	Online	General	Money	Bank	Computer	6	501	2
12	Barton (2015)	USA	Lab	Students	Effort		Computer	34	91	2
13	Bartoš et al. (2018)	Uganda	Field	General	Effort		Paper and pencil	10	6	2
14	Boonmanunt et al. (2018)	Thailand	Field	General	Money	Check	Paper and pencil	15	4	3
15	Bousquet (2016)	France	Lab	Students	Money	Paypal	Computer	40	21	10
16	Brocas et al. (2018)	USA	Lab	Students	Money	Gift card	Paper and pencil	45	11	9
17	Carvalho et al. (2016a)	USA	Online	General	Money	Bank	Computer	12	501	3
18	Carvalho et al. (2016b)	Nepal	Field	General	Money	Bank	Paper and pencil	4	3	3
19	Cerrone and Lades (2017)	UK	Lab	Students	Money	Mix	Computer	24	6	4
20	Chen et al. (2019)	China	Field	General	Money	Mix	Paper and pencil	24	6	4
21	Clot et al. (2017)	Uganda	Field	General	Money		Paper and pencil	15	3	3
22	Corbett (2016)	USA	Online	Students	Effort		Computer	9	11	3
23	Hvide and Lee (2016)	UK	Lab	Students	Money	Gift card	Paper and pencil	36	6	6
24	Imai and Camerer (2018)	USA	Online	General	Money		Computer	20	101	5
25	Imas et al. (2018)	USA	Online	General	Effort		Computer	8	11	2
26	Janssens et al. (2017)	Nigeria	Field	General	Money	Check	Paper and pencil	8	11	2
27	Kölle and Wenner (2018)	Germany	Lab	Students	Effort		Computer	12	51	2
28	Kuhn et al. (2017)	France	Lab	Students	Money	Gift card	Computer	45	17	9
29	Lindner and Rose (2017)	Germany	Lab	Students	Money	Gift card	Computer	24	6	4
30	Liu et al. (2014)	Taiwan; China	Lab	Students	Money	Gift card	Paper and pencil	10	61	2
31	Lührmann et al. (2018)	Germany	Class	Kids	Money	Check	Paper and pencil	21	4	3
32	Sawada and Kuroishi (2015a)	Philippines	Field	General	Money		Paper and pencil	24	5	4
33	Sawada and Kuroishi (2015b)	Japan	Field	General	Money		Paper and pencil	24		
34	Stango et al. (2017)	USA	Online	General	Money		Computer	24	101	4
35	Sun and Potters (2016)	Netherlands	Lab	Students	Money	Gift card	Computer	14		2
36	Yang and Carlsson (2016)	China	Field	General	Money		Paper and pencil	10	21	2

Note: Sun and Potters (2016) varied the number of tokens (i.e., number of options; 101, 201, 301, 401, 801) to manipulate the magnitude.

Table B.2: List of articles using the CTB protocol (without QHD parameter estimates).

#	Article	Country	Location	Subject	Reward	Delivery	Interface	# budgets	# options	# frame
37	Alan and Ertac (2015)	Turkey	Field	Kids	Other	Other	Paper and pencil	4	6	2
38	Alan and Ertac (2018)	Turkey	Field	Kids	Other	Other	Paper and pencil	4	6	2
39	Alan and Ertac (2017)	Turkey	Field	Kids	Other	Other	Paper and pencil	1	6	1
40	Andreoni and Sprenger (2012b)	USA	Lab	Students	Money	Bank	Computer	14	101	2
41	Andreoni et al. (2018)	USA	Lab	Students	Money	Gift card	Computer	16	8	2
42	Angerer et al. (2015)	Italy	Field	Kids	Other	Other	Paper and pencil	1	6	1
43	Atalay et al. (2014)	USA	Online	General	Money		Computer	9	5	
44	Batista et al. (2015)	Mozambique	Field	General	Money	Gift card	Paper and pencil	10	21	2
45	Blumenstock et al. (2018)	Afghanistan	Field	General	Money	Gift card	Paper and pencil	10	3	2
46	Bover et al. (2018)	Spain	Field	Kids	Other	Other	Paper and pencil	9	4	3
47	Bulte et al. (2016)	Vietnam	Field	General	Money		Paper and pencil	20		4
48	Cheung (2015)	Australia	Lab	Students	Money	Bank	Paper and pencil	14	101	2
49	Clot and Stanton (2014)	Uganda	Field	General	Money		Paper and pencil	10	3	2
50	de Oliveira and Jacobson (2017)	USA	Lab	Students	Effort		Computer	10	61	1
51	de Quidt et al. (2018)	USA	Online	General	Money	Other	Computer	1		1
52	Dertwinkel-Kalt et al. (2017)	Germany	Lab	Students	Money	Gift card	Paper and pencil	36	101	
53	Ersoy (2017)	USA	Online	Students	Money	Other	Computer	24	5	4
54	Esopo et al. (2018)	Kenya	Lab	General	Money	Gift card	Paper and pencil			
55	Franco and Mahadevan (2017)	Colombia	Lab	Students	Money		Computer	16	51	4
56	Giné et al. (2018)	Malawi	Field	General	Money	Cash	Paper and pencil	10	21	2
57	Grijalva et al. (2018)	USA	Lab	Students	Money		Paper and pencil	36	101	4
58	Hoel et al. (2016)	Ethiopia	Lab	Students	Money	Cash	Paper and pencil	12	6	1
59	Mayer et al. (2015)	USA	Field	General	Money		Computer	15	4	3
60	Miao and Zhong (2015)	Singapore	Lab	Students	Money	Bank	Paper and pencil	14	101	2
61	Mudzingiri (2017)	South Africa	Lab	Students	Money	Gift card	Paper and pencil	1	6	1
62	Penczynski and Santana (2016)	Philippines	Field	General	Money	Check	Paper and pencil	18	16	2
63	Potters et al. (2016)	Netherlands	Online	General	Money	Gift card	Paper and pencil	40	11	2
64	Rong et al. (2018)	USA	Lab	General	Money	Bank	Computer	9	101	1
65	Savani et al. (2018)	Singapore	Lab	Students	Money	Gift card	Paper and pencil	40	51	
66	Slonim et al. (2013)	Australia	Lab	Students	Money		Paper and pencil	6	6	2
67	Sutter et al. (2018)	Italy	Field	Kids	Other	Other	Paper and pencil	1	6	1

Table B.3: List of articles with some treatment variations.

Study	Treatment dimension
Abebe et al. (2017)	Incentive size
Alan and Ertac (2017)	Degree of optimism
Alan and Ertac (2018)	Educational intervention
Andreoni et al. (2018)	Salience of arbitrage
Ashton (2015)	Fatigue and hunger
Atalay et al. (2014)	Availability of a prize-linked savings account
Aycinena and Rentschler (2018)	Payoff display
Balakrishnan et al. (2017)	"Immediate" reward delivery timing
Bartoš et al. (2018)	Poverty priming
Bulte et al. (2016)	Male partner invited to join the training or not
Carvalho et al. (2016a)	Payday timing
Carvalho et al. (2016b)	Bank account assignment
Chen et al. (2019)	Hunger
Cheung (2015)	Probability of reward
Hoel et al. (2016)	Self-control fatigue
Hvide and Lee (2016)	Windfall or hard-earned money
Imai and Camerer (2018)	Budget set construction, fixed, random, or adaptive
Kuhn et al. (2017)	Cognitive resource depletion
Lindner and Rose (2017)	Time pressure
Liu et al. (2014)	Confucius priming
Lührmann et al. (2018)	Financial education
Penczynski and Santana (2016)	Future payment by microfinance or local money lender
Potters et al. (2016)	Stakes, time horizon, and frame
Yang and Carlsson (2016)	Separate or joint decision by couples

B.3 Coded Variables

TABLE B.4: List of coded variables.

Variable	Description
Atricle meta data	
`main.lastnames`	Last names of the authors
`main.firstnames`	First names of the authors
`main.title`	Title of the paper
`main.published`	1 if published or "in press"; 0 if unpublished; 9 if "do not circulate"
`main.yearpub`	Year of publication
`main.monthpub`	Month of publication
`main.journal`	Journal
`main.unpub.year`	Year this version was written (for unpublished papers)
`main.unpub.month`	Month this version was written (for unpublished papers)
`main.unpub.day`	Day this version was written (for unpublished papers)
`main.length`	Number of pages (main content; excluding appendices)
`main.length.appendix`	Number of pages (online appendices)
`main.affliations`	Affiliations of the authors
`main.fund`	Funding sources
`main.data.available`	1 if data is publicly available
`main.instructions`	1 if instructions available
Additional info about published article	
`pub.topfive`	1 if published in Top 5 (*AER*, *ECMA*, *JPE*, *QJE*, *REStud*)
`pub.firstyear`	Year of the first draft (or the oldest version identified)
Treatment and sample	
`treatment.neutral`	1 if control / neutral treatment
`treatment.nonneutral`	1 if some treatment variation
`treatment.dimension`	Description of treatment
`sample.all`	1 if estimation is based on all sample
`sample.sub`	1 if estimation is based on subsample
`sample.dimension`	Description of subsample

Variable	Description
Location of the experiment	
`location.lab`	1 if laboratory experiment
`location.field`	1 if field experiment
`location.amt`	1 if Amazon Mechanical Turk
`location.class`	1 if classroom experiment
`location.survey`	1 if online survey
`location.continent`	Continent
`location.country`	Country
`location.city`	City
`location.state`	State
Method	
`method.numbudget`	Number of budget lines
`method.numoption`	Number of available options on each budget
`method.corner`	1 if corners of the budget are available
`method.calendar`	1 if calendar is presented
`method.computer`	1 if computer interface was used; 0 if paper and pencil
`method.input`	1 if subjects entered desired allocation
`method.checkbox`	1 if subjects marked/clicked an option
`method.slider`	1 if subjects made an allocation decision by a slider
`method.physical`	1 if subjects allocated physical objects (e.g. marbles)
`method.timelimit`	Time limit (in second) in each decision
Time frame and budgets	
`ctb.time.unit`	Time unit for (t, k)
`ctb.sooner`	Potential sooner payment dates
`ctb.delay`	Potential delay length
`ctb.grossint`	Gross interest rate over k periods
`ctb.num.sooner`	Number of potential sooner payment dates (t)
`ctb.num.delay`	Number of potential delay length (k)
`ctb.num.frame`	Number of time frames (i.e. (t, k) pairs)
`ctb.num.slope`	Number of budget slopes (gross interest rates over k periods)

Variable	Description
Reward	
reward.real	1 if real reward
reward.certain	1 if all payments are certain
reward.risky	1 if payment risk is introduced (not about "random incentive system")
reward.correlated.risk	1 if payment risk is realized in a single lottery
reward.money	1 if monetary reward
reward.food	1 if food reward
reward.effort	1 if effort cost
reward.other	1 if other type of reward
Delivery of future reward	
delivery.pickup	1 if subjects came back to the lab to pickup reward
delivery.cash	1 if payments were made by cash
delivery.check	1 if payments were made by checks
delivery.paypal	1 if payments were made by PayPal
delivery.giftcard	1 if paymentts were made by gift card (e.g. Amazon)
delivery.bank	1 if payments were made by bank transfer
delivery.other	1 if other reward delivery method
delivery.notreported	1 if delivery method is not explained (or cannot be guessed)
Unit of time period presented	
time.minute	1 if time unit presented is "minute"
time.hour	1 if time unit presented is "hour"
time.day	1 if time unit presented is "day"
time.week	1 if time unit presented is "week"
time.month	1 if time unit presented is "month"
time.year	1 if time unit presented is "year"
time.mix	1 if time unit presented is mixture of the above
time.notreported	1 if time unit is not explained (or cannot be guessed)
Definition of "now"	
now.fedelay	1 if front-end-delay is introduced
now.mixed	1 if some choices involve "now" and some other don't
now.choice	1 if "now" payment is delivered right after choice
now.end	1 if "now" payment is delivered at the end of the experiment
now.day	1 if "now" payment is delivered within the same day of the experiment
now.notreported	1 if "now" payment timing is not explained

Variable	Description
Implementation	
`imp.deal.uncertainty`	1 if deal with uncertainty about future payment; 0 if not mentioned
`imp.deal.transactioncost`	1 if trying to equalize transaction costs; 0 if not mentioned
Subject pool	
`subject.child`	1 if subjects are children
`subject.teen`	1 if subjects are teenagers
`subject.university`	1 if subjects are university students
`subject.elderly`	1 if elderly population
`subject.gen`	1 if general population
`subject.farm`	1 if subjects are farmers
`subject.age.min`	Minimum age
`subject.age.max`	Maximum age
`subject.age.mean`	Mean age
`subject.age.median`	Median age
`subject.age.sd`	Standard deviation of age
`subject.male`	Fraction of male participants
Utility specifications	
`spec.u.est`	1 if utility curvature is simultaneously estimated
`spec.u.imputed`	1 if utility curvature is imputed by some other measure
`spec.u.crra`	1 if CRRA
`spec.u.cara`	1 if CARA
`spec.u.convex.effort`	1 if convex cost of effort utility
`spec.u.other`	1 if other functional form of u is assumed

Variable	Description
Estimation methods	
`est.ols`	1 if ordinary least squares
`est.nls`	1 if nonlinear least squares
`est.max.likelihood`	1 if Max Likelihood estimation
`est.tobit`	1 if Tobit regression
`est.mlogit`	1 if multinomial logit regression
`est.temperature`	1 if noise (temperature) parameter is estimated in logit specification
`est.invtemperature`	1 if noise (inverse temperature) parameter is estimated in logit specification
`est.fechner`	1 if noise (Fechner) parameter is estimated
`est.trembling`	1 if noise (trembling hand) parameter is estimated
`est.bgcons.fixed`	1 if background consumption is not fixed at zero
`est.bgcons.param`	1 if background consumption is estimated jointly with other parameters
`est.bgcons.sooner`	Level of background consumption for sooner period
`est.bgcons.later`	Level of background consumption for later period
`est.bgcons.sooner.se`	Standard error of estimated b.g. consumption for sooner period
`est.bgcons.later.se`	Standard error of estimated b.g. consumption for later period
`est.bgcons.same`	1 if sooner b.g. cons = later b.g. cons assumed
`est.bgcons.same.se`	Standard error of estimated b.g. consumption (sooner = later)
`est.bgcons.ind.report`	1 if background consumption is based on subject's report

Variable	Description
Aggregate results	
`ares.present`	1 if aggregate estimates is reported
`ares.units.discount`	Time unit for QHD model
`ares.drate`	Estimated discount rate
`ares.drate.error`	Standard error of estimated discount rate
`ares.dfactor`	Estimated discount factor
`ares.dfactor.error`	Standard error of estimated discount factor
`ares.pbias`	Estimated present bias
`ares.pbias.error`	Standard error of estimated present bias
`ares.ucurv`	Estimated utility curvature
`ares.ucurv.error`	Standard error of estimated utility curvature
`ares.convex.effort`	Estimated convex effort cost function
`ares.convex.effort.se`	Standard error of estimated convex effort cost function
`ares.temperature`	Estimated temperature parameter
`ares.temperature.error`	Standard error of estimated temperature parameter
`ares.invtemperature`	Estimated inverse temperature parameter
`ares.invtemperature.error`	Standard error of estimated inverse temperature parameter
`ares.fechner`	Estimated Fechner noise parameter
`ares.fechner.error`	Standard error of estimated Fechner noise parameter
`ares.trembling`	Estimated trembling hand parameter
`ares.trembling.error`	Standard error of estimated trembling hand parameter
`ares.rsquared`	(Adjusted) R-squared from regression
`ares.loglikelihood`	Log likelihood

C Additional Results

C.1 Funnel Plot

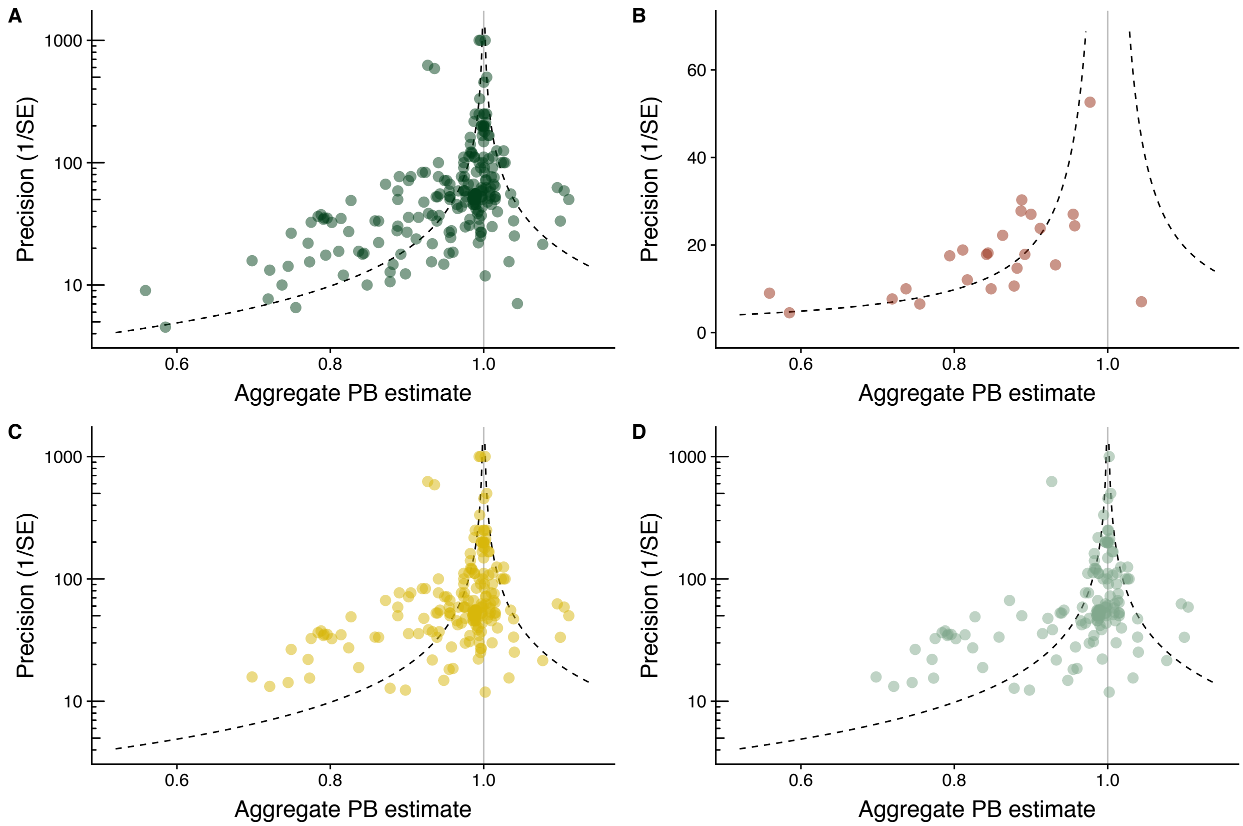


Figure C.1: Funnel plot of present bias parameter estimates *PB*. The y-axis is presented in the log-scale in the right panel. (A) All observations ($n = 220$). (B) Studies with real-effort task ($n = 24$). (C) Studies with monetary reward ($n = 196$). (D) Studies with monetary reward, neutral condition ($n = 142$). *Note*: The y-axis is presented in the log-scale except in panel B.

C.2 Study and Design Characteristics

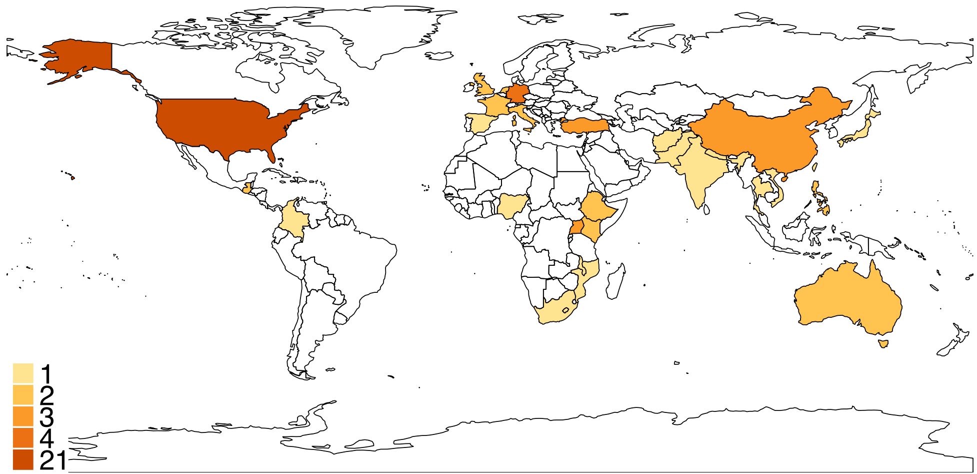


FIGURE C.2: Number of studies by country.

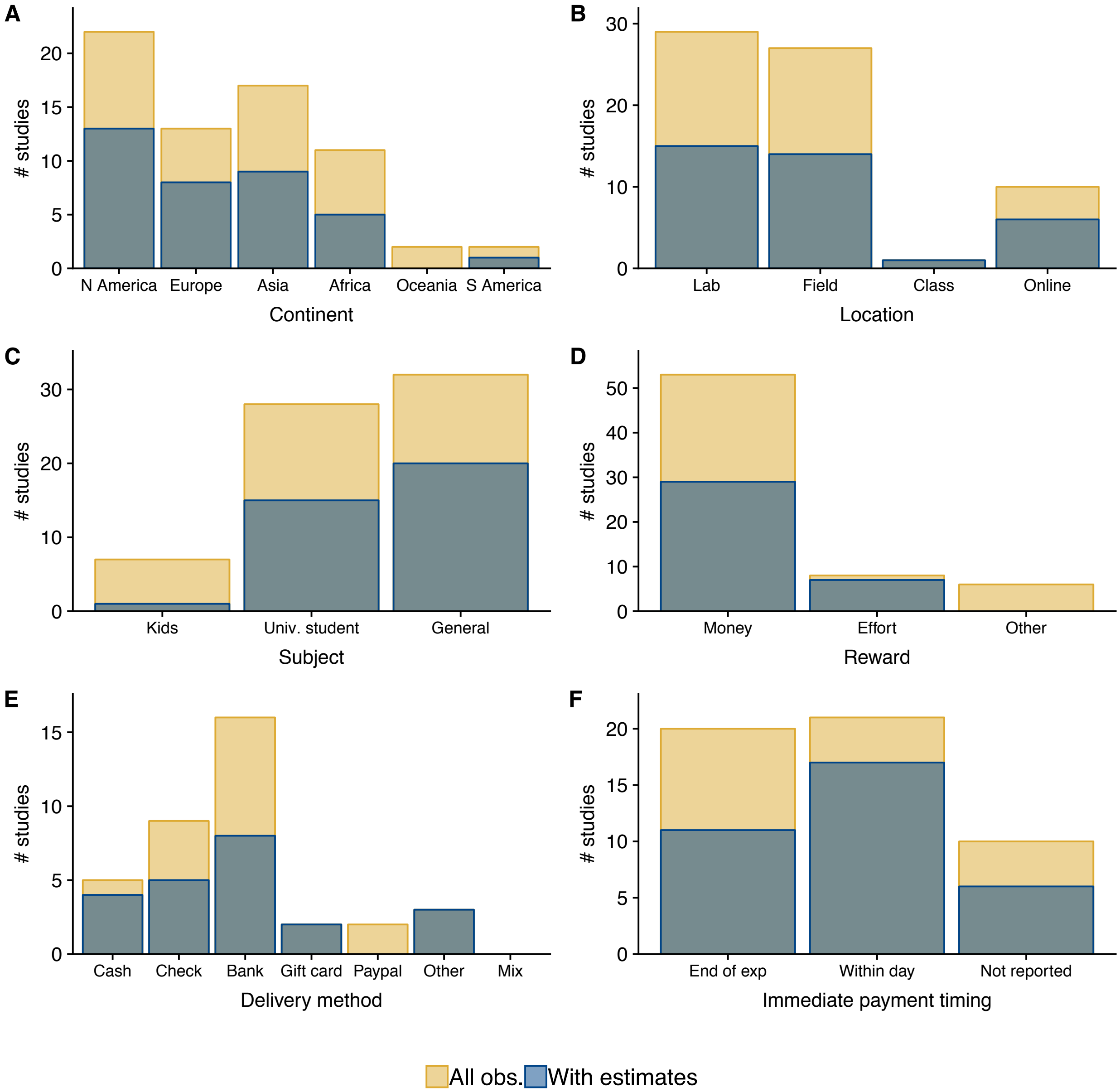


FIGURE C.3: CTB design characteristics.

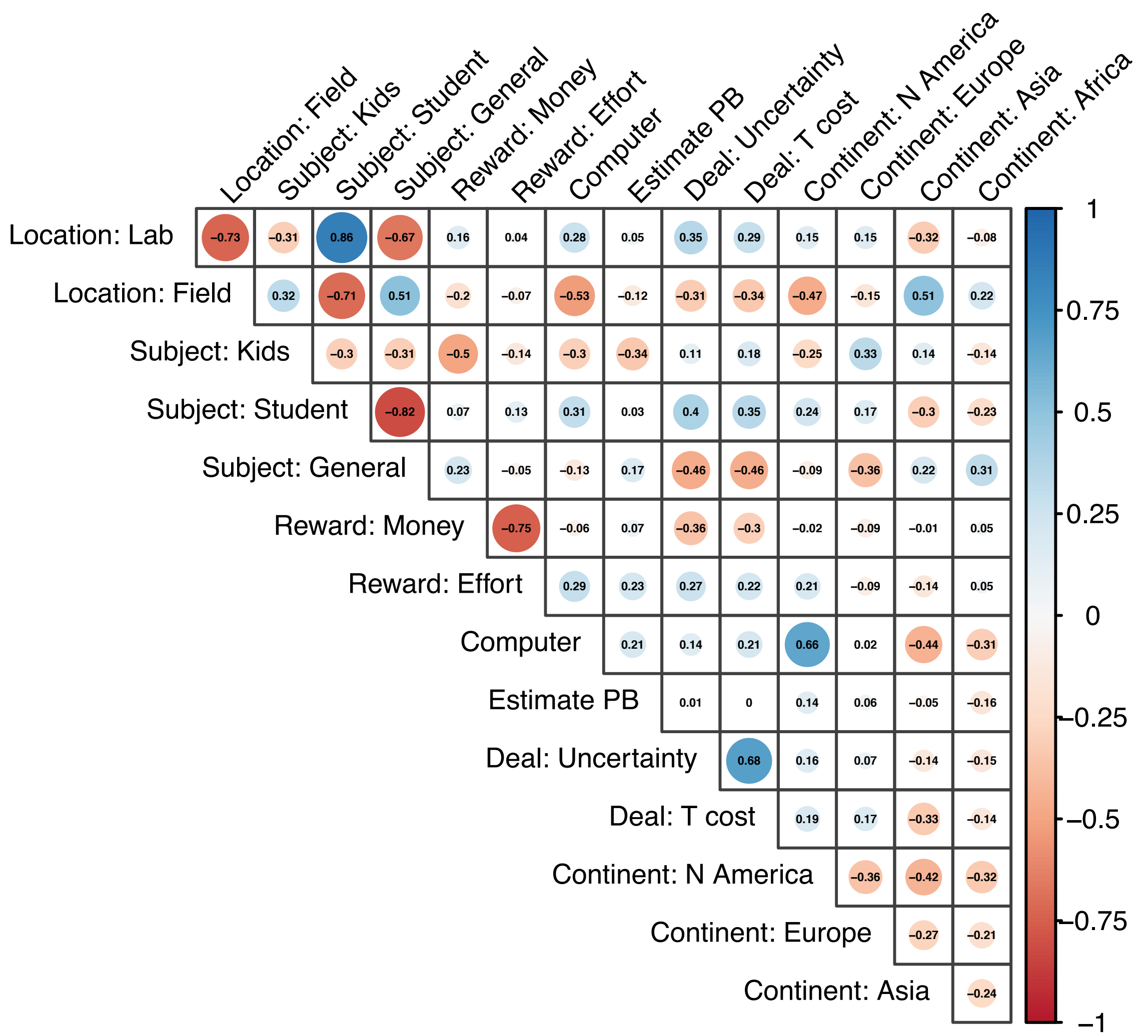


FIGURE C.4: Co-occurences of CTB design characteristics. Study-level data, with and without parameter estimates.

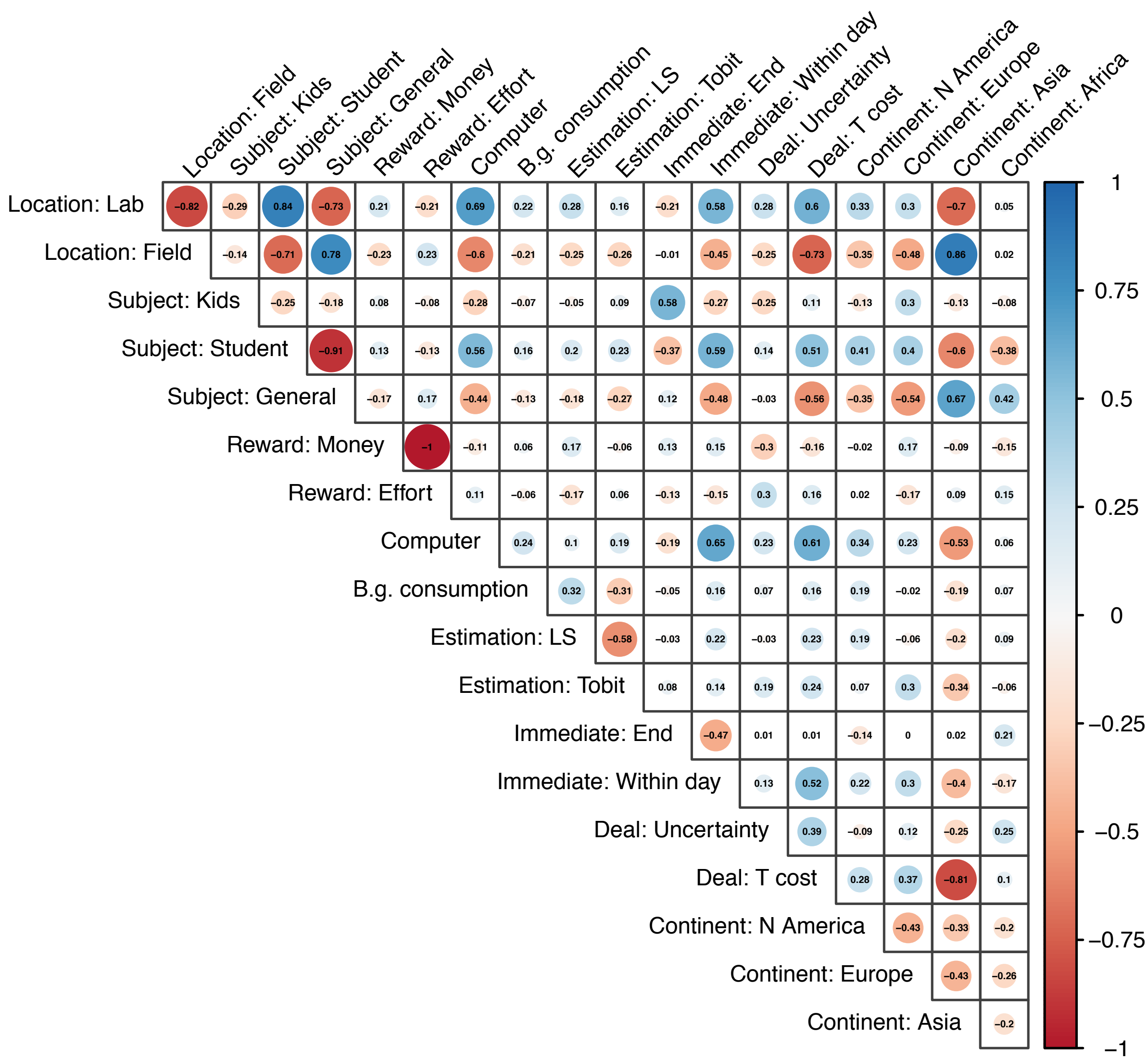


FIGURE C.5: Co-occurences of CTB design characteristics. Estimate-level data.

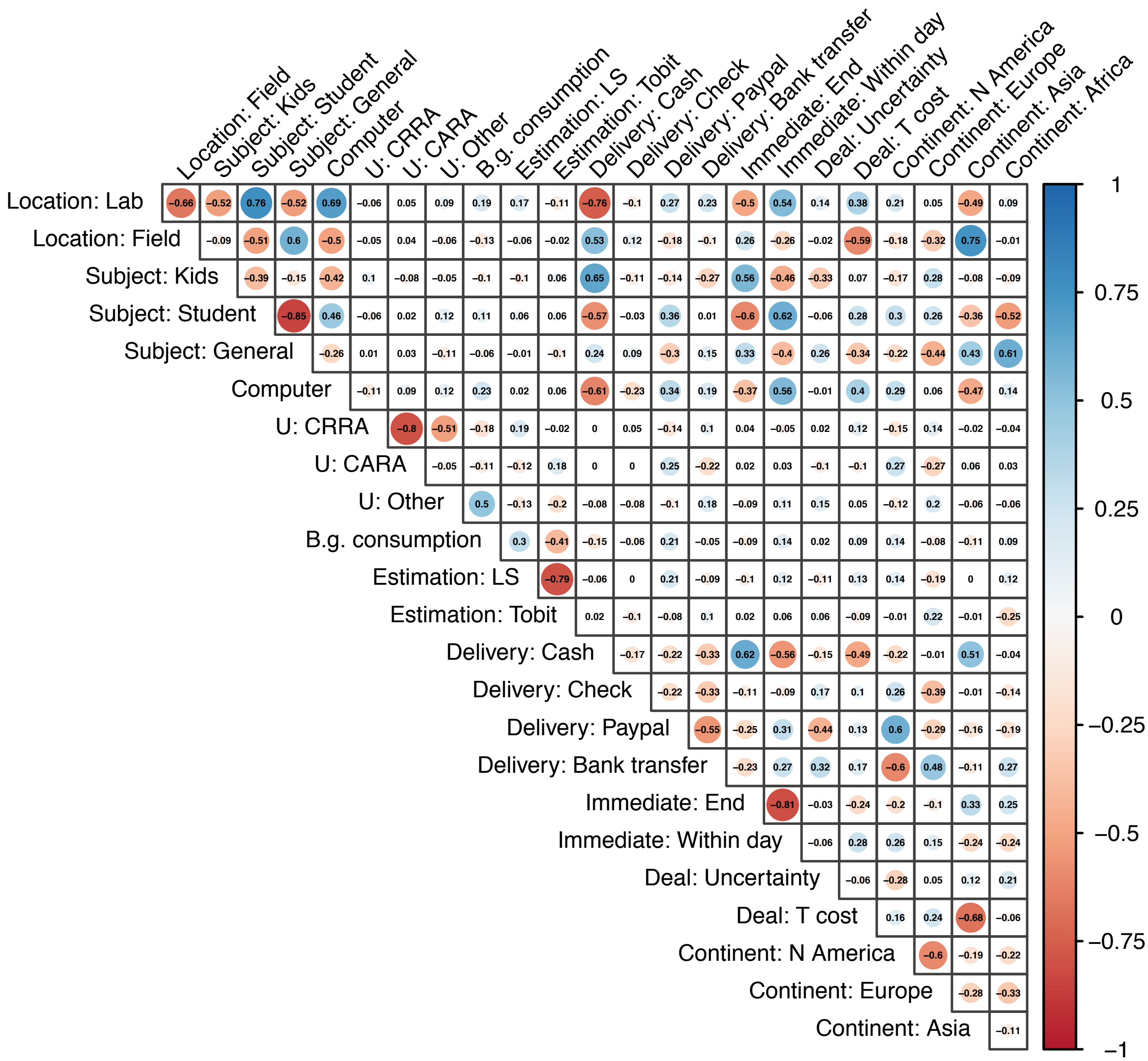


FIGURE C.6: Co-occurences of CTB design characteristics. Estimate-level data, monetary reward only.

C.3 Present Bias and Design Characteristics

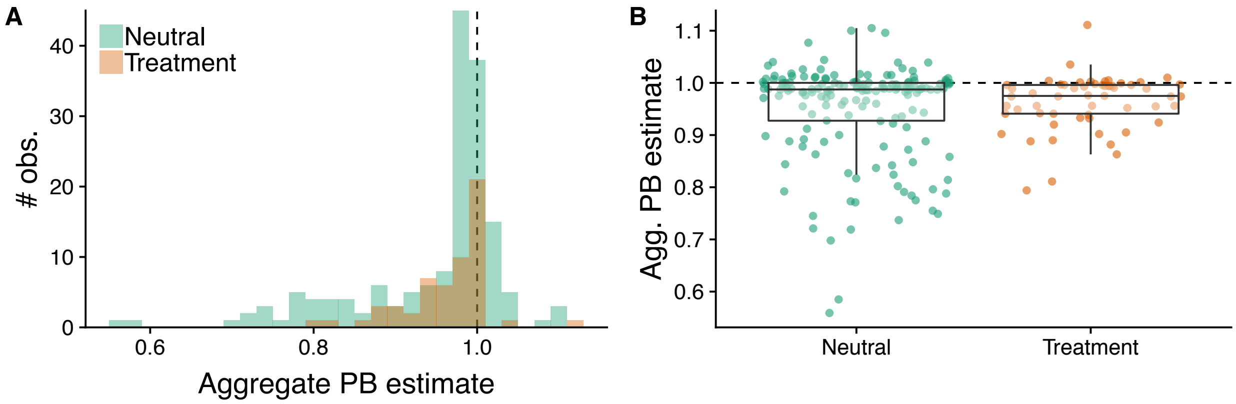


FIGURE C.7: Treatment type.

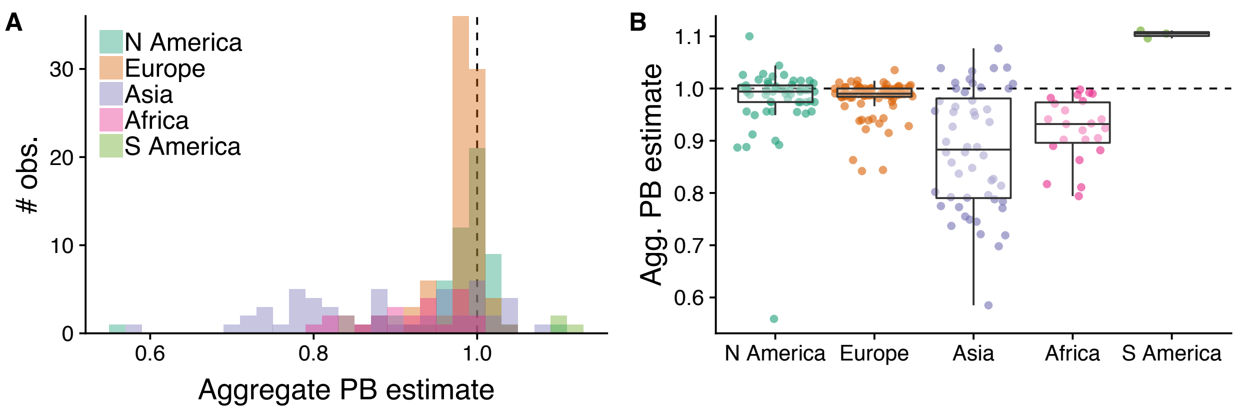


FIGURE C.8: Continent.

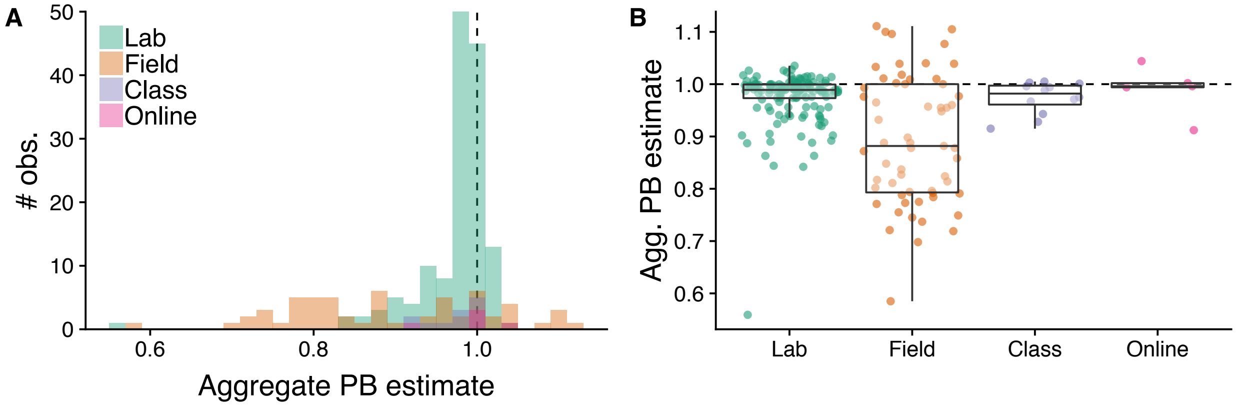


FIGURE C.9: Location of the experiment.

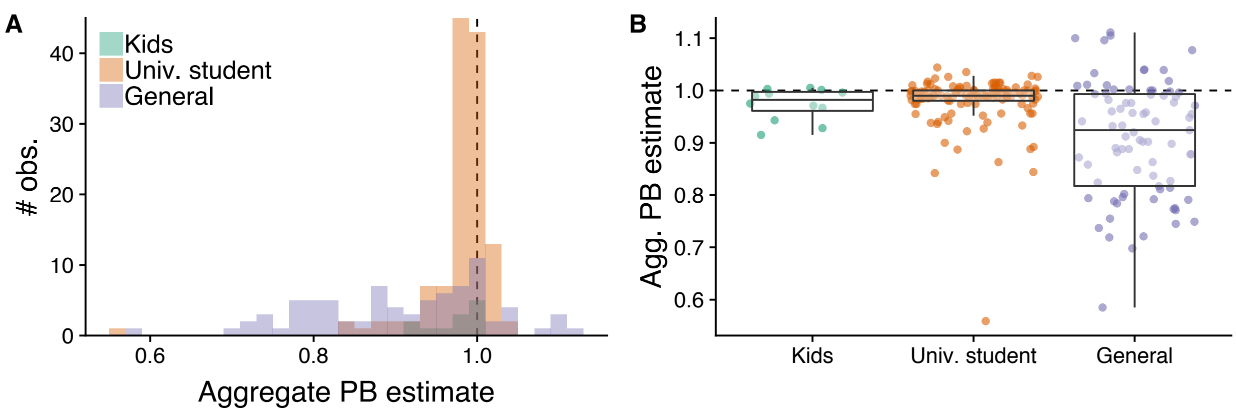


FIGURE C.10: Subject population.

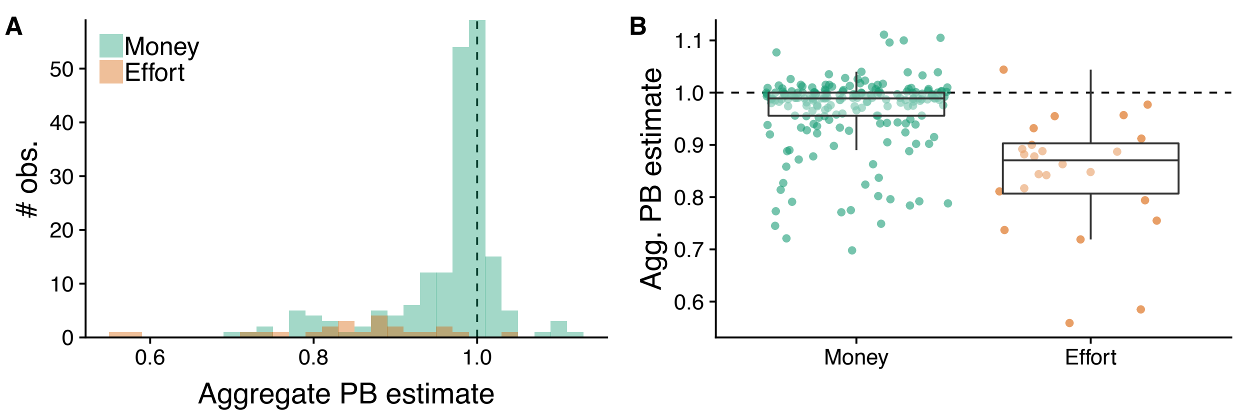


FIGURE C.11: Reward type.

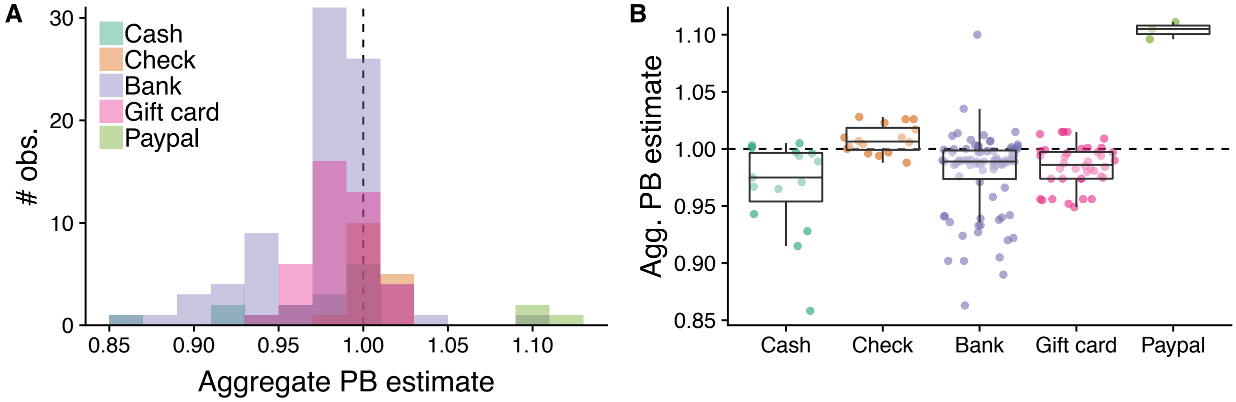


Figure C.12: Monetary reward delivery method.

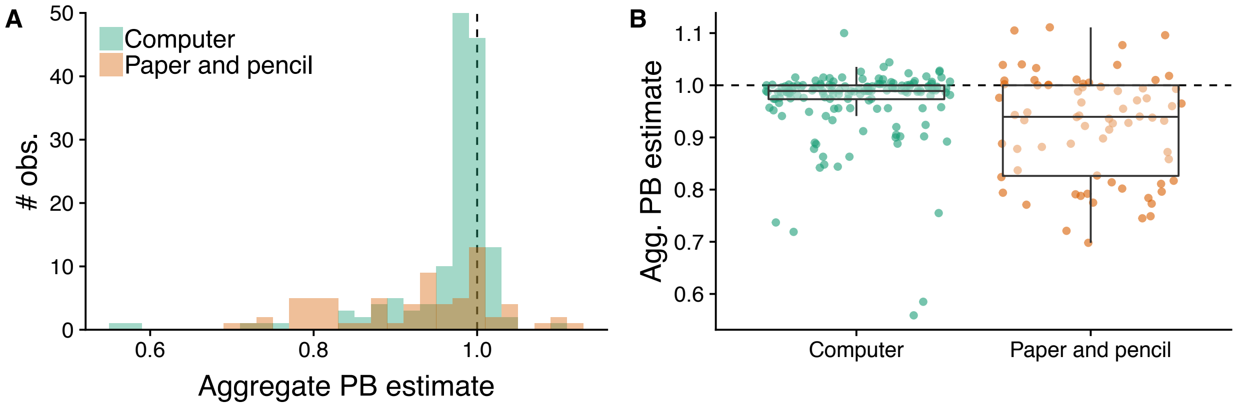


Figure C.13: Experimental interface.

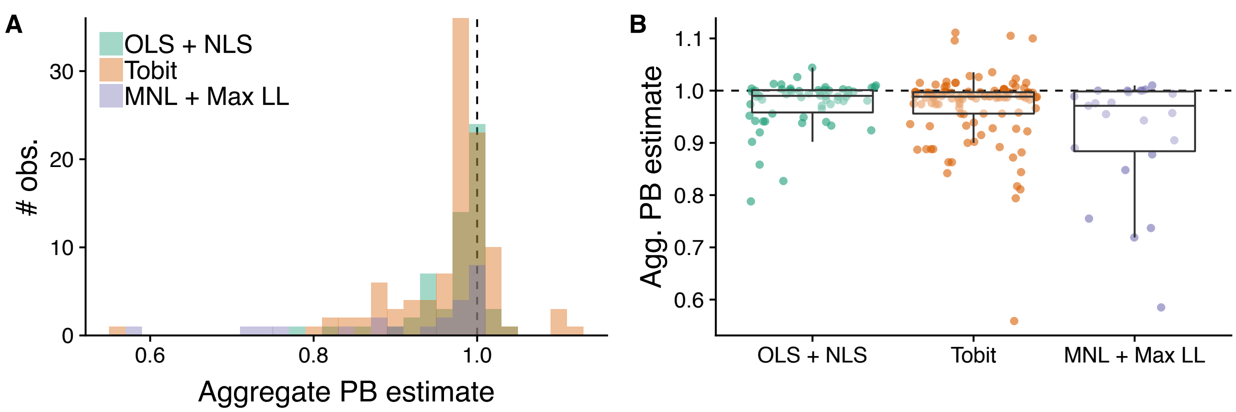


Figure C.14: Econometric approach.

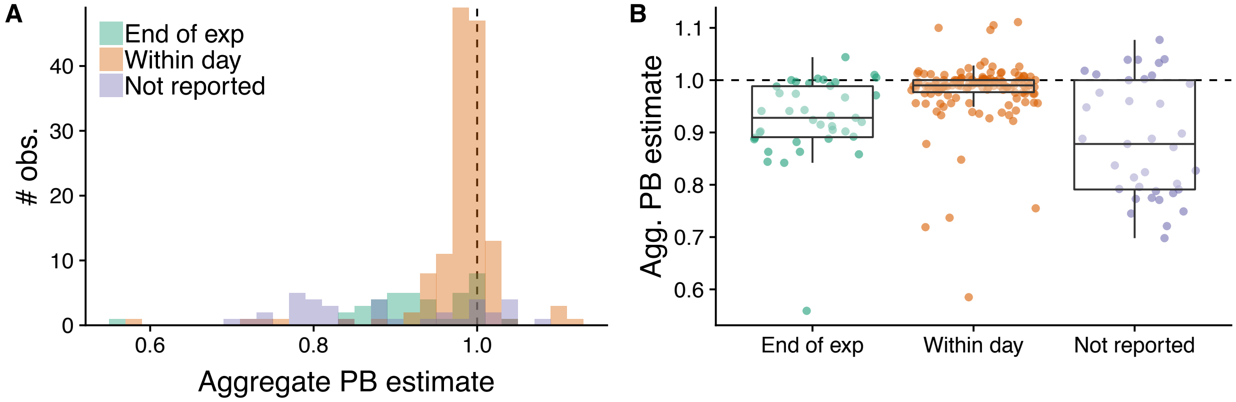


FIGURE C.15: Timing of immediate reward.

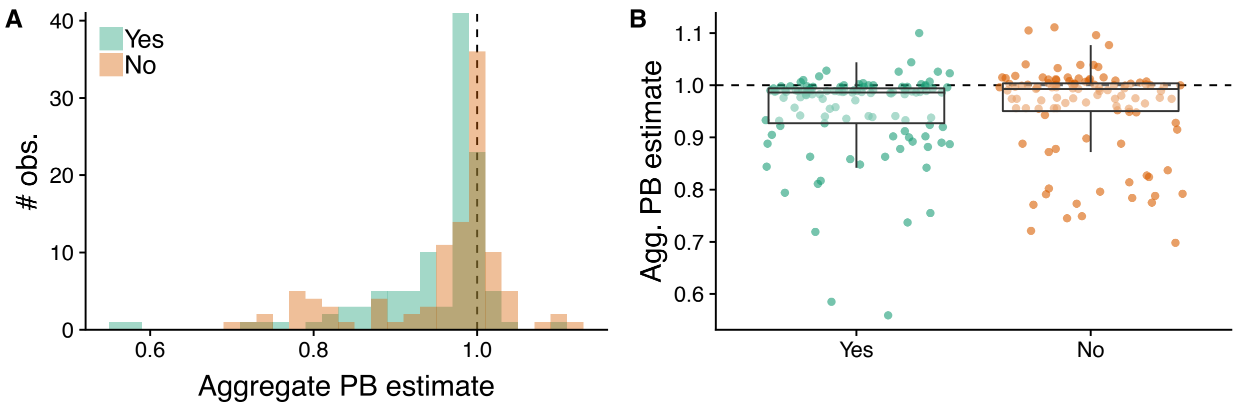


FIGURE C.16: Deal with uncertainty of future payment.

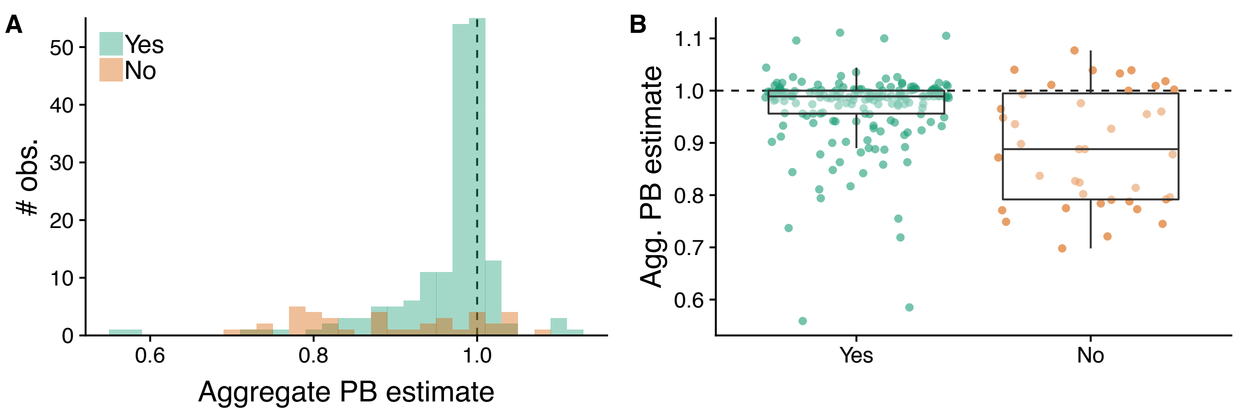


FIGURE C.17: Equalize transaction costs between two periods.

C.4 Meta-Regression Analysis

Common-effect estimate. In Section 4.1, we present meta-analytic averages $\overline{PB}_0$ estimated with the random-effects model and multi-level model. These models make sense given the sizable amount of heterogeneity between studies, but we present common-effect estimate here for completeness.

For the common-effect model, we assume that the sampling variance is known only up to some unknown multiplicative constant (i.e., $v_j^2 = \phi\tilde{v}_j^2$ for some $\phi > 0$). Equation (2) then becomes the unrestricted weighted least squares model (UWLS; Stanley and Doucouliagos, 2015). The common-effect and the unrestricted weighted least squares models give the same weighted average $\overline{PB}_0$ but their associated variances are different. The unknown constant ϕ is given by the residual variance from the standard weighted least squares.

Table C.1: Common-effect estimate of average present bias parameter (cf. Table 5).

	Without influential observations				All observations			
	(1) All	(2) Monetary (all)	(3) Monetary (neutral)	(4) Effort	(5) All	(6) Monetary (all)	(7) Monetary (neutral)	(8) Effort
$\overline{PB}_0$	0.9941	0.9943	0.9964	0.9072	0.9875	0.9876	0.9892	0.9072
	(0.0020)	(0.0020)	(0.0036)	(0.0242)	(0.0084)	(0.0084)	(0.0099)	(0.0242)
p-value	0.0069	0.0107	0.3317	0.0050	0.1444	0.1562	0.2873	0.0050
Observations	217	193	140	24	220	196	142	24
Studies	29	20	19	9	31	22	21	9

Note: p-values are from the two-sided test of the null hypothesis $H_0 : PB = 1$. Standard errors in parentheses are cluster-robust (Hedges et al., 2010).

Influential observations. Tables C.2-C.5 present the estimation results using all observations in the dataset, including three influential observations (cf. Tables 5-8).

TABLE C.2: Meta-analytic average of present bias parameter (cf. Tables 5).

	All studies		Monetary (all)		Monetary ("neutral")		Effort cost	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	RE	ML	RE	ML	RE	ML	RE	ML
$\overline{PB}_0$	0.9662	0.9532	0.9720	0.9750	0.9715	0.9754	0.8815	0.8802
	(0.0144)	(0.0139)	(0.0147)	(0.0141)	(0.0204)	(0.0148)	(0.0171)	(0.0208)
p-value	0.0261	0.0021	0.0708	0.0912	0.1786	0.1112	0.0001	0.0004
τ^2	0.0031		0.0028		0.0037		0.0021	
I^2	98.6540		98.6865		98.5134		45.9587	
I^2_{within}		0.8028		0.9985		0.4772		9.2236
I^2_{between}		98.4187		98.1379		98.2989		39.5324
Observations	220	220	196	196	142	142	24	24
Studies	31	31	22	22	21	21	9	9

Notes: p-values are from the two-sided test of the null hypothesis $H_0 : PB = 1$. Standard errors in parentheses are cluster-robust (Hedges et al., 2010). τ^2 in the random-effects model is estimated using the restricted maximum likelihood method. Three observations with large influence measure ($|DFBETAS| > 1$) are included.

Table C.3: Funnel plot asymmetry and precision effect testing (cf. Table 6).

		All studies		Monetary (all)		Monetary ("neutral")		Effort cost	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SE of PB estimate	α_1	−0.9623	−0.3649	−0.7849	−0.2447	−0.6752	−0.1872	−2.0571	−1.8720
		(0.8164)	(0.3342)	(0.9410)	(0.3422)	(1.2438)	(0.3934)	(0.4412)	(0.1093)
Constant	α_0	0.9907		0.9902		0.9920		0.9931	
		(0.0100)		(0.0103)		(0.0132)		(0.0255)	
FAT ($H_0 : \alpha_1 = 0$)	p-value	0.24781	0.2836	0.4136	0.4825	0.5932	0.6394	0.0016	0.0000
PET ($H_0 : \alpha_0 = 1$)	p-value	0.3566		0.3491		0.5527		0.7931	
Study fixed effect		No	Yes	No	Yes	No	Yes	No	Yes
Observations		220	220	196	196	142	142	24	24
Number of studies		31	31	22	22	21	21	9	9
R^2		0.0326	0.9372	0.0193	0.9384	0.0146	0.9644	0.5100	0.9503
Adjusted R^2		0.0282	0.9269	0.0142	0.9305	0.0076	0.9582	0.4877	0.9183

Notes: Estimated by weighted least squares. Standard errors are clustered at the study level. Three observations with large influence measure ($|DFBETAS| > 1$) are included. In the specification with study fixed effects, the constant term is dropped and all the dummy variables for the studies are included.

Table C.4: Explaining the heterogeneity of reported estimates (monetary reward; including overly influential estimates; cf. Table 7).

	(1)	(2)	(3)	(4)	(5)	(6)
SE of PB estimate	0.852	0.926	−1.011	−0.205	−0.878	−0.331
	(1.082)	(0.885)	(0.626)	(0.455)	(0.510)	(0.369)
Non-neutral condition	-0.009^{**}	-0.008^{**}	−0.005	−0.004	−0.004	−0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Location: Field	0.042	0.039	0.130^{***}	0.259^{***}	0.201^{***}	0.113^{**}
	(0.026)	(0.040)	(0.033)	(0.061)	(0.040)	(0.043)
Location: Class	−0.008	−0.023	0.069^{**}	0.138^{***}	0.060^{***}	−0.004
	(0.017)	(0.020)	(0.022)	(0.040)	(0.018)	(0.031)
Location: Online	−0.0002	0.027^{*}	-0.019^{*}	−0.009	−0.013	−0.010
	(0.004)	(0.013)	(0.010)	(0.007)	(0.009)	(0.009)
"Immediate" pay: Within day	0.037^{**}	0.043	0.053^{***}	0.030^{***}	0.016	0.023
	(0.013)	(0.022)	(0.015)	(0.009)	(0.012)	(0.016)
"Immediate" pay: Not reported	−0.083	−0.078	0.059	−0.066	-0.160^{*}	−0.046
	(0.064)	(0.067)	(0.068)	(0.054)	(0.079)	(0.065)
Delivery: Cash	0.015	0.047	0.029	0.013	−0.003	0.008
	(0.019)	(0.030)	(0.020)	(0.014)	(0.018)	(0.019)
Delivery: Bank	-0.039^{*}	-0.028^{*}	−0.014	0.034	0.003	-0.034^{**}
	(0.019)	(0.013)	(0.008)	(0.020)	(0.009)	(0.013)
Delivery: Other	-0.011^{**}	-0.010^{*}	-0.014^{**}	0.009	−0.001	-0.013^{*}
	(0.004)	(0.005)	(0.005)	(0.010)	(0.006)	(0.005)
Estimation: Tobit		-0.036^{**}	0.006	−0.006	−0.002	0.007
		(0.014)	(0.008)	(0.007)	(0.007)	(0.007)
Estimation: Other		0.011	0.003	−0.001	−0.014	−0.001
		(0.012)	(0.008)	(0.011)	(0.010)	(0.008)
Estimation: B.g. consumption		−0.004	−0.003	−0.003	−0.004	0.007
		(0.010)	(0.007)	(0.007)	(0.006)	(0.007)
Deal uncertainty			−0.005	0.022^{*}	0.009	0.003
			(0.004)	(0.009)	(0.006)	(0.006)
Deal transaction cost			0.123^{**}	0.096^{**}	0.006	0.075
			(0.043)	(0.036)	(0.051)	(0.054)
Paper and pencil			-0.070^{***}	-0.099^{***}	-0.044^{***}	-0.035^{*}
			(0.012)	(0.020)	(0.011)	(0.015)
Credit card				0.179^{***}		
				(0.053)		
Withdrawal					0.333^{**}	
					(0.112)	
Emergency funds impossible						-0.236^{*}
						(0.092)
Constant	0.966^{***}	0.967^{***}	0.845^{***}	0.759^{***}	0.675^{***}	0.978^{***}
	(0.014)	(0.025)	(0.059)	(0.047)	(0.078)	(0.089)
Observations	196	196	196	196	196	196
R^2	0.436	0.607	0.798	0.865	0.867	0.845
Adjusted R^2	0.405	0.579	0.780	0.852	0.854	0.831

Note: Observations with large influence measure ($|DFBETAS| > 1$) are included. Study fixed effects are not included in the model. Standard errors are clustered at the study level. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table C.5: Explaining the heterogeneity of reported estimates (monetary vs. non-monetary rewards; cf. Table 8).

	(1)	(2)	(3)	(4)	(5)	(6)
Constant (*PB* from effort-CTB)	0.907^{***}	0.907^{***}	0.907^{***}	0.993^{***}	0.993^{***}	0.993^{***}
	(0.023)	(0.023)	(0.023)	(0.024)	(0.024)	(0.024)
SE of PB estimates				-2.057^{***}	-2.057^{***}	-2.057^{***}
				(0.414)	(0.414)	(0.414)
Reward: Money	0.082^{**}	0.075^{**}	0.092^{***}	−0.0002	−0.015	0.001
	(0.025)	(0.027)	(0.023)	(0.027)	(0.031)	(0.024)
× Non-neutral condition	−0.003	-0.013^{**}	-0.011^{**}	−0.005	-0.013^{**}	-0.011^{**}
	(0.009)	(0.005)	(0.003)	(0.008)	(0.005)	(0.003)
× Location: Field		0.063^{*}	0.061		0.057^{*}	0.054
		(0.027)	(0.040)		(0.026)	(0.040)
× Location: Class		0.027	0.030		0.022	0.024
		(0.018)	(0.030)		(0.022)	(0.033)
× Location: Online		0.024	0.048^{**}		0.027	0.052^{***}
		(0.015)	(0.015)		(0.019)	(0.015)
× "Immediate": By end of exp		−0.021	−0.016		−0.024	−0.019
		(0.023)	(0.030)		(0.021)	(0.031)
× "Immediate": Not reported		−0.115	−0.122		−0.121	−0.129
		(0.063)	(0.067)		(0.064)	(0.068)
× Estimation: Tobit			-0.042^{**}			-0.043^{***}
			(0.015)			(0.013)
× Estimation: Other			−0.007			−0.007
			(0.007)			(0.008)
× *SE* of PB estimates				1.175	2.881	2.988^{**}
				(1.058)	(1.482)	(0.986)
Observations	220	220	220	220	220	220
R^2	0.019	0.262	0.519	0.047	0.280	0.540
Adjusted R^2	0.010	0.237	0.498	0.030	0.249	0.516
$H_0 : PB_{\text{effort}} = 1$		$p = 0.0004$			$p = 0.7743$	

Note: Observations with large influence measure ($|DFBETAS| > 1$) are included. Study fixed effects are not included in the model. Standard errors are clustered at the study level. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

C.5 Access to Financial Markets

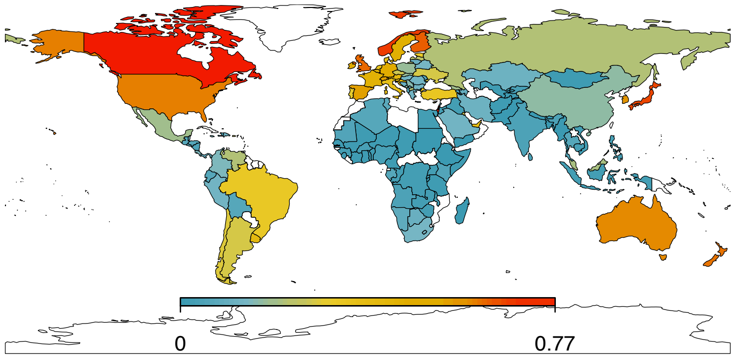


Figure C.18: Credit card ownership.

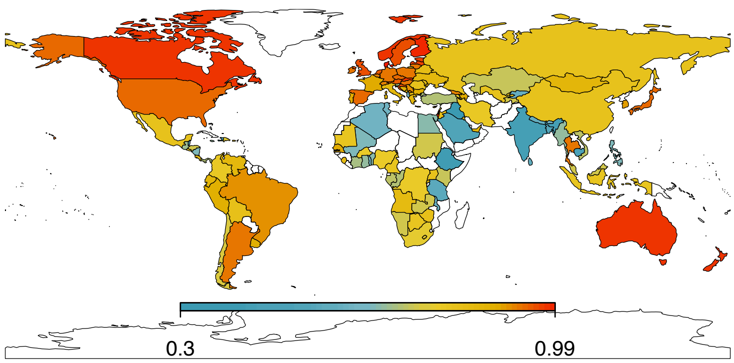


Figure C.19: Withdrawal from account in last year.

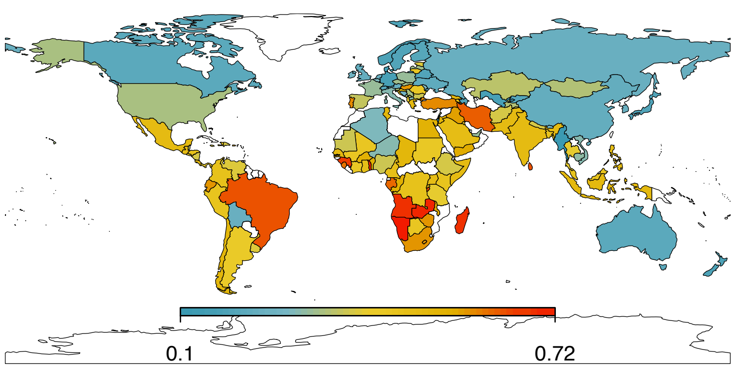


Figure C.20: Impossible to come up with emergency funds.

C.6 Bayesian Model Averaging

In Section 4.3, we estimate a meta-regression model of the form:

$$y_i = \gamma_0 + \boldsymbol{\gamma} \boldsymbol{X}_i + \varepsilon_i.$$

A problem arises when the set of potential explanatory variables $\mathfrak{X}$ is large and a researcher is not sure which variables should be included in the model.

Bayesian model averaging (BMA) approaches such model uncertainty by estimating models for all possible combination of potential explanatory variables in $\mathfrak{X}$ and constructing a weighted average (Hoeting et al., 1999; Moral-Benito, 2015; Steel, forthcoming). Suppose the size of $\mathfrak{X}$ is q. Then, there are 2^q candidate models M_m, indexed by m, to be estimated.

Let $P(M_m)$ be a prior model probability. It is typically assumed to be uniform ($P(M_m) \propto 1$) to represent the lack of knowledge. We can calculate the posterior model probability using Bayes' rule as:

$$P(M_m \mid \boldsymbol{y}) = \frac{f(\boldsymbol{y} \mid M_m) P(M_m)}{f(\boldsymbol{y})},$$

where f denotes a (conditional) likelihood of observation $\boldsymbol{y}$. Since each model M_m depends on parameters $\boldsymbol{\gamma}^m$, we can calculate the posterior for the parameters associated with M_m as:

$$g(\boldsymbol{\gamma}^m \mid \boldsymbol{y}, M_m) = \frac{f(\boldsymbol{y} \mid \boldsymbol{\gamma}^m, M_m) g(\boldsymbol{\gamma}^m \mid M_m)}{f(\boldsymbol{y} \mid M_m)}.$$

Combining these observations, we now obtain the posterior of the parameters for all the models under consideration:

$$g(\boldsymbol{\gamma} \mid \boldsymbol{y}) = \sum_{m=1}^{2^q} g(\boldsymbol{\gamma}^m \mid \boldsymbol{y}, M_m) P(M_m \mid \boldsymbol{y}).$$

Following figures represent results from BMA. In each plot, columns denote individual models where variables are sorted by posterior model probability in a descending order. Blue cells (darker cells in grayscale) indicate that the variable is included in the model and has a positive coefficient, while red cells (lighter cells in grayscale) indicate that the variable has a negative coefficient. White cells indicate that the variable is not included in the model.

Figure C.21 and Figure C.22 use observations both from monetary-CTB and effort-CTB, while Figures C.23 and C.24 (reported as Figure 5 in the main paper) focus only on monetary CTB. The top panel in each plot uses observations both from neutral and non-neutral conditions and the bottom panel discards data from non-neutral conditions. Observations with large influence measure ($|DFBETAS| > 1$) are excluded.

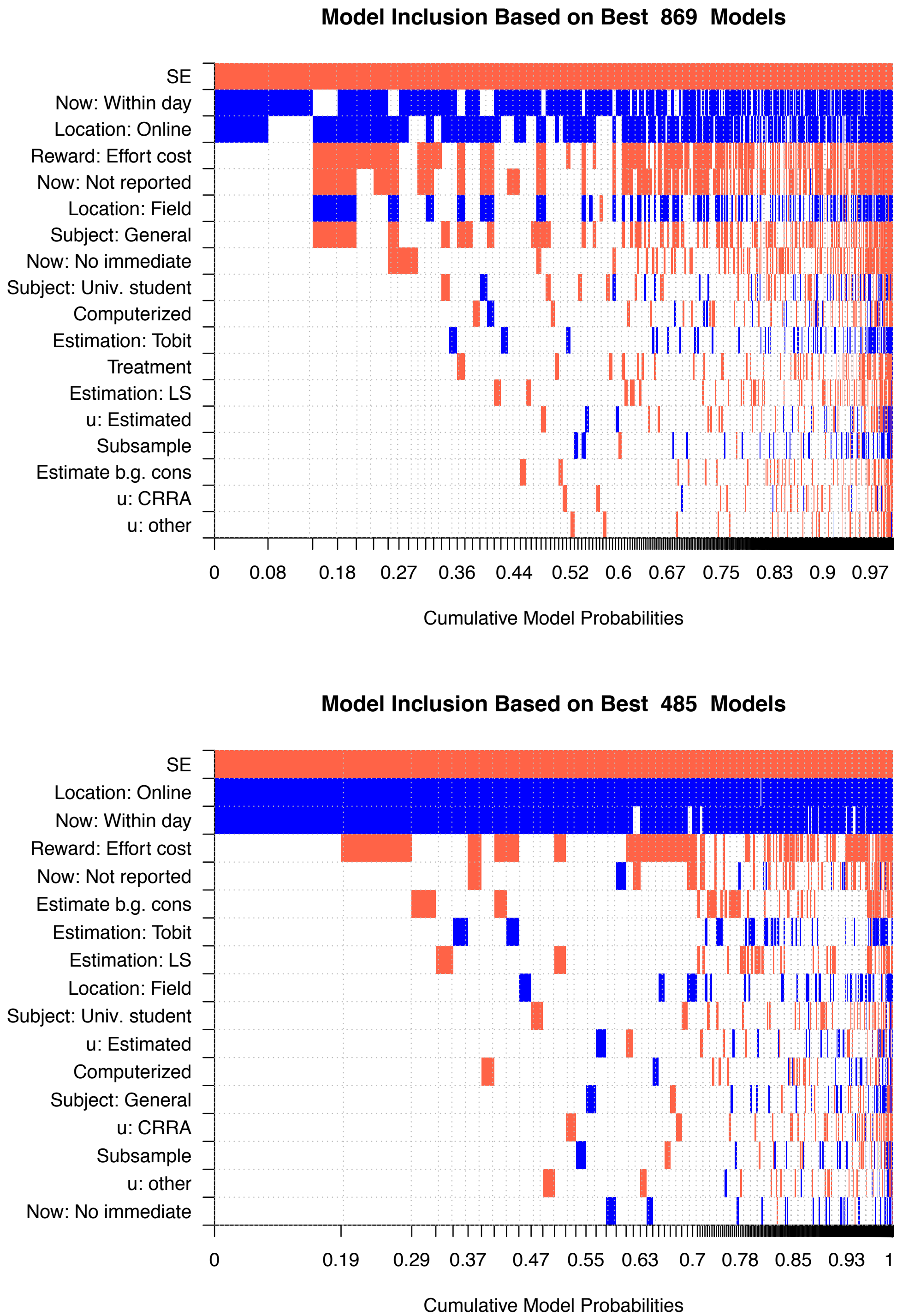
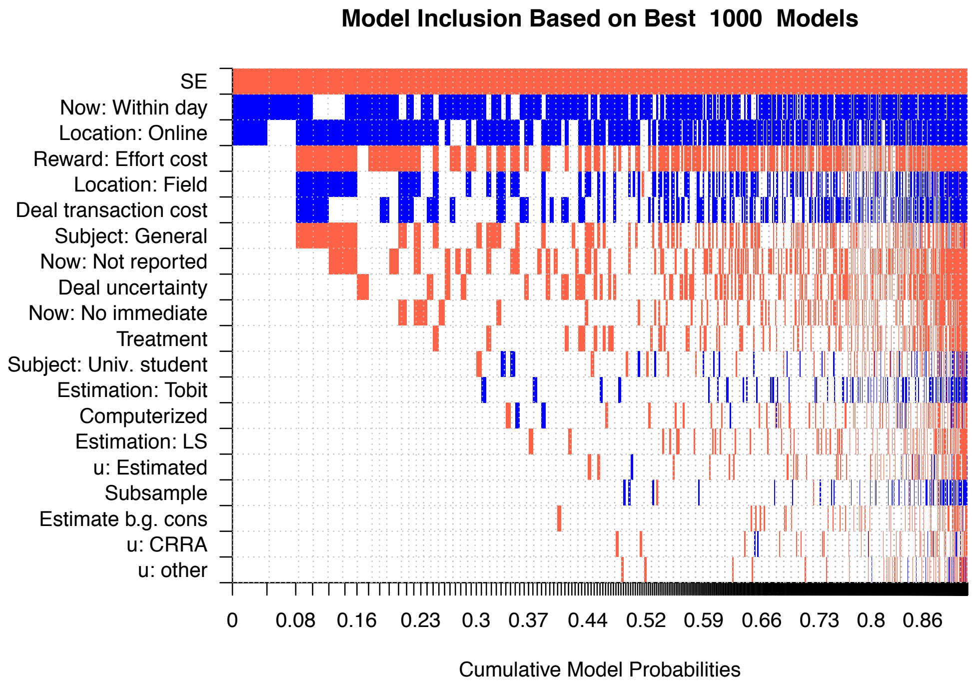


Figure C.21: Model inclusion. Observations from both monetary-CTB and effort-CTB studies. The top panel of the figure uses observations from both neutral and non-neutral conditions, while the bottom panel discards data from non-neutral conditions.



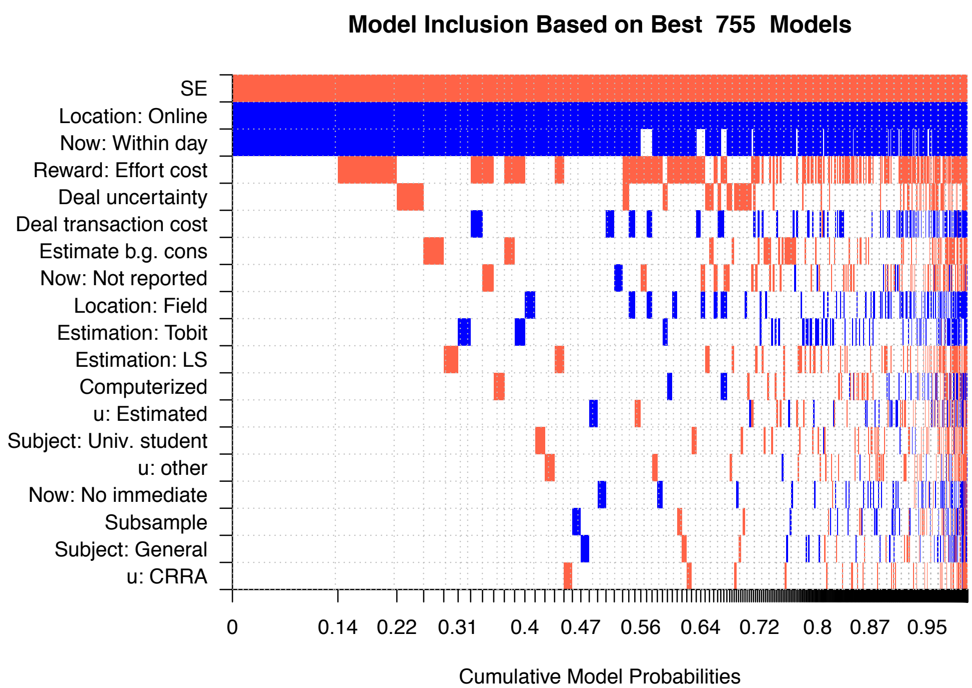
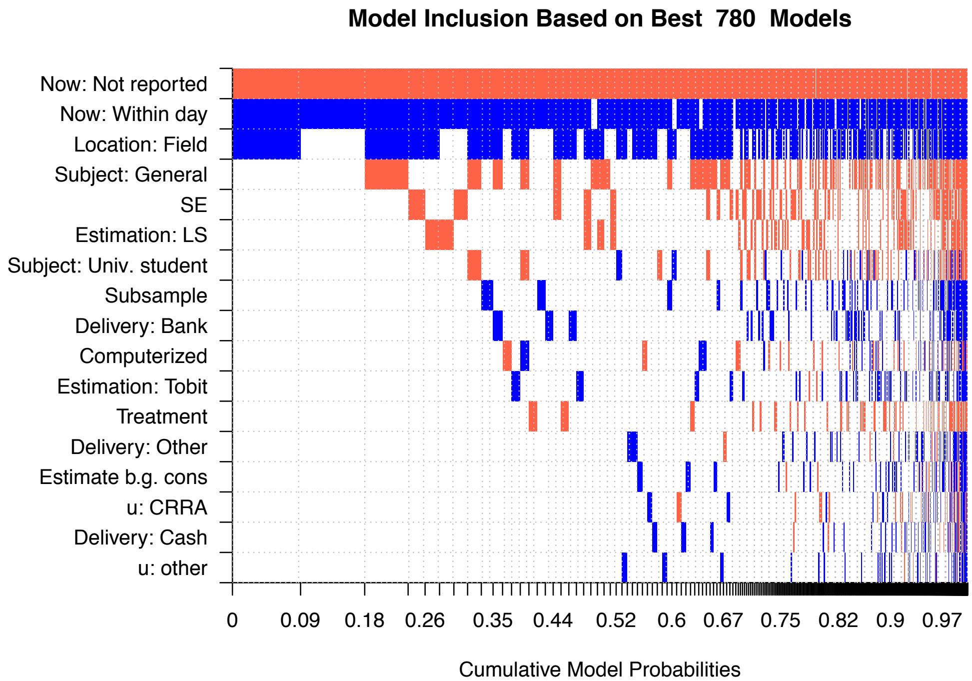


Figure C.22: Model inclusion. Observations from both monetary-CTB and effort-CTB studies. The top panel uses observations from both neutral and non-neutral conditions, while the bottom panel discards data from non-neutral conditions.



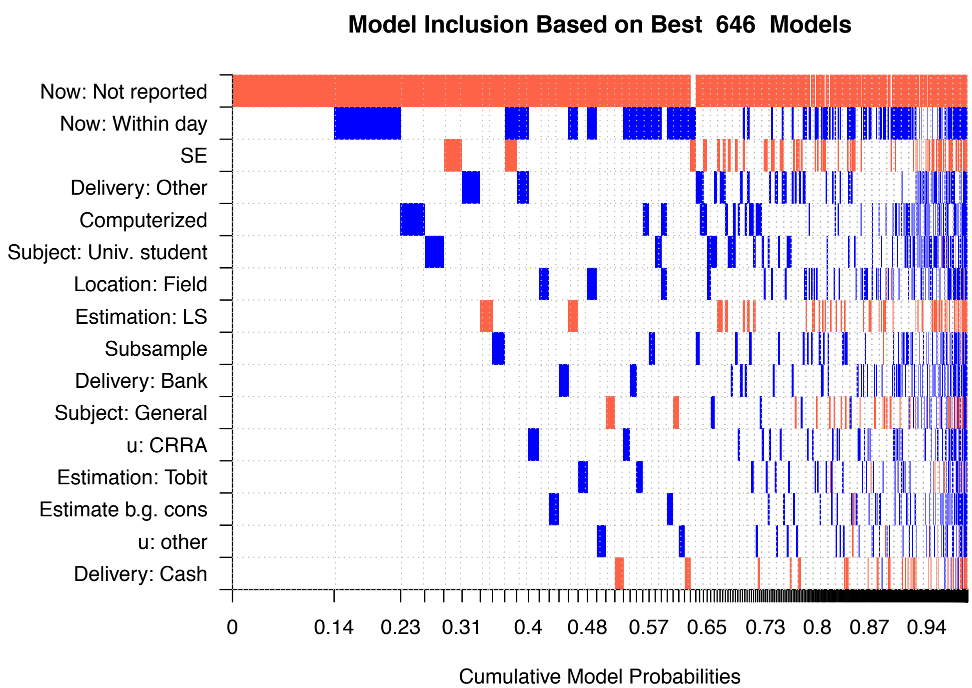
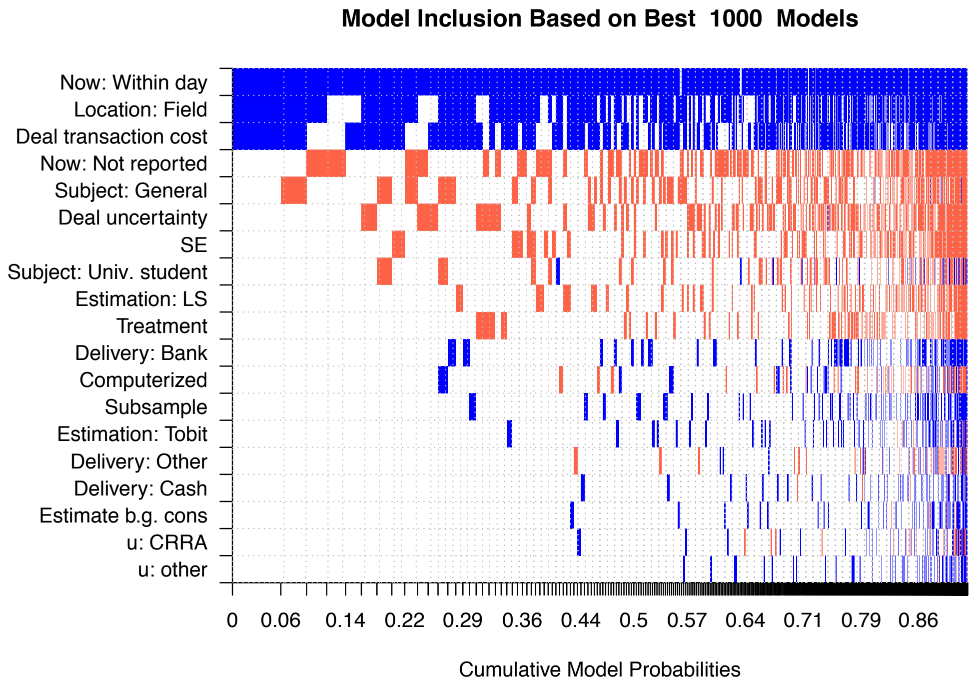


FIGURE C.23: Model inclusion. Observations from monetary-CTB studies only. The top panel uses observations from both neutral and non-neutral conditions, while the bottom panel discards data from non-neutral conditions.



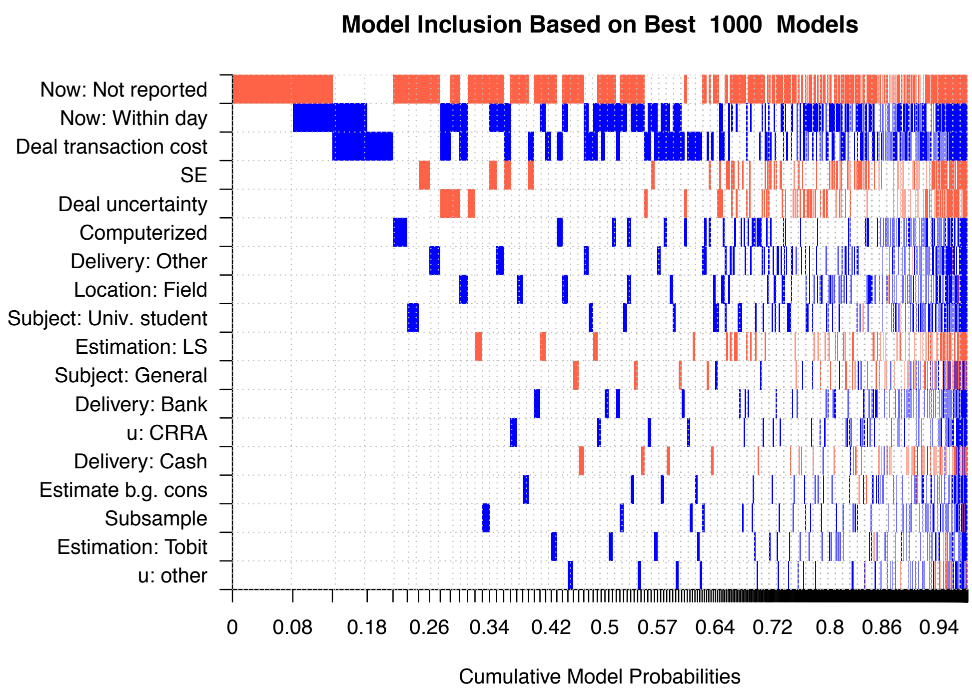


FIGURE C.24: Model inclusion. Observations from monetary-CTB studies only. The top panel uses observations from both neutral and non-neutral conditions, while the bottom panel discards data from non-neutral conditions.

C.7 Latent Studies Model

Andrews and Kasy (2019), hereafter AK, propose using the collected data from a meta-analysis to model the conditional probability of publication as a function of a study's results. The conditional publication probabilities can then be used to generate publication-bias-corrected estimates for the reported results from each study, along with associated confidence intervals.[1]

The setup for their nonparametric estimator is to assume that there exists a population of latent studies indexed by i. The *true* parameter that study i attempts to estimate is denoted Θ_i^*, and is drawn from distribution μ_Θ, such that it may vary across studies.

The result for latent study i, denoted X_i^*, is drawn from the normal distribution $N(\Theta_i^*, \Sigma_i^{*2})$, where Σ_i^* is the (fixed) standard deviation of the estimate X_i^* in latent study i. AK then assume that we only observe "published" studies, with latent studies published with probability $p(Z_i^*)$, where $Z_i^* = X_i^*/\Sigma_i^*$.

We use the *degree* of present bias $X_i^* = 1 - PB_i$, the deviation of estimated present-bias parameter from one, as the variable of interest. [2] Figure C.25A shows the density plot of the Z-statistics. The plot does exhibit jumps in the density around the cutoffs -1.96 and 1.96, unlike many applications discussed in Andrews and Kasy (2019). Figure C.25B is the funnel plot and carries the same information as Figure 1.

AK show that we can identify the conditional publication probability $p(\cdot)$ up to scale using the data collected in a meta-analysis, and then use the estimated $p(\cdot)$ to derive median unbiased estimators and valid confidence intervals for $\Omega_i = \Theta_i/\Sigma_i$ for "published" studies (random variables relating only to "published" studies are denoted by the lack of an asterisk). The intuition behind this identification result is that, in the presence of publication bias, we can glean information on the probability of a given result being published by comparing the observed distribution of results from studies with different standard deviations to see if there are areas of the distribution of estimates with fewer results than would be expected given the results from other studies and the standard deviation of estimates in this area of the distribution.

[1] Mirroring our discussion in Section 4.2, these conditional "publication" probabilities do not necessarily represent the probability of the result being published in an academic journal, but instead reflect the probability of the (latent) result being observed by the meta-analyst (including, for example, as part of unpublished work that was made available).

[2] This means that estimates of present bias less than one will yield *positive* Z-statistics.

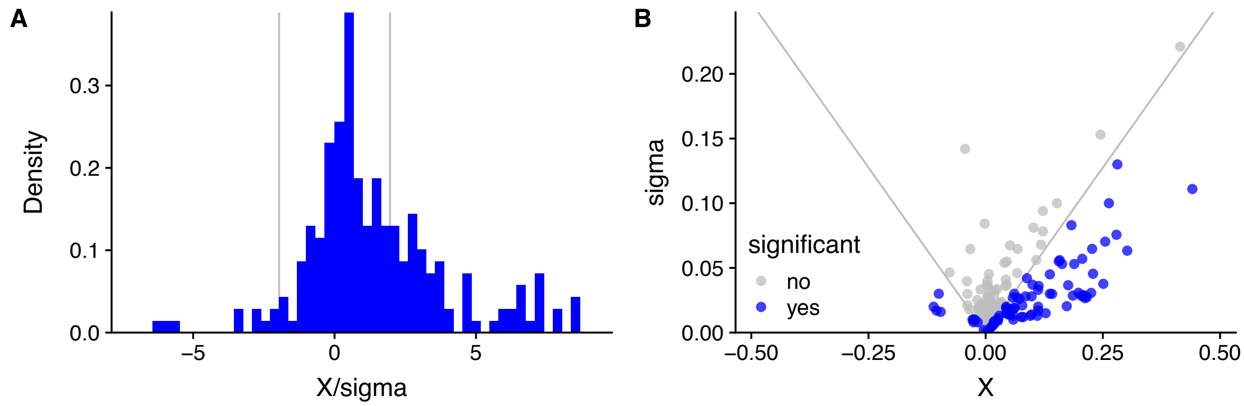


FIGURE C.25: (A) Binned density plot for the z-statistics $Z^* = X^*/\Sigma^*$. (B) Joint distribution of the estimated degree of present bias and the standard error. The grey lines mark $|X^*|/\Sigma^* = 1.96$. Overly influential ($|DFBETAS| > 1$) observations are excluded. The figure is generated with the package provided by AK.

We use the following specification for the likelihood of publication, also considered by AK:

$$\Theta^* \sim \mathcal{N}(\bar{\theta}, \tilde{\tau}^2), \quad p(Z) \propto \begin{cases} \psi_{p,1} & Z < -1.96 \\ \psi_{p,2} & -1.96 \leq Z < 0 \\ \psi_{p,3} & 0 \leq Z < 1.96 \\ 1 & Z \geq 1.96 \end{cases}.$$

The results from this specification are provided in Table C.6. They indicate the intuitive result that studies showing statistically significant *future bias* are less likely to be reported than studies showing either statistically significant present bias (reflected in $\psi_{p,1} < 1$) or studies showing no significant present *or* future bias (reflected in $\psi_{p,1} < \psi_{p,2}$ and $\psi_{p,1} < \psi_{p,3}$). [3] The estimate $\bar{\theta}$ for the mean present-biasedness in the the population of latent estimates is small and statistically indistinguishable from zero at the 5% level. When we estimate the model with a small subset of data using the real-effort version of the CTB, the mean latent effect becomes large ($\bar{\theta} = 0.096$) and is significantly different from zero.

[3]It is tempting to think that there are simply no latent studies in which the aggregate estimate of the present-bias parameter indicates future bias, but in individual results for present bias, such as those provided by Andreoni and Sprenger (2012), a surprisingly large proportion of individuals *do* exhibit choices consistent with future bias, so it is not unlikely that there are many latent studies indicating aggregate future bias.

Table C.6: Selection estimates.

			Monetary		
		All	All	"Neutral"	Effort
Mean latent effect	$\bar{\theta}$	0.026	0.013	0.061	0.096
		(0.040)	(0.051)	(0.064)	(0.016)
$\frac{\Pr[\text{Report}\mid Z<-1.96]}{\Pr[\text{Report}\mid Z\geq 1.96]}$	$\psi_{p,1}$	0.259	0.229	0.896	0.000
		(0.369)	(0.376)	(1.577)	(5.291)
$\frac{\Pr[\text{Report}\mid -1.96\leq Z<0]}{\Pr[\text{Report}\mid Z\geq 1.96]}$	$\psi_{p,2}$	1.809	2.112	6.116	0.191
		(1.432)	(1.847)	(5.483)	(0.201)
$\frac{\Pr[\text{Report}\mid 0\leq Z<1.96]}{\Pr[\text{Report}\mid Z\geq 1.96]}$	$\psi_{p,3}$	3.869	4.539	10.769	0.534
		(2.243)	(2.797)	(7.071)	(0.460)
Mean *PB*	$1-\bar{\theta}$	0.974	0.987	0.939	0.904
Test of selective reporting	$H_0 : \psi_{p,1} = \psi_{p,2} = \psi_{p,3} = 1$	0.019	0.005	0.392	0.000
Test of a true effect	$H_0 : \theta = 0$	0.511	0.804	0.342	0.000
Observations		217	193	140	24
Number of studies		29	20	19	9

Note: Three observations with large influence measure ($|DFBETAS| > 1$) are excluded. Z-values are defined such that estimates of the present bias parameter below one yield *positive* Z-values. Publication likelihood ψ_p's are measured relative to omitted category of positively significant (at 5% level) estimates. Standard errors in parentheses are clustered at study level.

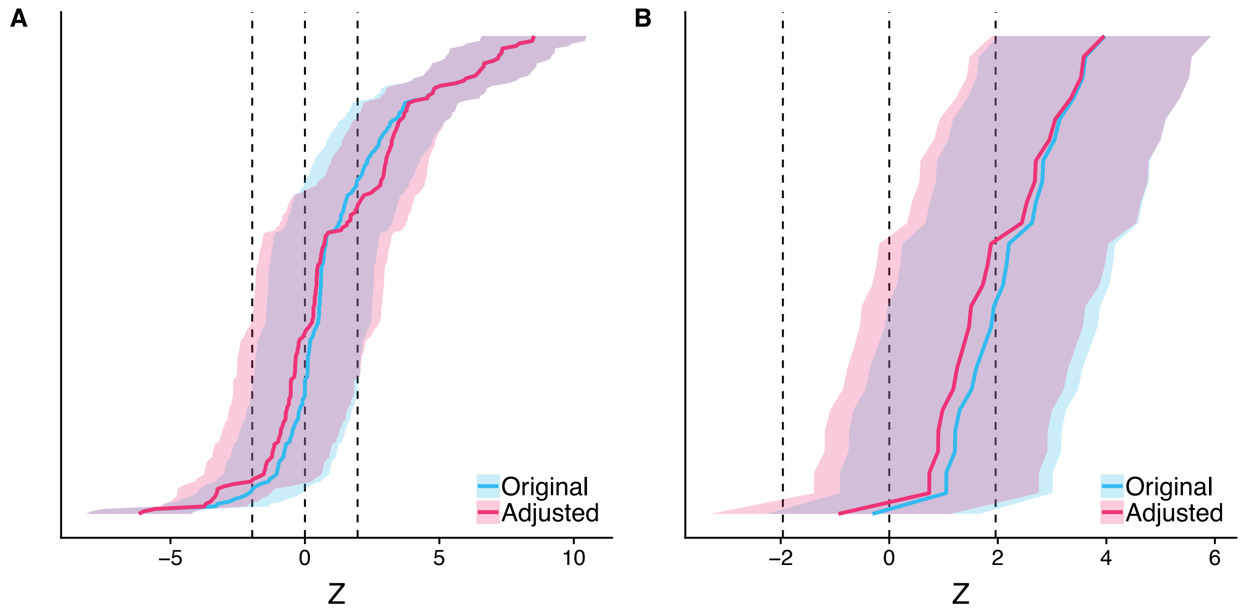


Figure C.26: The original Z-statistics and bias-corrected Z-statistics. (A) Monetary-CTB, all observations. (B) Effort-CTB. *Note*: Since we look at the deviation of estimated present-bias parameter from one, $Z > 0$ corresponds to $PB < 1$.

C.8 Stem-Based Bias Correction

Furukawa (2019) shows that a range of underlying processes—not just the biased preferences of researchers and journal editors—could lead to publication bias, and proposes a "stem-based" bias correction method for meta-analyses based on weaker assumptions regarding the selection process for reported results.

This estimator uses the studies with the highest precision to estimate a bias-corrected average effect for the hypothetical population of latent studies, since the studies with high precision are generally the least affected by publication bias (since there is simply less variation in study results for selection to occur on). The number of studies to include in the estimate is determined by minimizing the estimated mean squared error of the resulting estimator. In this way, this estimator is a generalization of the method suggested by Stanley et al. (2010) whereby the most precise 10% of all studies are averaged.

TABLE C.7: Stem-based correction.

		Monetary		
	All	All	"Neutral"	Effort
	(A)	(B)	(C)	(D)
PB	0.9910	0.9910	0.9992	0.9266
	(0.0029)	(0.0029)	(0.0036)	(0.0253)
Observations	217	193	140	24
Number of stems	56	56	55	7
% information used	0.4312	0.4497	0.5405	0.4664

Note: Three observations with large influence measure ($|DFBETAS| > 1$) are excluded. Column identifiers A-D indicate the panels in Figure C.27.

The results show that, averaging over the most precise studies, the estimated present-bias parameter is statistically different from one, indicating aggregate evidence of present bias (PB = 0.991; Table C.7 column 1, Figure C.27A). When restricting the sample to estimates without any treatment variations, the estimated present bias parameter is indistinguishable from one (PB = 0.999; Figure C.27B). These results are consistent with the simple meta-analytic average presented in Table 5, columns (1) and (4).

Similar to the other meta-analytic methods we employ, Figure C.27E show that when only studies where subjects make decisions over allocations of effort are included, there is significant aggregate evidence of present bias ($PB = 0.927$), which is in stark contrast with monetary-reward CTB ($PB = 0.999$).

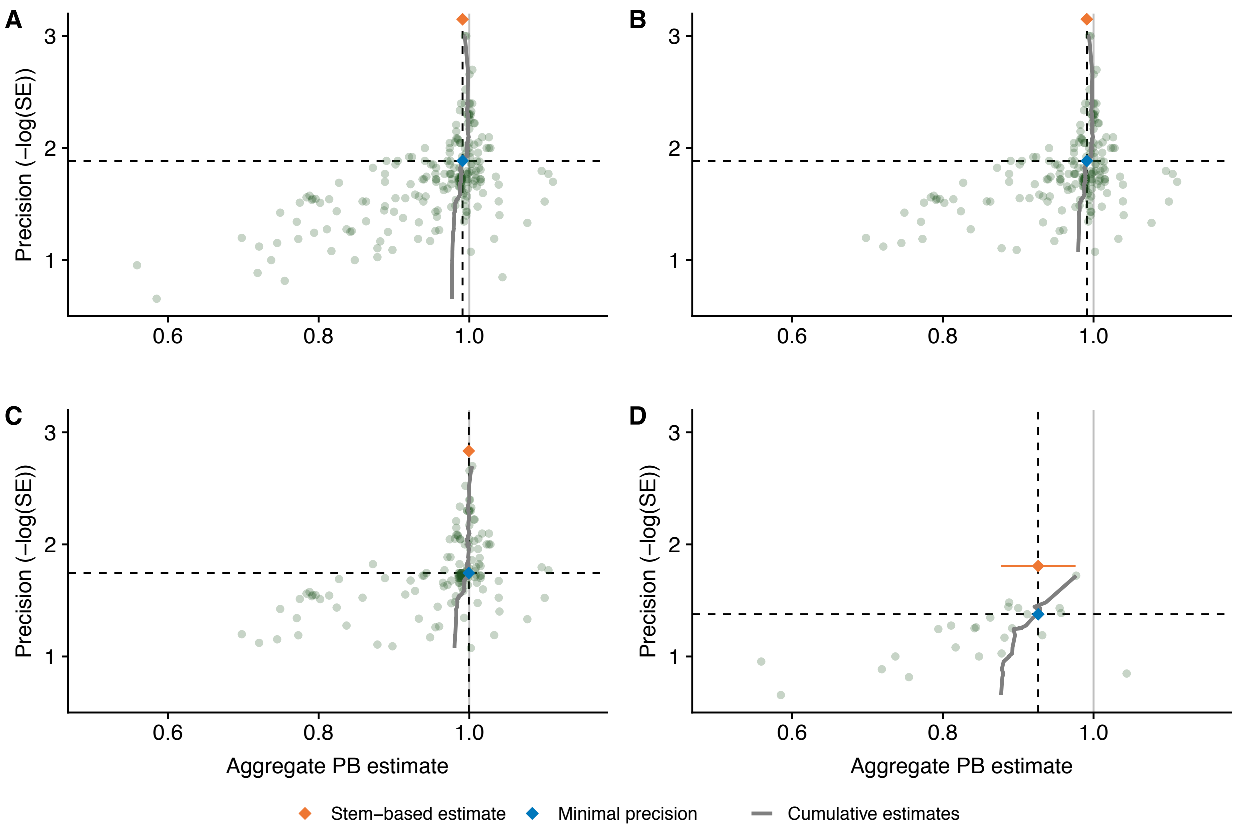


FIGURE C.27: Stem-based estimates. Overly influential ($|DFBETAS| > 1$) observations are excluded. (A) All observations. (B) Monetary-CTB, all observations. (C) Monetary-CTB, neutral condition only. (D) Effort-CTB.

C.9 P-Values of *PB* Estimates

We calculated p-values from the reported estimates and their associated standard errors since not all articles reported the p-value from the test against the null hypothesis of "no present bias" ($H_0 : PB = 1$). The distribution of p-values are shown as a boxplot for each study and in empirical CDFs split by the condition of the experiment (neutral or some treatment variation) in Figure C.28.

Just under 40% ($84/220 = 0.38$; 73 of them in the direction of present bias) of all the *PB* estimates are significantly different from one (Table C.8 in Online Appendix). The proportion of estimates with $p < 0.05$ is higher in experiments with some treatment variation than in neutral experiments, but the difference in proportions is not large (50% in treatment and 34% in neutral; two-sample z-test for proportion, $p = 0.031$). Note, however, that our classifications of "treatment" and "neutral" are made somewhat arbitrarily in some cases. [4] There are 16 studies that reported at most three *PB* estimates (eight of them reported only one estimate) and 75% (12/16) of them reported only significant estimate(s). Eight studies (out of 31) reported only insignificant result(s).

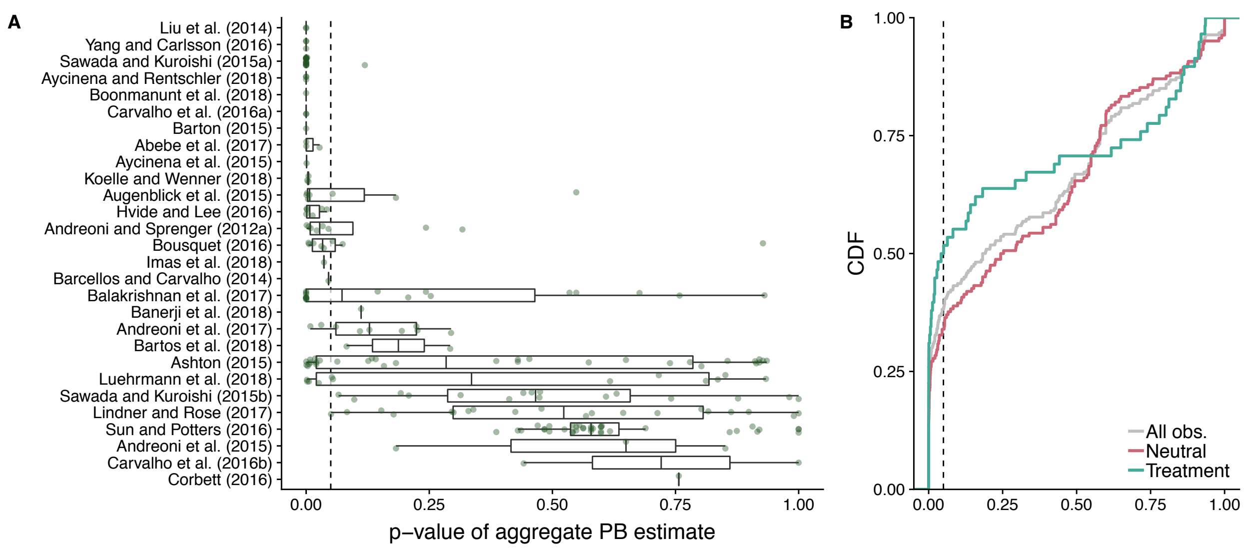


Figure C.28: P-values of present bias parameter estimates. The vertical dotted lines indicate the 5% significance level.

[4]Focusing on 84 significant ($p < 0.05$) estimates, we can make a p-curve introduced by Simonsohn et al. (2014) to detect p-hacking (which will produce disproportionately many estimates just below the desired threshold such as $p < 0.05$. The shape of the p-curve does not indicate evidence of aggressive p-hacking (Figures C.16 and C.17 in Online Appendix).

Table C.8: Re-calculaed p-values of *PB* estimates.

	All		Neutral		Treatment	
	Freq.	Prop. (%)	Freq.	Prop. (%)	Freq.	Prop. (%)
Total # estimates	220	100.0	162	100.0	58	100.0
$PB < 1$	170	77.3	121	74.7	49	84.5
with $p < 0.05$	73	42.9	45	37.2	28	57.1
$PB \geq 1$	50	22.7	41	25.3	9	15.5
with $p < 0.05$	11	22.0	10	24.4	1	11.1

Note: Proportions of statistically significant *PB* estimates ($p < 0.05$) are conditional on either $PB < 1$ or $PB \geq 1$ depending on the row.

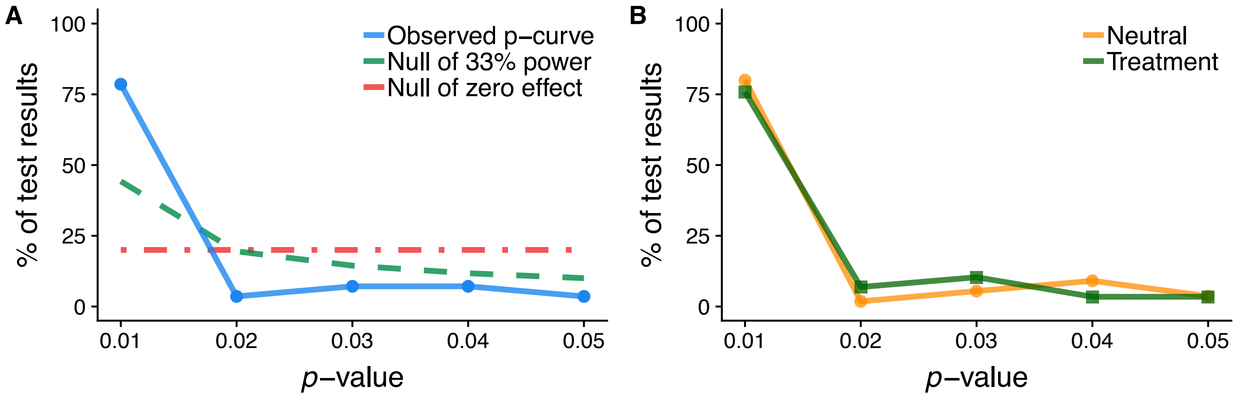


FIGURE C.29: *P*-curves (significant estimates split by the treatment type). (A) All observations. (B) Treatment type.

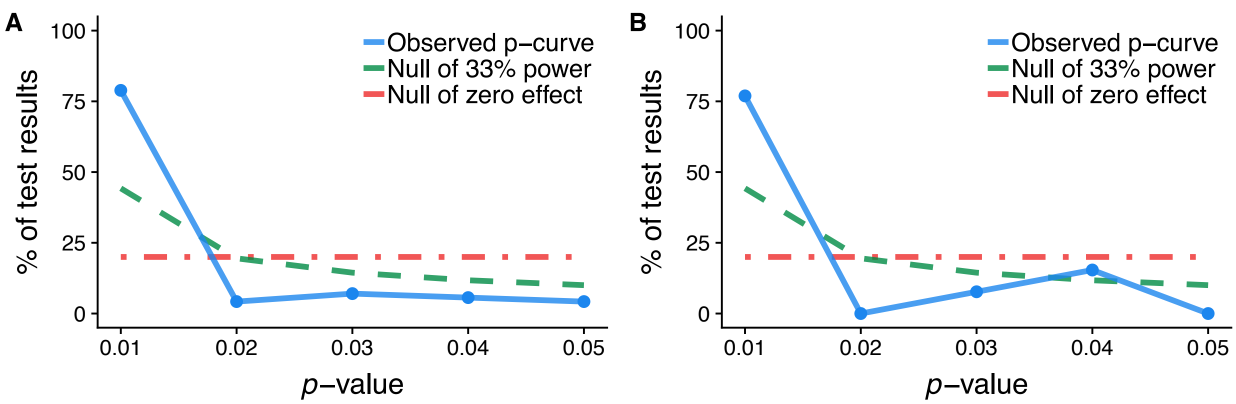


FIGURE C.30: *P*-curves (significant estimates split by the reward type). (A) Monetary-CTB. (B) Effort-CTB.

D List of Articles Included in the Master Data

Abebe, G., S. Caria, and E. Ortiz-Ospina (2017): "The Selection of Talent: Experimental and Structural Evidence from Ethiopia," Unpublished manuscript.

Alan, S. and S. Ertac (2015): "Patience, Self-Control and the Demand for Commitment: Evidence from a Large-Scale Field Experiment," *Journal of Economic Behavior & Organization*, 115, 111–122.

——— (2017): "Belief in Hard Work and Prosocial Behavior: Evidence from a Randomized Field Experiment," Unpublished manuscript.

——— (2018): "Fostering Patience in the Classroom: Results from Randomized Educational Intervention," *Journal of Political Economy*, 126, 1865–1911.

Andreoni, J., M. Callen, Y. Khan, K. Jaffar, and C. Sprenger (2017): "Using Preference Estimates to Customize Incentives: An Application to Polio Vaccination Drives in Pakistan," NBER Working Paper No. 22019.

Andreoni, J., C. Gravert, M. A. Kuhn, S. Saccardo, and Y. Yang (2018): "Arbitrage or Narrow Bracketing? Experimental Tests of Money as a Primary Reward," NBER Working Paper No. 25232.

Andreoni, J., M. A. Kuhn, and C. Sprenger (2015): "Measuring Time Preferences: A Comparison of Experimental Methods," *Journal of Economic Behavior & Organization*, 116, 451–464.

Andreoni, J. and C. Sprenger (2012a): "Estimating Time Preferences from Convex Budgets," *American Economic Review*, 102, 3333–3356.

——— (2012b): "Risk Preferences are not Time Preferences," *American Economic Review*, 102, 3357–3376.

Angerer, S., P. Lergetporer, D. Glätzle-Rützler, and M. Sutter (2015): "How to Measure Time Preferences in Children: a Comparison of Two Methods," *Journal of the Economic Science Association*, 1, 158–169.

Ashton, L. (2015): "Hunger Games: Does Hunger Affect Time Preferences?" Unpublished manuscript.

Atalay, K., F. Bakhtiar, S. Cheung, and R. Slonim (2014): "Savings and Prize-Linked Savings Accounts," *Journal of Economic Behavior & Organization*, 107, 86–106.

Augenblick, N., M. Niederle, and C. Sprenger (2015): "Working over Time: Dynamic Inconsistency in Real Effort Tasks," *Quarterly Journal of Economics*, 130, 1067–1115.

Aycinena, D., S. Blazsek, L. Rentschler, and B. Sandoval (2015): "Risk, Discounting and Demand for Intra-Household Control for Recipients of Conditional Cash Transfers," Unpublished manuscript.

Aycinena, D. and L. Rentschler (2018): "Discounting and Digit Ratio: Low 2D:4D Predicts Patience for a Sample of Females," *Frontiers in Behavioral Neuroscience*, 11, 1–13.

Balakrishnan, U., J. Haushofer, and P. Jakiela (2017): "How Soon is Now? Evidence of Present Bias from Convex Time Budget Experiments," NBER Working Paper No. 23558.

Banerji, A., J. Goto, H. Ishizaki, T. Kurosaki, K. Lai, S. Paul, Y. Sawada, and S. Tsuda (2018): "Entrepreneurship in Micro and Small Enterprises: Empirical Findings from Resurveys in Northeastern Areas of Delhi, India," Unpublished manuscript.

Barcellos, S. H. and L. Carvalho (2014): "Information about Self-Control and Intertemporal Choices," Unpublished manuscript.

Barton, B. (2015): "Interpersonal Time Inconsistency and Commitment," Unpublished manuscript.

Bartoš, V., M. Bauer, J. Chytilová, and I. Levely (2018): "Effects of Poverty on Impatience: Preferences or Inattention?" Unpublished manuscript.

Batista, C., D. Silverman, and D. Yang (2015): "Directed Giving: Evidence from an Inter-Household Transfer Experiment," *Journal of Economic Behavior & Organization*, 118, 2–21.

Blumenstock, J., M. Callen, and T. Ghani (2018): "Why Do Defaults Affect Behavior? Experimental Evidence from Afghanistan," *American Economic Review*, 108, 2868–2901.

Boonmanunt, S., T. Lauer, B. Rockenbach, and A. Weiss (2018): "Field Evidence on the Contested Role of Time Preferences in Resource Management," Unpublished manuscript.

Bousquet, L. (2016): "Measuring Time Preference and Anticipation: A Lab Experiment," Unpublished manuscript.

Bover, O., L. Hospido, and E. Villanueva (2018): "The Impact of High School Financial Education on Financial Knowledge and Choices: Evidence from a Randomized Trial in Spain," Unpublished manuscript.

Brocas, I., J. D. Carrillo, and J. Tarrasó (2018): "How Long Is a Minute?" *Games and Economic Behavior*, 111, 305–322.

Bulte, E., R. Lensink, and N. Vu (2016): "Gender Training and Female Empowerment: Experimental Evidence from Vietnam," *Economics Letters*, 145, 117–119.

Carvalho, L. S., S. Meier, and S. W. Wang (2016a): "Poverty and Economic Decision-Making:

Evidence from Changes in Financial Resources at Payday," *American Economic Review*, 106, 260–284.

Carvalho, L. S., S. Prina, and J. Sydnor (2016b): "The Effect of Saving on Risk Attitudes and Intertemporal Choices," *Journal of Development Economics*, 120, 41–52.

Cerrone, C. and L. K. Lades (2017): "Sophisticated and Naïve Procrastination: An Experimental Study," Unpublished manuscript.

Chen, Y., M. Jiang, and E. L. Krupka (2019): "Hunger and the Gender Gap," *Experimental Economics*, 22, 885–917.

Cheung, S. L. (2015): "Comment on 'Risk preferences are not Time Preferences': On the Elicitation of Time Preference under Conditions of Risk," *American Economic Review*, 105, 2242–2260.

Clot, S. and C. Y. Stanton (2014): "Present Bias Predicts Participation in Payments for Environmental Services: Evidence from a Behavioral Experiment in Uganda," *Ecological Economics*, 108, 162–170.

Clot, S., C. Y. Stanton, and M. Willinger (2017): "Are Impatient Farmers more Risk-Averse? Evidence from a Lab-in-the-Field Experiment in Rural Uganda," *Applied Economics*, 49, 156–169.

Corbett, C. H. (2016): "Preferences for Effort and Their Applications," Ph.D. thesis, University of Oregon.

de Oliveira, A. C. M. and S. Jacobson (2017): "(Im)Patience by Proxy: Making Intertemporal Decisions for Others," Unpublished manuscript.

de Quidt, J., J. Haushofer, and C. Roth (2018): "Measuring and Bounding Experimenter Demand," *American Economic Review*, 108, 3266–3302.

Dertwinkel-Kalt, M., H. Gerhardt, G. Riener, F. Schwerter, and L. Strang (2017): "Concentration Bias in Intertemporal Choice," Unpublished manuscript.

Ersoy, F. (2017): "Effects of Perceived Productivity on Study Effort: Evidence from a Field Experiment," Unpublished manuscript.

Esopo, K., D. Mellow, C. Thomas, H. Uckat, J. Abraham, P. Jain, C. Jang, N. Otis, M. Riis-Vestergaarda, A. Starcev, K. Orkin, and J. Haushofer (2018): "Measuring Self-Efficacy, Executive Function, and Temporal Discounting in Kenya," *Behaviour Research and Therapy*, 101, 30–45.

Franco, C. and M. Mahadevan (2017): "Behavioral Dynamics in Transitions from College to the Workforce," Unpublished manuscript.

Giné, X., J. Goldberg, D. Silverman, and D. Yang (2018): "Revising commitments: Field evidence on the adjustment of prior choices," *The Economic Journal*, 128, 159–188.

Grijalva, T. C., J. L. Lusk, R. Rong, and W. D. Shaw (2018): "Convex Time Budgets and Individual Discount Rates in the Long Run," *Environmental and Resource Economics*, 71, 259–277.

Hoel, J. B., B. Schwab, and J. Hoddinott (2016): "Self-Control Exertion and the Expression of Time Preference: Experimental Results from Ethiopia," *Journal of Economic Psychology*, 52, 136–146.

Hvide, H. K. and J. H. Lee (2016): "Does Source of Income Affect Risk and Intertemporal Choices?" Unpublished manuscript.

Imai, T. and C. F. Camerer (2018): "Estimating Time Preferences from Budget Set Choices Using Optimal Adaptive Design," Unpublished manuscript.

Imas, A., M. Kuhn, and V. Mironova (2018): "Waiting to Choose," Unpublished manuscript.

Janssens, W., B. Kramer, and L. Swart (2017): "Be Patient when Measuring Hyperbolic Discounting: Stationarity, Time Consistency and Time Invariance in a Field Experiment," *Journal of Development Economics*, 126, 77–90.

Kölle, F. and L. Wenner (2018): "Time-Inconsistent Generosity: Present Bias across Individual and Social Contexts," Unpublished manuscript.

Kuhn, M. A., P. Kuhn, and M. C. Villeval (2017): "Decision-Environment Effects on Intertemporal Financial Choices: How Relevant are Resource-Depletion Models?" *Journal of Economic Behavior & Organization*, 137, 72–89.

Lindner, F. and J. Rose (2017): "No Need for more Time: Intertemporal Allocation Decisions under Time Pressure," *Journal of Economic Psychology*, 60, 53–70.

Liu, E. M., J. Meng, and J. T.-y. Wang (2014): "Confucianism and Preferences: Evidence from Lab Experiments in Taiwan and China," *Journal of Economic Behavior & Organization*, 104, 106–122.

Lührmann, M., M. Serra-Garcia, and J. Winter (2018): "The Impact of Financial Education on AdolescentsâĂŹ Intertemporal Choices," *American Economic Journal: Economic Policy*, 10, 309–332.

Mayer, S. E., A. Kalil, P. Oreopoulos, and S. Gallegos (2015): "Using Behavioral Insights to Increase Parental Engagement: The Parents and Children Together (PACT) Intervention," Unpublished manuscript.

Miao, B. and S. Zhong (2015): "Comment on 'Risk Preferences are not Time Preferences': Separating Risk and Time Preference," *American Economic Review*, 105, 2272–2286.

Mudzingiri, C. (2017): "Incentivized Time Preferences, Level of Education in a Household and

Financial Literacy: Laboratory Evidence," Unpublished manuscript.

Penczynski, S. and M. Santana (2016): "Measuring Trust in Institutions: A Lab-in-the-Field Study Using Time Preference Elicitation," Unpublished manuscript.

Potters, J., A. Riedl, and R. Smeets (2016): "Towards a Practical and Scientifically Sound Tool for Measuring Time and Risk Preferences in Pension Savings Devisions," Unpublished manuscript.

Rong, R., M. Gnagey, and T. Grijalva (2018): "'The Less You Discount, the More It Shows You Really Care': Interpersonal Discounting in Households," *Journal of Economic Behavior and Organization*, 154, 1–23.

Savani, K., A. Sengupta, and D. Wang (2018): "The Cancellation Heuristic in Intertemporal Choice," Unpublished manuscript.

Sawada, Y. and Y. Kuroishi (2015a): "How Does a Natural Disaster Affect People's Preference? The case of a Large Scale Flood in the Philippines Using the Convex Time Budget Experiments," in *Disaster Risks, Social Preferences, and Policy Effects: Field Experiments in Selected ASEAN and East Asian Countries*, ed. by Y. Swada and S. Oum, Economic Research Institute for ASEAN and East Asia, chap. 2, 27–56.

——— (2015b): "How to Strengthen Social Capital in Disaster Affected Communities? The Case of the Great East Japan Earthquake," in *Disaster Risks, Social Preferences, and Policy Effects: Field Experiments in Selected ASEAN and East Asian Countries*, ed. by Y. Swada and S. Oum, Economic Research Institute for ASEAN and East Asia, chap. 6, 163–169.

Slonim, R., C. Wang, E. Garbarino, and D. Merrett (2013): "Opting-In: Participation Biases in Economic Experiments," *Journal of Economic Behavior & Organization*, 90, 43–70.

Stango, V., J. Yoong, and J. Zinman (2017): "The Quest for Parsimony in Behavioral Economics: New Methods and Evidence on Three Fronts," NBER Working Paper No. 23057.

Sun, C. and J. Potters (2016): "Magnitude Effect in Intertemporal Allocation Tasks," Unpublished manuscript.

Sutter, M., S. Angerer, D. Glatzle-Rutzler, and P. Lergetporer (2018): "Language Group Differences in Time Preferences: Evidence from Primary School Children in a Bilingual City," *European Economic Review*, 106, 21–34.

Yang, X. and F. Carlsson (2016): "Influence and Choice Shifts in Households: An Experimental Investigation," *Journal of Economic Psychology*, 53, 54–66.

References

Andreoni, J., M. A. Kuhn, and C. Sprenger (2015): "Measuring Time Preferences: A Comparison of Experimental Methods," *Journal of Economic Behavior & Organization*, 116, 451–464.

Andreoni, J. and C. Sprenger (2012): "Estimating Time Preferences from Convex Budgets," *American Economic Review*, 102, 3333–3356.

Andrews, I. and M. Kasy (2019): "Identification of and Correction for Publication Bias," *American Economic Review*, 109, 2766–2794.

Augenblick, N., M. Niederle, and C. Sprenger (2015): "Working Over Time: Dynamic Inconsistency in Real Effort Tasks," *Quarterly Journal of Economics*, 130, 1067–1115.

Furukawa, C. (2019): "Publication Bias under Aggregation Frictions: Theory, Evidence, and a New Correction Method," Unpublished manuscript.

Hedges, L. V., E. Tipton, and M. C. Johnson (2010): "Robust Variance Estimation in Meta-Regression with Dependent Effect Size Estimates," *Research Synthesis Methods*, 1, 39–65.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999): "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14, 382–401.

Moral-Benito, E. (2015): "Model Averaging in Economics: An Overview," *Journal of Economic Surveys*, 29, 46–75.

Simonsohn, U., L. D. Nelson, and J. P. Simmons (2014): "*P*-Curve: A Key to the File-Drawer." *Journal of Experimental Psychology: General*, 143, 534–547.

Stanley, T. D. and H. Doucouliagos (2015): "Neither Fixed nor Random: Weighted Least Squares Meta-Analysis," *Statistics in Medicine*, 34, 2116–2127.

Stanley, T. D., S. B. Jarrell, and H. Doucouliagos (2010): "Could It Be Better to Discard 90% of the Data? A Statistical Paradox," *The American Statistician*, 64, 70–77.

Steel, M. F. J. (forthcoming): "Model Averaging and its Use in Economics," *Journal of Economic Literature*.